DOKUMENTATIONEN
58/2015

Umweltinformations-systeme

Big Data – Open Data – Data Variety

Für Mensch & Umwelt

Umwelt Bundesamt

DOKUMENTATIONEN 58/2015

Umweltinformationssysteme

Big Data – Open Data – Data Variety

Ergebnisse des 21. und 22. Workshops des Arbeitskreises „Umweltinformationssysteme“ der Fachgruppe „Informatik im Umweltschutz“, veranstaltet an der Hochschule Karlsruhe - Technik und Wirtschaft (HsKA) am 22./23. Mai 2014 sowie am Wissenschaftlichen Zentrum für Informationstechnik-Gestaltung (ITeG) der Universität Kassel am 07./08. Mai 2015

Im Auftrag des Umweltbundesamtes

Impressum

Herausgeber:
Umweltbundesamt
Wörlitzer Platz 1
06844 Dessau-Roßlau
Tel: +49 340-2103-0
Fax: +49 340-2103-2285
info@umweltbundesamt.de
Internet: www.umweltbundesamt.de

/umweltbundesamt.de
/umweltbundesamt

Redaktion:
Abteilung IV 2 Arzneimittel, Chemikalien, Stoffuntersuchungen
Uta Zacharias

Publikationen als pdf:
http://www.umweltbundesamt.de/publikationen/umweltinformationssysteme-5

ISSN 2199-6571

Dessau-Roßlau, September 2015

Vorwort

Die vorliegende Publikation ist ein Beitrag zur Vermittlung von Informationen und Wissen aus dem Bereich der angewandten Umweltinformatik. Die seit nun mehr 20 Jahren stattfindenden Workshops des Arbeitskreises Umweltinformationssysteme bieten eine Plattform für den Wissenstransfer zu interdisziplinären Themen unter Anwendung innovativer Methoden und Werkzeuge der Informatik. Die Beiträge der Workshops von 2014 und 2015 werden neben der Publikation in einem Sammelband auch in dem internationalen Informationssystem ict-ensure „Informationssystem für Nachhaltige Umweltinformatik“ referenziert (http://www.iai.kit.edu/ictensure/). Damit wird auch das Auffinden dieser Beiträge über Suchmaschinen gefördert.

Der 21. Workshop des Arbeitskreises fand am 22./23. Mai 2014 an der Hochschule Karlsruhe - Technik und Wirtschaft (HsKA) statt. Unter dem Titel „Big Data, Open Data und Datenqualität“ wurde in 21 Beiträgen mit mehr als 35 Beteiligten reger Erfahrungsaustausch in zahlreichen Diskussionen praktiziert.

Die Ausrichtung dieses Workshops übernahm Prof. Dr. Detlef Günther-Diringer der HsKA, organisatorisch unterstützt von der Fachschaft Informationsmanagement und Medien (IMM). Die Sprecherin des Arbeitskreises Ulrike Freitag (Condat AG) koordinierte inhaltlich die Themen, wesentlich unterstützt von Dr. Andreas Abecker (Disy GmbH).

Das Thema „Big Data“ wurde mit einem Mini-Tutorium eingeführt und dann später in zahlreichen Diskussionen wieder aufgegriffen und in Beziehung zu den aktuellen Vorträgen gestellt. Dazu gibt der Beitrag über den Einsatz der Mongo-Datenbank (MongoDB) bei der Verwaltung meteorologischer Massendaten einen Einblick in praktische Erfahrungen einer umweltrelevanten Big Data-Anwendung. Vorgestellt wurde außerdem ein Vergleich verschiedener kommerzieller Cloud-Computing-Angebote zur Auslagerung von rechenintensiven Simulationen.

Einen weiteren Themenschwerpunkt bildeten Beiträge über das Anbieten und die Integration von Kartendarstellungen und Kartendiensten sowie einigen speziellen Diensten des OGC (Open Geospatial Consortium). Insbesondere der Einsatz von WPS (Web Processing Service) war Gegenstand eines regen Meinungsaustausches. Ein Thema, das 2015 wieder mit aktuellen Erkenntnissen im Fokus stand.

Der 22. Workshop wurde am 07./08. Mai 2015 von Prof. Dr. Gerd Stumme am Wissenschaftlichen Zentrum für Informationstechnik-Gestaltung (ITeG) der Universität Kassel ausgerichtet, inhaltlich koordiniert wieder von Ulrike Freitag, unterstützt von Dr. Andreas Abecker. Das Workshop-Motto „Umwelt.Daten.Vielfalt“ spiegelte sich in den vorgestellten Beiträgen anschaulich wider. Die fachlichen Inhalte symbolisieren die Bandbreite der Umweltinformatik und reichen von Energiewende, Landmanagement,

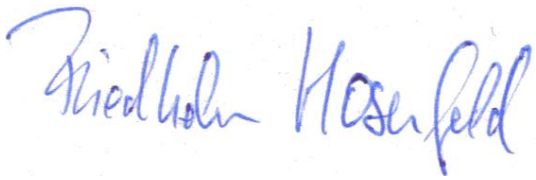
Forstwirtschaft, Meeresstrategie, Hochwasser und Katastrophenschutz über Flughafenlärm und Logistikanalysen der Bahn bis hin zu Lösungsmitteln in der Automobilbranche.

Aus technischer Sicht wurden einige Aspekte wie INSPIRE und OGC-Protokolle, die Integration von verschiedenen Informationsquellen in Umweltportale sowie der Einsatz der Auswertungs- und Data-Warehouse-Plattform Cadenza aus dem Vorjahr wieder aufgegriffen. Die Präsentation erfolgte jedoch aus einem anderen Blickwinkel oder mit unterschiedlichen Anwendungsfällen, so dass neue Diskussionsimpulse gegeben wurden.

Zusätzlich standen Apps für mobile Endgeräte auf dem Programm. Dies betraf Beispiele zum Crowdsourcing, zur Katastrophenschutz-Alarmierung aber auch für den Vorort-Einsatz in der Forstwirtschaft. Weitere Apps zu Datenaustauschportalen auf der Basis von OpenSource, Business Intelligence-Systemen mit Geovisualisierung sowie die semantische Wissens-repräsentation durch RDF (Ressource Description Framework) zeigten das breite Spektrum der Einsatzgebiete.

Nicht von allen Vorträgen der beiden Workshops liegen Langfassungen vor. Die Foliensätze aller freigegebenen Vorträge der beiden Workshops stehen jedoch zum Download auf der Homepage des Arbeitskreises http://www.ak-uis.de/ zur Verfügung.

Dort finden sich unter anderem auch Links auf die Tagungsbände der Workshops vorangegangener Jahre.



Dipl.-Inform. Friedhelm Hosenfeld, Sprecher des Arbeitskreises „Umweltinformationssysteme“

Dr. Gerlinde Knetsch, Umweltbundesamt,
Fachgebiet Informationssysteme Chemikaliensicherheit

August 2015

Inhaltsverzeichnis

Workshop 2014

22./23. Mai 2014
an der Hochschule Karlsruhe - Technik und Wirtschaft (HsKA)

Online-Zugriff auf INSENSUM Umwelt-Sensordaten unter Nutzung von OGC SWE Standards

Dr.-Ing. Siegbert Kunz, Fraunhofer Institut für Optronik, Systemtechnik und Bildauswertung IOSB,
Siegbert.Kunz@iosb.fraunhofer.de

Abstract

The Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB has designed for R&D purposes a concept and realized an integrated sensor net for the environment (INSENSUM) on its site in Karlsruhe. INSENSUM aimed in designing and realizing a completely integrated sensor system beginning from the Sensors itself up to the end user web interface by an overall web based approach.

1 Einleitung

Das Fraunhofer Institut für Optronik, Systemtechnik und Bildauswertung IOSB hat für Forschungszwecke ein integriertes Sensornetz für die Umwelt (INSENSUM) auf seinem Institutsfreigelände in Karlsruhe konzipiert und eingerichtet.

2 Zielsetzung INSENSUM

Zielsetzung war hierbei die Konzeption und Realisierung eines durchgehenden, integrierten Sensorsystems beginnend vom Sensor bis zur Benutzerschnittstelle in einem ganzheitlichen, webbasierten Ansatz.

INSENSUM dient dazu, Erfahrungen mit Sensor Web Enablement (SWE)-Standards des Open Geospatial Consortiums (OGC) zu sammeln und zu vertiefen, bei dem das IOSB als Technical Comittee Member vertreten ist. Neben der Bereitstellung einer Integrations- und Testplattform [Kunz, 2007], [Kazakos, 2009], [Kunz, 2010] in unserer Abteilung Informationsmanagement und Leittechnik sowie für Kooperationspartner war

INSENSUM auch ein Experimental-Testbed für unsere EU Projekte (z.B. SANY Sensor ANYwhere, FP6 EU-Projekt 2006-2009, siehe [Kunz, 2009], [Usländer, 2010], und EO2HEAVEN Earth Observation and ENVironmental modelling for the mitigation of HEAlth risks, FP7 EU-Projekt 2010-2013, siehe [Richter, 2010] ,[Kunz, 2013], [Jirka, 2013], [Brauner, 2013]).

Bei INSENSUM geht es auch um die Verarbeitung und Visualisierung der Unsicherheit von Daten, um die Entwicklung von Algorithmen für die Sensor Daten Fusion („Fusion4Decision"), um den mobilen Zugriff auf Umwelt- und Wetterdaten sowie um die Entwicklung offener Geo-Service Architekturen und integrierter Systeme [Kunz, 2009], [Kunz, 2009a], [Kunz, 2014].

3 Aufbau des integrierten Sensornetzes

INSENSUM als Testbed und Integrationsplattform umfasst auf dem IOSB Gelände mehrere ortsfeste Sensoren für die Domänen Wasser – Boden – Luft. In einem fischbesetzten institutseigenen Teich werden mittels einer Unterwasser-Sonde diverse Messgrößen (engl. observed properties) wie pH-Wert, Wassertemperatur, Wasserstand, Sauerstoffkonzentration und Trübung gemessen. Im Boden werden auf dem IOSB-Freigelände an einer Stelle in bis zu acht verschiedenen Tiefen Bodentemperaturen und Bodenfeuchtewerte erfasst. In der Domäne Luft wurden verteilt auf das Institutsfreigelände insgesamt vier Messstationen unterschiedlicher Hersteller mit Sensoren für Wetterdaten wie Lufttemperatur, Luftdruck/-feuchtigkeit, Windstärke und -richtung, Regenmenge, Sonneneinstrahlung, UVA/UVB-Einstrahlung sowie Energiestrahlung (up- und downlink in kurz- und langwelligen Frequenzbereichen) aufgebaut.

Um aufwändige Kabelverlegungsarbeiten ganz zu vermeiden und größtmögliche Flexibilität hinsichtlich evtl. Änderungen von Messstationsstandorten zu ermöglichen, wurden im INSENSUM-Architekturkonzept ausschließlich einerseits ganzjährig energieautarke Solarstromversorgungen und andererseits drahtlose Mobilfunkdatenübertragungen (via GPRS) vorgesehen und realisiert.

Da die Hersteller bei ihren Messstationen und Datenloggern typischerweise jeweils ihre meist nicht offengelegten, spezifischen Datenformate und -protokolle verwenden, mussten zusätzliche INSENSUM Applikationen erstellt werden, welche die Sensoren ins Web bringen (m.a.W. "*Sensor Web Enablement*"). Die INSENSUM-Architektur sieht dabei eine konsequente Nutzung von OGC SWE Standards vor. Mittels spezifischer Datenaufbereitungsapplikationen werden die Datenlogger-Messdaten geeignet nach OGC Standards und nach dem *Observation and Measurement Model (O&M)* gewandelt und eine SensorML-Beschreibung der Messstation mit allen meta-Daten (wie z.B. Messgrößen, Einheiten, Postionen) automatisch generiert. Diese startet initial eine *RegisterSensor* Operation zur Registrierung auf einem SOS-Server (open source Implementierung der Fa. 52North) und sendet fortwährend die O&M formatierten Messdaten zum SOS-Server via *InsertObservation* Operationen, wobei eine Pufferung bei evtl. fehlender Verbindung und späteres Nachholen bis zum Erfolg für die zuverlässige Datenübertragung sorgen.

4 Webbasierte Visualisierung der Sensordaten

Sobald die Messdaten auf dem SOS Server vorliegen und in dessen Datenbank gespeichert sind, kann mittels diversen am IOSB entwickelten Applikationen darauf online zugegriffen werden [Kunz, 2014], um an der in einer zoombaren Karte markierten Messstation beispielsweise den letzten aktuell verfügbaren Messwert einer oder mehrere Messgrößen z.B. in einem Browser (vorzugsweise Chrome) auf stationären oder mobilen Geräten anzuzeigen, siehe Abbildung 1.

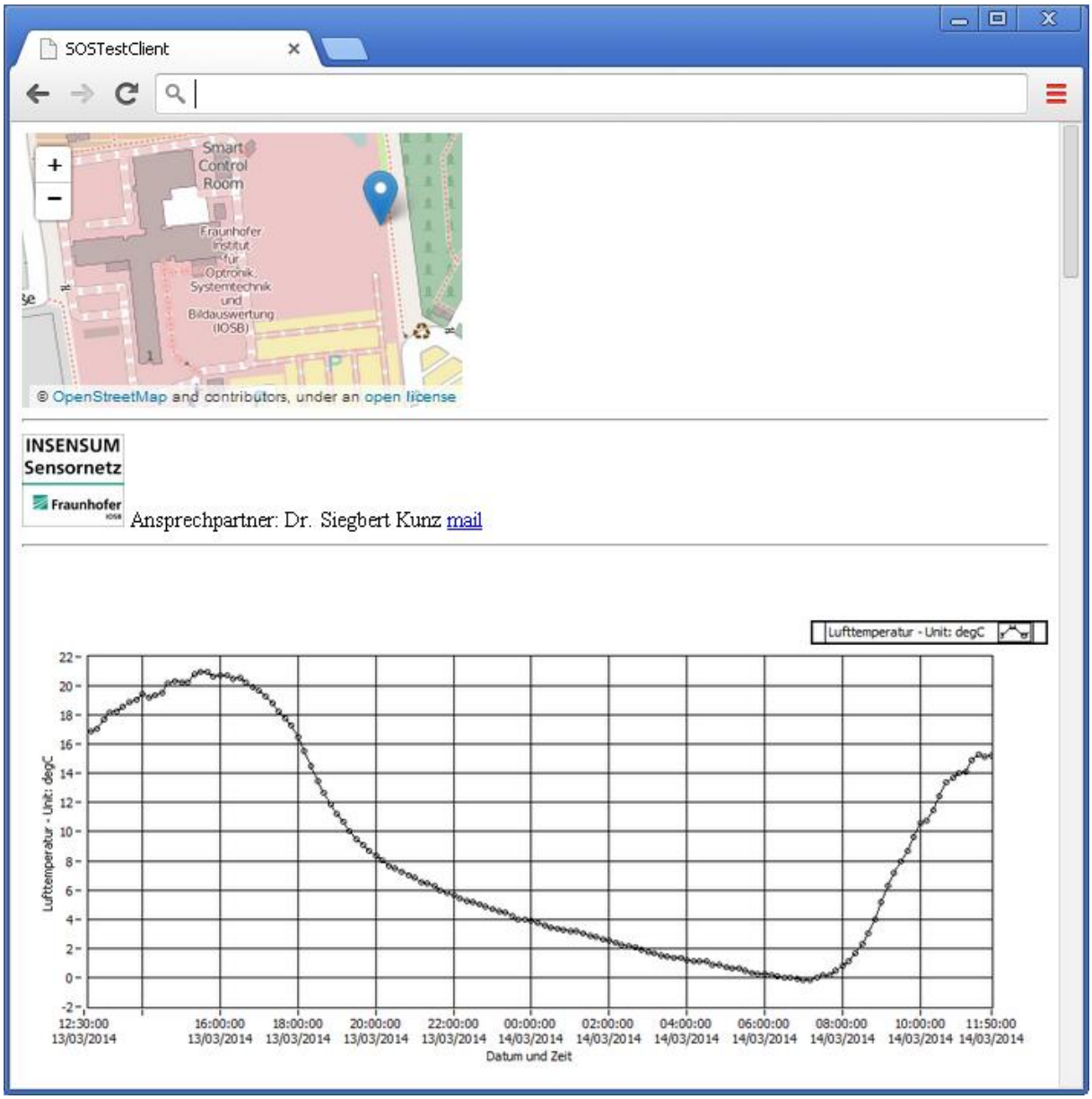


Abbildung 1: Lage der INSENUM Messstation und 24h-Messkurve (z.B. hier Lufttemperatur)

Daneben sind auch alle Messwerte über einen wählbaren Zeitraum bzw. eine gewählte Anzahl von Messwerten einer Messgröße abrufbar, die danach entweder in Tabellenform oder als Zeitverlaufskurven visualisierbar sind. Die Niederschlagswerte lassen sich als Momentanwerte sowie alternativ als kumulative tägliche Niederschlagsmenge z.B. über einen Monatszeitraum darstellen, zudem kann die Windsituation z.B. während der letzten 2 Stunden als eine sog. Windrose aufgezeigt werden, woraus man z.B. die Vorzugsrichtungen ablesen kann, woher der Wind in diesem Zeitraum überwiegend herkam, siehe Abbildung 2.

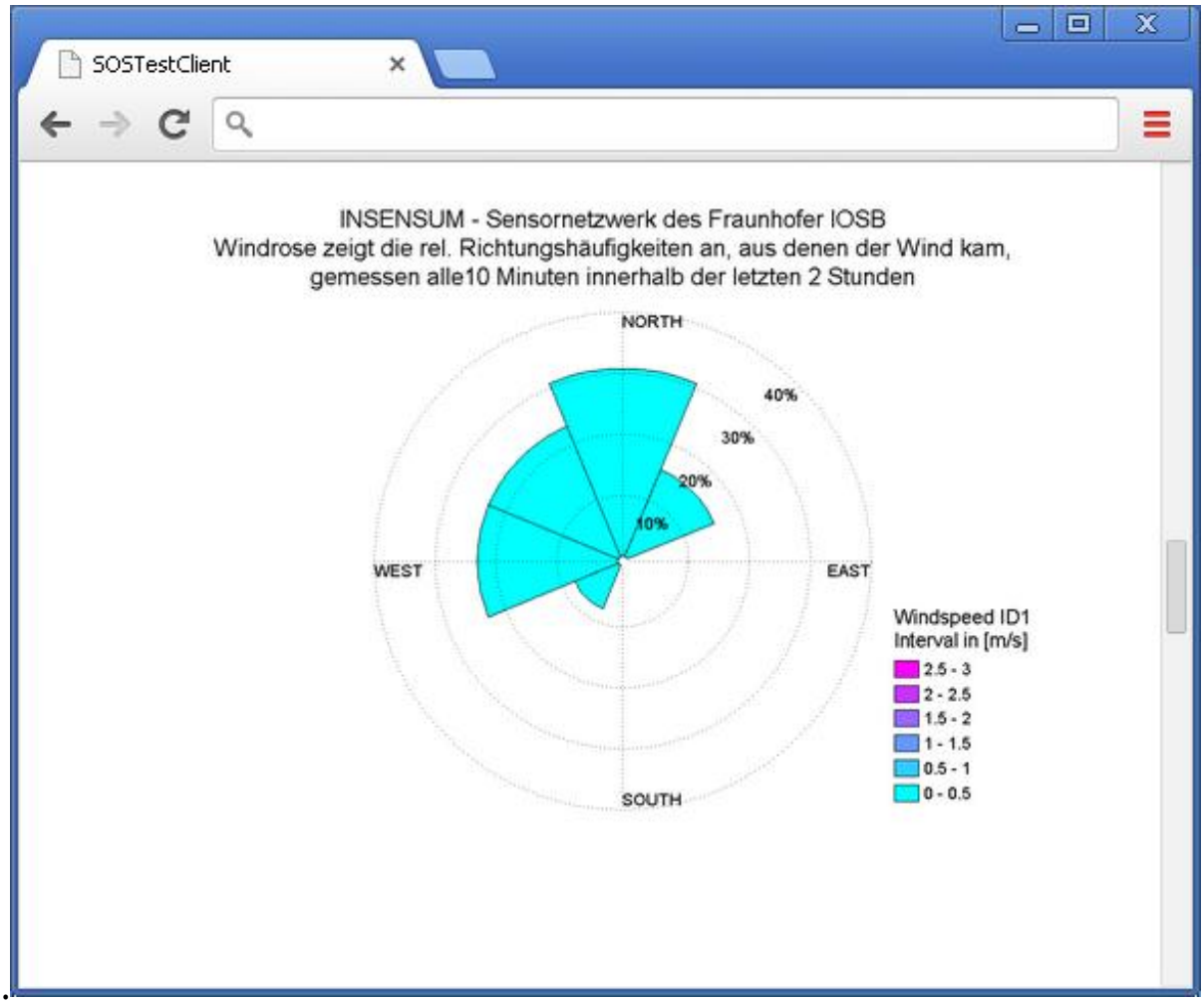


Abbildung 2: Die Darstellung einer Windrose illustriert die Wind-Vorzugsrichtungen

Auch lassen sich z.B. Messwerte an verschiedenen Messorten verknüpfen (Fusion) und beispielsweise daraus flächenhaft interpolierte Windverteilungen mittels eines Windfeld-Pfeildarstellung anschaulich visualisieren, optional auch eingefügt und als zeitliche Abfolge animiert in bekannte Tools wie z.B. Google Earth oder Google Maps. Jüngste Entwicklungen zielen auf die Verwendung von mobilen Endgeräten wie Smartphones und Tablets zur Darstellung der von INSENSUM bereitgestellten Messwerte ab, wobei aktuell eine prototypische App für die beiden Betriebssysteme IOS und Android entwickelt wird. Die eingeschränkte Performanz der Mobilendgeräte und die typischerweise oft begrenzte Datenübertragungsleistung in Mobilfunk-systemen wird dabei durch eine merklich effizientere, serverseitige Vorverarbeitungs-unterstützung und Datenübertragungsreduktion (z.B. via JSON) gut kompensiert.

5 Literaturverzeichnis

[Kunz, 2007]

Kunz, S.; Watson, K. (Eds.); Sensor Taxonomy, Public report D2.2.1 of the European Integrated Project SANY, 72 S., 2007. www.sany-ip.eu/biblio

[Kunz, 2009]

Kunz, S.; Watson, K. et al.; SANY Fusion and Modelling Architectural Design; Public report D3.3.2.3 of the European Integrated Project SANY, 159 S., 2009. www.sany-ip.eu/biblio

[Kunz, 2009a]

Kunz, S.; Usländer, T.; Watson, K.; A Testbed for Sensor Service Networks and the Fusion SOS: towards plug & measure in sensor networks for environmental monitoring; 18th World IMACS / MODSIM Congress, 13.07,2009, Cairns/Australia, S. 973-979, 2009. www.mssanz.org.au/modsim09/C4/kunz.pdf

[Kazakos, 2009]

Kazakos, W.; Heidmann, C.; Klenke, M.; Kunz, S.; Einführung von OGC SOS-Diensten für den Austausch und die Darstellung numerischer Daten in PortalU; GfI, Workshop AK Umweltdatenbanken/Umweltinformationssysteme, 4.-5-06.2009 in Hof. Veröffentlicht in UBA-Texte 01/2010. http://www.umweltbundesamt.de/uba-info-medien/mysql_medien.php?anfrage=Kennummer&Suchwort=3901, http://www.ak-uis.de/ws2009/05_Kazakos.pdf

[Kunz, 2010]

Kunz, S.; Watson, K. et al.; Sensor-Dienste-Architekturen auf der Grundlage von OGC Sensor-Web-Enablement-Diensten; Tagungsband des Workshops “ENERGIEAUTARKE SENSORNETZWERKE”, am Fraunhofer IZM, München, 16.11.2010 http://www.izm.fraunhofer.de/de/news_events/rueckblicke/workshop_energieautarkesensornetzwerke.html

[Richter, 2010]

Richter, S.; Brauner, J.; Kunz, S.; Integration von Fernerkundungs-, Geo- und Krankheitsdaten in Geodateninfrastrukturen für die Vorhersage von umweltbedingten Gesundheitsgefährdungen; Workshop Umweltinformationssysteme 2010, Leipzig, 27./28.05.2010, Veröffentlicht in UBA Texte 57/2010 http://www.ak-uis.de/ws2010/pdf/d_07_Richter_Silke_EO2HEAVEN_UIS-Leipzig_SR_final.pdf, http://www.umweltbundesamt.de/uba-info-medien/4042.html

[Usländer, 2010]

Usländer, T.; Kunz, S.; Watson, K. et al.; Specification of the Sensor Service Architecture, V3. Public report D2.3.4 of the European Integrated Project SANY, 2009. www.sany-ip.eu/biblio

[Kunz, 2013]

Kunz, S; Usländer, T. (Eds.); Specification of the SII Implementation Architecture; Public report D4.1.3 of the European Integrated Project EO2HEAVEN, 2013. www.eo2heaven.org/sites/default/files/content-files/documents/d413-specification-sii-implementation-architecture.pdf

[Jirka, 2013]

Jirka, S.; Kunz, S.; Usländer, T.; Watson, K. et al.; Specification of the Advanced Sensor Web Enablement SWE Concepts; Public report D4.1.4 of the European Integrated Project EO2HEAVEN http://www.eo2heaven.org/sites/default/files/content-files/documents/d414specificationoftheadvancedsweconceptsissue4.pdf

[Brauner, 2013]

Brauner, J.; Kunz, S.; Usländer, T.; Watson, K. et al.; Specification of the Advanced Geo-processing Services; Public report D4.1.5 of the European Integrated Project EO2HEAVEN. http://www.eo2heaven.org/sites/default/files/content-files/documents/d415specificationoftheadvancedgeo-processingservicesissue4.pdf

[Kunz, 2014]

Kunz, S; Günter, F.; Präsentationsfolien des IOSB-Beitrags “Online-Zugriff auf INSENSUM Umwelt-Sensordaten unter Nutzung von OGC SWE Standards” auf dem 21. Workshop AK Umweltinformationssysteme – UIS 2014 in Karlsruhe 22.-23.5.2014. http://www.ak-uis.de/ws2014/Vortraege.zip

Flexible Darstellung von geobezogenen Daten in Webseiten über ein ereignis-orientiertes Karten-Widget

Jan Düpmeier, Clemens Düpmeier, Thorsten Schlachter (KIT-IAI)
Wolfgang Schillinger (LUBW)

Abstract

Basing on the Google Maps API a web widget was developed in order to generate various digital maps combining different map layers. This web widget can be integrated within the web portal framework LUPO presenting spatial search results on maps complementing the list of fulltext search results.

Seit der Entwicklung des ersten Prototyps für ein Landesumweltportal (LUPO) im Jahr 2003, das vom Institut für Angewandte Informatik (IAI) des Karlsruher Institut für Technologie (KIT) für das Umweltinformationssystem (UIS) Baden-Württemberg im Auftrag des Ministerium für Umwelt, Klima und Energiewirtschaft Baden-Württemberg in Kooperation mit der Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württembergs entwickelt wird, gab es mehrere große Evolutionsschritte. Eine zentrale Rolle spielten dabei die Entwicklungen rund um die Volltextsuche, die inzwischen auch Möglichkeiten zum Zugriff auf Inhalte von Fachsystemen bietet, welche von externen Internet-Suchmaschinen nicht gefunden werden können.

Im Zuge der Entwicklung der nächsten Generation des Fachportal-Frameworks LUPO sollen nun geobasierte Suchergebnisse dynamisch in Form von Layerdarstellungen über einer Google Maps Grundkarte in Trefferlisten bei der Volltextsuche eingebettet und über eine gleichzeitig eingeblendete thematische und ortsbezogene Facettierung verfeinert werden können. Hierzu wurde auf Basis der Google Maps API ein Javascript-basiertes Web-Widget zur Kartendarstellung entwickelt, das basierend auf einem

zugrundeliegenden REST-Service als Konfigurationsserver dynamisch nahezu beliebige Kartenkonfigurationen anzeigen kann, die sich aus gängigen Google Maps Layertypen kombinieren lassen. Zurzeit unterstützt das Karten-Widget hierbei u.a. Google Maps Engine (GME) Layer, über die sich beliebige Raster- und Vektordaten, wie z.B. Schutzgebietslayer, darstellen lassen, Layer aus openGIS-Standard konformen WMS-Servern, Layer, die aus Google Fusion-Tables generiert werden, KML und GeoRSS-Layer und als neueste Google Maps Layerform auch die sogenannten Data Layer, die auf dem GeoJSON-Format basieren. Hiermit lassen sich alle geobezogenen Umweltdaten, die bei der LUBW verfügbar sind, über das Widget anzeigen.

Das Widget und der zugehörige REST-basierte Konfigurationsservice unterstützen dabei die logische Zusammenstellung einer Karte aus verschiedenen „virtuellen Layern“, die aus mehreren physikalischen Google Maps Layern bestehen können. So ist beispielsweise der virtuelle Layer zur Beurteilung des Solarenergiepotentials von Hausdächern in Baden-Württemberg aufgrund seiner großen Anzahl von Features in mehrere physikalische GME-Layer aufgeteilt, erscheint auf der Karte aber nur als ein integrierter virtueller Layer „Solarpotential von Dächern“. Des Weiteren lassen sich Layer im Konfigurationsservice zu Layergruppen zusammenfassen. So stellt z.B. die Layergruppe „Energie“ Energie-relevante virtuelle Layer zusammen, während die Layergruppe „Schutzgebiete“ alle Schutzgebietslayer, wie Naturschutzgebiete, Landschaftsschutzgebiete, etc. enthält.

Das Kartenwidget erlaubt es, über eine ereignisorientierte Schnittstelle zu jeder Zeit, dass einzelne Layer oder vollständige Layergruppen einer angezeigten Karten hinzugefügt, entfernt, angezeigt oder verdeckt werden können. Gibt ein Nutzer z.B. einen Suchbegriff bei der Volltextsuche ein, den das Backend der Suche unter den Oberbegriff „Energie“ einordnet, so erhält das Kartenwidget bei der Anzeige der Suchergebnisse über ein Ereignis die Anweisung die Layergruppe „Energie“ zu laden. Lässt sich der Suchbegriff (z.B. „solar“) direkt auf einen Layer in der Gruppe abbilden, wird zusätzlich dieser Layer der Layergruppe direkt angeschaltet und auf der Karte visualisiert. Die gesamte Layergruppe „Energie“ wird in diesem konkreten Beispiel (siehe Bild 1) als eine Facettierung der Suche links neben der Karte als Auswahlliste

„Energie“ angezeigt. Der Benutzer kann dann bei Bedarf weitere Layer der Gruppe auf der Karte anzeigen oder wieder ausblenden.

Bild 1: Suchergebnisseite nach der Suche nach den Begriffen „solar Neuhausen“.

Neben Ereignissen, die Layergruppen oder Layer im Kartenwidget anzeigen oder ausblenden, unterstützt das Kartenwidget auch ortsbezogene Ereignisse. So wurde bei der Suche im Portal neben dem thematischen Begriff „solar“ auch der ortsbezogene Begriff „Neuhausen“ eingegeben. Dieser wurde vom Geolokationsservice auf der Serverseite des LUPO-Portals in die zwei Orte „Neuhausen auf den Fildern“ und „Neuhausen ob Eck“ aufgelöst, zu denen jeweils Boundingboxen ermittelt wurden. Wie man man in Bild 1 sieht, werden beide Alternativen als Facetten links neben der Karte angezeigt, wobei die Karte über ein Positionierungsereignis auf den ersten Ort fokussiert wurde. Durch Klick auf die andere Ortsbezeichnung in der Facettierung triggert der Benutzer ein Ereignis, das die Karte auf den anderen Ort positioniert.

Wie im obigen Beispiel verdeutlicht wurde, lässt sich das Kartenwidget durch Senden von Ereignissen flexibel von anderen Interaktionselementen in der Suchergebnisseite aus steuern. Umgekehrt lösen Interaktionen in der Karte, z.B. das Klicken auf ein Layerobjekt Aktionen aus, die auch Ereignisse an andere Teile der Webseite senden

können. So kann das Klickereignis z.B. innerhalb eines Informationsbereiches der Webseite objektspezifische Informationen anzeigen oder direkt über ein Overlay auf der Karte visualisieren. Weitere Funktionalitäten sind durch dieses Ereignismodell denkbar und sollen in Zukunft realisiert werden.

Im Rahmen des Vortrags wurden der aktuelle Stand des Kartenwidgets, des dazugehörigen Konfigurationsservices und deren technische Implementierung und Anwendungsmöglichkeiten an Hand von konkreten Anwendungsbeispielen erläutert. Zum Abschluss wurde ein Ausblick auf den weiteren Ausbau der Funktionalitäten gegeben.

Professionelles Metadaten-Management in Geodateninfrastrukturen mit Preludio und eENVplus – Status und Perspektiven –

Andreas Abecker[1], Riccardo Albertoni[2], Carsten Heidmann[1], Wassilios Kazakos[1], Monica de Martino[2], Radoslav Nedkov[1], Yashar Moradiafkan[1], Roman Wössner[1]

[1] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe (DE), *vorname.nachname@disy.net*

[2] CNR-IMATI, Via De Marini, 6, Torre di Francia, 16149 Genova (IT), *vorname@ge.imati.cnr.it*

Abstract

With its Web-based software solution Preludio, Disy Informationssysteme GmbH offers one of the most powerful and at the same time user-friendly solutions for standards-compliant metadata management in Spatial Data Infrastructures (SDI). Preludio supports editing, discovery, management and exchange of metadata for data, services and applications in an SDI. In this article, we describe the most important features of Preludio. Furthermore, we describe ongoing research and development in the European Research project eENVplus aiming at powerful new tools for multi-thesaurus management for metadata.

Einleitung

Das Software-Werkzeug Preludio der Disy Informationssysteme GmbH ist seit vielen Jahren eine der mächtigsten und zugleich benutzerfreundlichsten Web-basierten Lösungen zum standard-konformen Metadaten-Management in Geodateninfrastrukturen (GDI). Preludio unterstützt Erfassung, Suche, Verwaltung und Austausch von Metadaten zu Daten, Diensten und Anwendungen in einer GDI. In diesem Artikel stellen

wir einige der wichtigsten Charakteristika von Preludio. Weiterhin skizzieren wir aktuelle Forschungsarbeiten im EU-Projekt eENVplus, die darauf abzielen, Fachthesauri für das Metadatenmanagement besser nutzen zu können.

1 Einleitung: Metadaten in Geodateninfrastrukturen

Metadaten sind „Daten über Daten". Sie stellen dem Nutzer eine strukturierte Beschreibung der eigentlichen Geodaten und Geodienste zur Verfügung und sind somit ein elementarer Baustein einer Geodateninfrastruktur. Wichtige Aspekte der Metadaten zu einem Geodatensatz sind z.B. Aktualitäts- und Qualitätsangaben, Format, Raum- und Zeitbezug sowie eine inhaltliche Beschreibung einschließlich der Sachdaten und Attribute. Metadaten ermöglichen es zunächst einmal, für einen bestimmten Zweck die passenden Geodaten zu finden. Darüber hinaus können Metadaten helfen bei der Vermeidung redundanter Datenerfassung, beim Aufdecken vorhandener Lücken in den Datenbeständen, beim Vergleich von Datenbeständen u.a.m. Für Metadatenerfassung und -verwaltung gibt es kommerzielle wie auch Open Source Werkzeuge. In diesem Artikel stellen wir mit Preludio[1] das entsprechende Werkzeug der Disy Informationssysteme GmbH vor. Die Bereitstellung und Nutzung von Metadaten erfolgt über sogenannte Katalogdienste (Catalog Services for the Web, CSW). Wenn es um die inhaltliche Beschreibung von Geodatensätzen geht, verwendet man i.d.R. Schlagwörter, die häufig aus kontrollierten Vokabularen stammen, also aus Fachthesauri, Fachontologien, festen Code-Listen, etc. (sog. *Knowledge Organization Systems,* KOS). Da bei der anwendungsübergreifenden Suche, bei komplexen Fragestellungen mit Bezug zu mehreren Fachbereichen oder bei grenzüberschreitenden Anwendungen u.U. mehrere solcher KOS zusammen betrachtet werden müssten, stellt sich die Frage nach dem Zusammenhang zwischen unterschiedlichen KOS bzw. nach der gleichzeitigen Nutzung mehrerer KOS. Diese Themen werden (neben vielen anderen) im EU-Projekt

[1] http://www.disy.net/produkte/preludio.html

eENVplus[2] behandelt. Die entsprechenden Ansätze und Entwicklungen aus eENVplus werden ebenfalls im vorliegenden Artikel dargestellt.

2 Das Metadaten-Management Werkzeug Disy Preludio

Disy Preludio ist eine Web-basierte und hochgradig konfigurierbare (Metadatenschema und GUI-Layout) orientierte Software-Lösung für Erfassung, Verwaltung und Nutzung von Metadaten in Geodateninfrastrukturen. Preludio unterstützt alle relevanten Metadaten-Standards wie ISO 19115, ISO 19119 und ISO 19139. Ferner werden die wesentlichen OGC und W3C Schnittstellen-Standards für Datenbereitstellung und -zugriff implementiert. Preludio kann als Metadaten-Server in eine umfassendere GDI eingebunden werden und implementiert dazu den OGC CSW 2.0.2 Standard mit ISO API 1.0. Der integrierte Map-Viewer für kartenbasierte Datendarstellung und -suche basiert auf dem OGC WMS-Standard. Über den Map-Viewer kann auch auf externe Kartendienste zugegriffen werden, um Geodaten anderer Organisationen im Web darzustellen. Metadatenprofile in Preludio sind kundenspezifisch anpassbar. Ebenso kann in derselben Anwendung mit mehreren Metadatenprofilen gleichzeitig gearbeitet werden. Für die Suche nach Metadaten wird die Recherche im Volltextmodus sowie nach Raum, Zeitbezug oder Stichwörtern angeboten. Die Datenhaltung von Preludio erfolgt in einer relationalen Datenbank (HSQL als Standard; andere Datenbanken wie Oracle oder PostgreSQL können leicht einkonfiguriert werden).

Die Web-Anwendung ermöglicht nicht nur Volltextsuche und strukturierte Metadatensuche im Geodatenbestand; durch den integrierten Open Source Map Viewer (realisiert mit Open Layers und Legato) können auch geographische Suchen nach Datensätzen mit einem bestimmten Orts- oder Raumbezug einfach realisiert werden.

[2] http://www.eenvplus.eu/

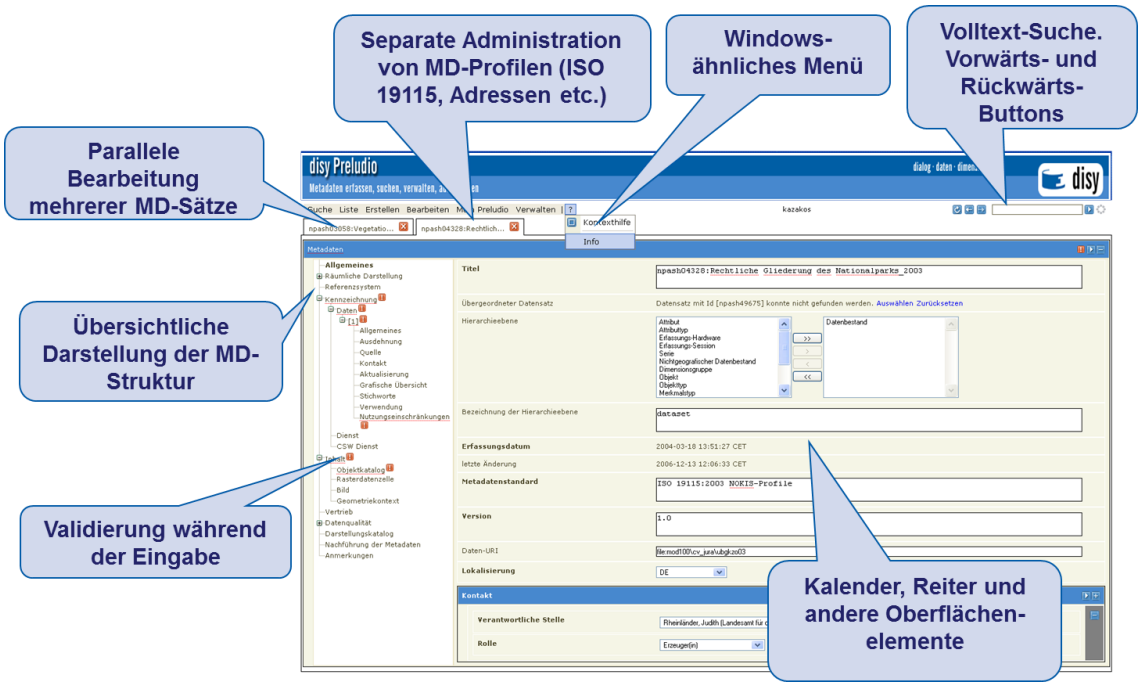


Abbildung 1: Preludio-GUI zur Metadaten-Erfassung

Abbildung 1 erläutert die wichtigsten Maßnahmen zur Erhöhung der Gebrauchstauglichkeit, **Benutzerfreundlichkeit und Benutzungseffizienz**: Windows-ähnliche Menüs und Strukturbäume erleichtern dem Windows-Nutzer das Zurechtfinden. Mehrere Metadatensätze können gleichzeitig erfasst und manipuliert werden; Während der Eingabe greifen verschiedene Mechanismen zur Gültigkeitsprüfung und Metadatenvalidierung. Diese Mechanismen reichen von der Angabe von Default-Werten für Metadatenfelder über die Spezifikation bedingter Defaults (Werte können abhängig von anderen Eingaben automatisch berechnet werden) bis zur Prüfung gegen die INSPIRE-Implementierungsregeln (vgl. Abbildung 2). Kürzlich wurde weiterhin die GDI-DE Testsuite[3] zur Validierung eingebunden.

Um den Anwender bei der Erzeugung **qualitativ hochwertiger Metadaten** zu unterstützen, kommen verschiedene Mechanismen zum Einsatz: Zunächst erhält der Benutzer durch das Werkzeug eine Reihe von direkten Hinweisen und Informationen. Diese betreffen insbesondere **Warnmeldungen zur Gültigkeit und Vollständigkeit** von Metadaten. Abbildung 2 zeigt, wie auf die entsprechenden Informationen auf der Benutzeroberfläche hingewiesen wird. In Abbildung 3 wiederum ist die **Kontexthilfe** zu

[3] http://www.geoportal.de/DE/GDI-DE/Komponenten/GDI-DE-Testsuite/gdi-de-testsuite.html?lang=de

sehen, mit der insbesondere Erläuterungen zu allen Metadaten-Elementen präsentiert werden können.

Abbildung 2: Warnmeldungen zur Gültigkeit und Vollständigkeit von Metadaten

Weiterhin unterstützt Preludio verschiedene **Freigabeebenen**, mit denen beispielsweise ein 4-Augen-Prinzip realisiert werden kann. Grundsätzlich sind einzelne Nutzer in der Regel authentifiziert und mit spezifischen **Benutzerrollen** versehen.

Ferner hat jedes Dokument einen **Freigabestatus**. Für einen bestimmten Nutzer und ein bestimmtes Element (Datensatz, Dokument, Information) ergibt sich aus der jeweiligen Kombination von Benutzerrolle und Freigabestatus die Sichtbarkeit bzw. Änderbarkeit für bzw. durch diesen Nutzer.

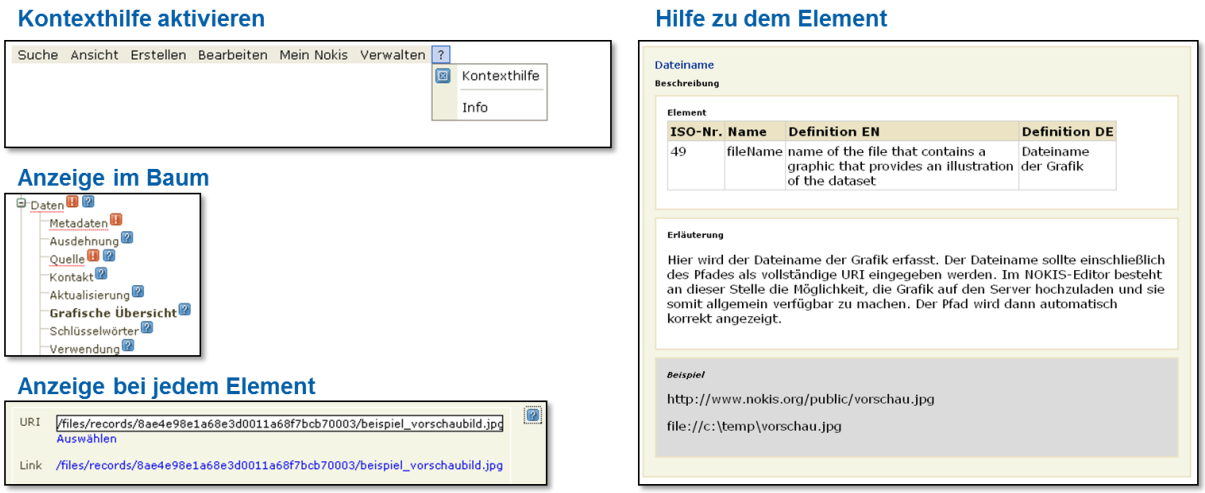


Abbildung 3: Kontexthilfe in Preludio

Preludio unterscheidet folgende Benutzerrollen und Freigabekonzepte:

Gast (nicht angemeldeter Benutzer)	Kann nur suchen und anzeigen.
User	Kann „eigene“ Datensätze ändern. Kann deren Bearbeitung als abgeschlossen markieren.
Superuser	Kann alle Datensätze ändern. Kann Datensätze intern und extern freigeben.
Gruppenadministrator	Administration aller Gruppenmetadaten.
Admin	Kann neue Benutzer und Rollen anlegen. Verwaltungsfunktionen.

Zur Illustration des Rollenkonzepts betrachten wir Abbildung 4 und Abbildung 5: Erstere zeigt den Bildschirm eines „Gasts“: dieser sieht nur extern freigegebene Datensätze und erhält ein minimales Menü und Kontextmenü sowie eine reduzierte Listendarstellung.

Abbildung 4: Listenansicht eines Gasts (Anfrage "Bayern")

Dagegen enthält die Listenansicht eines Superusers (Abbildung 5) sowohl intern und extern freigegebene als auch „eigene“ Datensätze, ein erweitertes Kontextmenü und eine erweiterte Listendarstellung.

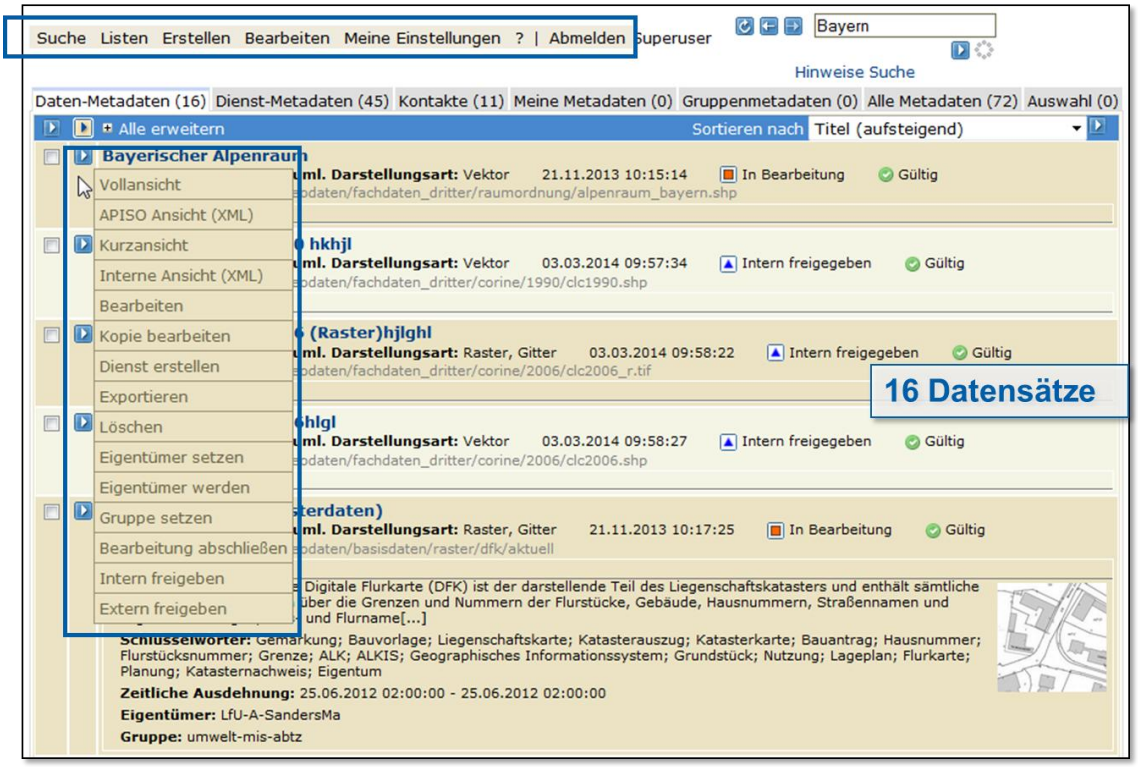


Abbildung 5: Listenansicht eines Superusers (Anfrage "Bayern")

Eine weitere, für die Praxis häufig bedeutende, Eigenschaft von Preludio betrifft die Tatsache, dass Metadatenprofile sehr einfach individuell angepasst werden können und dass innerhalb derselben Preludio-Installation für verschiedene Inhalte auch mit unterschiedlichen Metadatenprofilen gearbeitet werden kann. Zum Beispiel zeigt Abbildung 6 einen kundenspezifisch angepassten Metadaten-Editor, wie er beim Bayerischen Landesamt für Umwelt (LfU) im Einsatz ist.

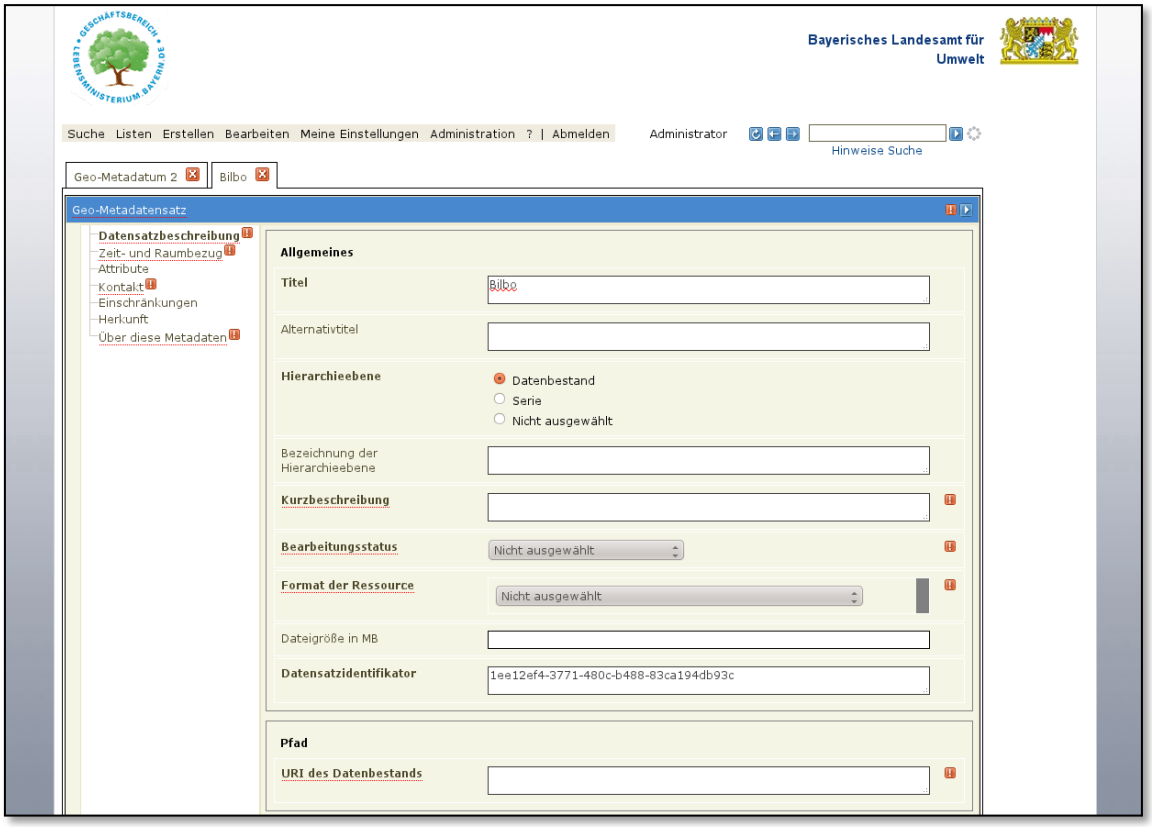


Abbildung 6: Preludio im LfU Bayern – Editor für angepasstes Metadatenprofil

Ab Version Preludio 2015 gibt es auch Plug-ins für Disy Cadenza und für ArcGIS Desktop. Über diese GIS-Anbindung kann anhand der Metadaten direkt aus einem GIS heraus nach Geodaten gesucht und diese können dann direkt ins GIS geladen werden. Zur Vereinfachung der Eingabe besteht auch die Möglichkeit, Metadaten direkt aus dem GIS zu erzeugen – natürlich einschließlich automatischer Erstellung entsprechender technischer Metadaten.

3 Multi-Thesaurus Management im eENVplus Projekt

Vorerst noch unabhängig von Preludio entwickelt, aber bei Bedarf einfach ankoppelbar, sind die prototypischen Ansätze zum Multi-Thesaurus Management im Forschungsprojekt eENVplus.

3.1 Idee und Grobkonzept

In eENVplus wird ein Thesaurus Framework (TF) entwickelt, das den dienste-basierten Zugriff auf ein Netzwerk von **als Linked-Open-Data dargestellten, miteinander**

verlinkten Fachvokabularen bzw. Fachthesauri ermöglicht. Das TF macht Fachvokabulare (KOS), die im **Semantic Web Format SKOS** / **RDF** repräsentiert sind, über Web-Dienste verfügbar und erlaubt es, Verbindungen zwischen diesen Vokabularen herzustellen, so dass Sprach- und Domänengrenzen zwischen unterschiedlichen Thesauri überbrückt werden können.

Das Thesaurus-Framework bietet eine Reihe von **Thesaurus-Nutzungsdiensten** (TF Exploitation Services, TFES) an. Dies sind Web-Dienste, die das in den verlinkten Thesauri enthaltene Wissen **zur Unterstützung der Metadaten-Eingabe, zur Vereinfachung der Metadaten-Pflege und zur metadatenbasierten Suche** nutzen. Diese Dienste können von anderen Werkzeugen aufgerufen werden, z.B. von Metadaten-Editoren, Metadaten-Validatoren, Katalogdiensten, Informationsportalen oder spezifischen Anwendungslösungen. Einige beispielhafte mögliche Verwendungen der TFES:

- Die eENVplus TFES können Synonyme, allgemeinere, speziellere oder verwandte Begriffe (*synonyms, broader, narrower, related terms* in SKOS) aus verschiedenen verbundenen Thesauri (oder anderen KOS) für Suchanwendungen anbieten, um so die themenübergreifende, korpus-übergreifende oder mehrsprachige Suche zu unterstützen. Wie die Suchanwendung genau die angebotene Hintergrundinformation nutzt und wie sie das auf der GUI darstellt, bleibt dieser Drittanwendung überlassen. Durch die Mächtigkeit des im Thesaurus Framework hinterlegten Wissens sind alle Verwendungsarten denkbar, wie man sie aus der **semantischen Suche** [8] kennt.
- Wenn ein Anwender, der auf einem bestimmten Gebiet unsicher ist, dennoch Metadaten auf diesem Feld erstellen muss, kann ein **Visualisierungsdienst zur explorativen Navigation** durch die konzeptuellen Räume der verschiedenen verbundenen Thesauri genutzt werden. Weiterhin können die hinterlegten Begriffsdefinitionen und Übersetzungen das Verständnis der Domäne und der verwendbaren Thesauri und Fachvokabulare verbessern.

- Automatisches **"Cross-walking"** zwischen verschiedenen verlinkten Fachvokabularen kann dabei helfen, leichter die Einschränkungen einzelner Thesauri/ Vokabulare zu verstehen bzw. zu überwinden.

Abbildung 7 illustriert das Zusammenspiel der einzelnen Komponenten: Ausgangspunkt ist die von Haus aus multikulturelle, multilinguale und multidisziplinäre Natur der Anwendungsdomäne „Umwelt". Zum Management heterogener Umweltdaten innerhalb von Initiativen wie *"INSPIRE – Infrastructure for Spatial Information in the European Community"* oder *"SEIS - Shared Environmental Information System"* werden terminologische Ressourcen wie Thesauri weithin genutzt als gemeinsame Grundlage der Kommunikation zwischen unterschiedlichen Anwendergruppen in umweltbezogenen Anwendungsgebieten. Sie ermöglichen es, wissenschaftlich-technische Fachbegriffe auszutauschen, sich auf ihre Bedeutung und Nutzung zu einigen und sie in unterschiedliche Sprachen zu übersetzen. In den vergangenen Jahren wurden viele solcher Thesauri von verschiedenen Nutzergruppen mit einem breiten Spektrum von Kompetenzen entwickelt. Sie unterscheiden sich in Umfang/ Abdeckung und in ihrer Granularität, teilweise auch in der Perspektive, die man zu einem bestimmten Sachgegenstand einnimmt. All diese Thesauri sind wertvoll und ihre Nutzung und Wiederverwendung in Geodateninfrastrukturen (GDI) ist wichtig. Sie können aber durchaus fachlich oder syntaktisch überlappend, redundant oder gar widersprüchlich sein. Gleichzeitig wäre es sowohl ein enormer Aufwand als auch fachlich unangemessen, alle existierenden Fachthesauri in *eine einzige*, allumfassende „richtige" Wissensstruktur zu überführen. Daher verbleiben die bereits existierenden Thesauri zur Pflege beim Ersteller, sofern er die Veröffentlichung entsprechend den Linked Open Data (LOD) Prinzipien sicherstellen kann (als SKOS / RDF) [3]. Im Thesaurus Framework werden dann nur die Verlinkungen repräsentiert, die auf die terminologischen Originalressourcen (per URI) verweisen. Kann ein Thesaurus-Anbieter die Veröffentlichung als LOD nicht realisieren, kann natürlich auch der gesamte Thesaurus-Inhalt als RDF im TF gespeichert werden.

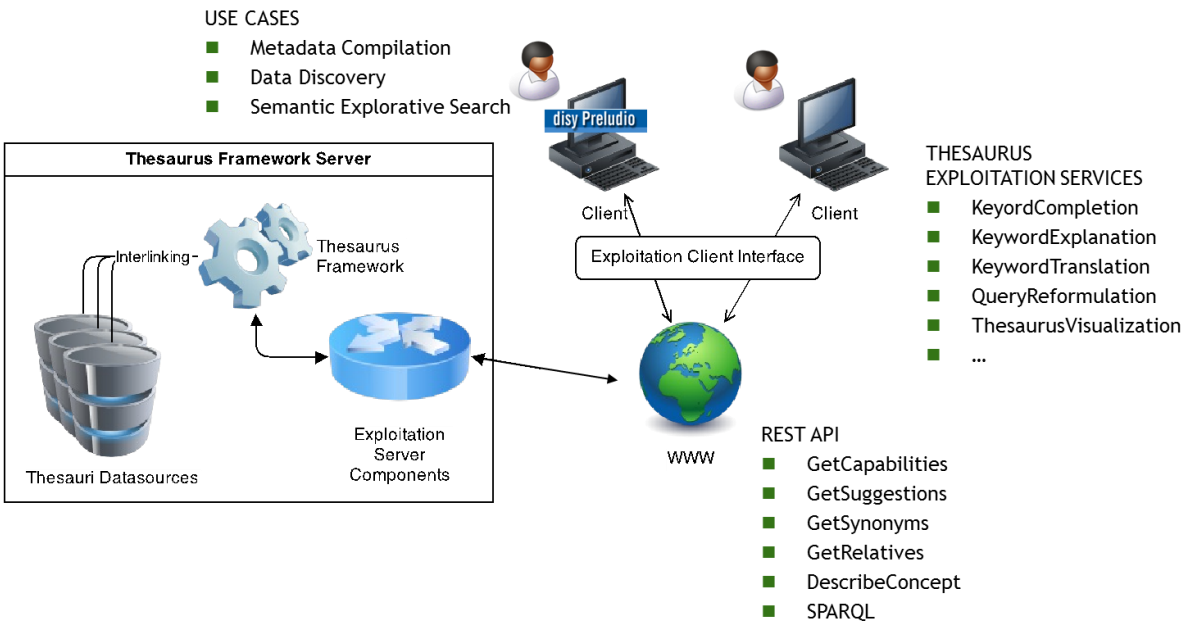


Abbildung 7: Illustration des Zusammenspiels der eENVplus Komponenten zur Nutzung des Thesaurus-Frameworks

Die Infrastruktur des Thesaurus Frameworks adressiert die Bedürfnisse verschiedener Nutzergruppen beim grenzübergreifenden Austausch digitaler Informationen, d.h. insbesondere Konzept-Interoperabilität und Konzept-Verfügbarkeit in verschiedenen Sprachen, wie es für die Erzeugung von Metadaten und die Suche nach Informationen benötigt wird. Weiterhin unterstützt das Software-Framework die **Kombination verschiedener existierender Thesauri** zum Management umweltbezogener digitaler Inhalte. Dieses Framework betrachtet die Heterogenität von Umweltthesauri in Umfang und Detaillierungsgrad als Kapital, wenn es um den Umgang von Umweltdaten geht; daher kommen Best Practices aus dem Bereich der Linked Data [3, 4] zum Einsatz, um eine Multi-Thesaurus Lösung für umweltbezogene INSPIRE-Datenthemen anzubieten. Das eENVplus TF erweitert die bestehende Lösung für den Naturschutz [5] aus dem NatureSDIplus[4] Projekt. Es basiert auf existierenden KOS, die entsprechend SKOS (Simple Knowledge Organization System) strukturiert und im Datenmodell RDF (Resource Description Framework) enkodiert sind. Das TF bietet eine integrierte Sicht auf verschiedene verfügbare KOS. Es ist eine offene Lösung, mithilfe derer allgemeine KOS ebenso wie domänenspezifische Terminologien hinzugefügt, zusammengeführt

[4] www.nature-sdi.eu

und ausgetauscht werden können. Das Framework ist insofern dynamisch, als Thesauri in einfacher Weise geändert, erweitert oder auch neue hinzugefügt werden können.

Wir sehen weiterhin in Abbildung 7, dass die TF Nutzungsdienste vom Exploitation Server über eine REST-Schnittstelle den Client-Lösungen zugänglich gemacht werden. Nutzungsdienste sind an den Bedürfnissen der Endnutzer orientierte Funktionalitäten wie Schlüsselwort-Vervollständigung, -Erklärung, -Übersetzung, aber auch komplexere Dienste wie die Umformulierung einer Suchanfrage mithilfe von Thesauruswissen usw. Diese Funktionalitäten werden in Dritt-Software eingebunden – dies könnte zum Beispiel für ein Metadaten-Managementsystem wie Preludio geschehen. Die Anwendungsfälle auf der Nutzerseite sind dann insbesondere die Metadaten-Erstellung, Metadatensuche und explorative Suche durch Datenbestände.

Beim Design des Thesaurus Framework wurden folgende Ziele verfolgt:

- **Modularität**. Jedes KOS sollte modular in die Menge der vom Framework verwalteten Thesauri einklinkbar sein. Modularität ist auch erforderlich, um mit eventuellen zukünftigen Änderungen in existierenden Terminologien umgehen zu können.
- **Offenheit.** Jedes KOS sollte in einfacher Weise erweiterbar sein, um z.B. neue Konzepte und Terme einzufügen, ohne den Original-Thesaurus zu tangieren.
- **Nutzbarkeit.** Jedes KOS sollte in einem flexiblen und standardisierten Format vorliegen, um die Nutzung und eventuell auch Erweiterung durch Drittsysteme anzuregen.
- **Verlinktheit.** Terme und Konzepte in existierenden KOS sollten miteinander vernetzt werden (verlinkt), um die Termverwendung aus multikultureller Sicht zu harmonisieren.

Technologien aus dem Gebiet des Semantic Web können helfen, diese Anforderungen zu erfüllen: Daher liegt das Standardmodell SKOS (Simple Knowledge Organization

System[5]) der Kodierung aller KOS zugrunde und die Prinzipien von Linked Data[6] werden angewandt, um Thesauri mithilfe von weltweit im WWW auffindbaren Uniform Resource Identifiers (URIs) zu publizieren, zu teilen und zu verbinden. Die Übersetzung terminologischer Ressourcen nach SKOS unterstützt insbesondere die oben geforderte Modularität – während die Ressourcenverfügbarkeit aufgrund der Linked Data Realisierung einen Schlüsselaspekt für Offenheit und Nutzbarkeit darstellt.

Im jeweils aktuellen Entwicklungsstatus ist das eENVplus Thesaurus Framework im Web zugreifbar.[7] Es enthält verschiedene SKOS/RDF Ressourcen, wie insbesondere den **Thist** Thesaurus über Geologie, die **EUNIS** Species und Habitat Typen, die Digitale Karte der ökologischen Regionen in der EU (**DMEER**) und den Referenzthesaurus für Umweltanwendungen (**EARTh**), welcher als Basisthesaurus im Framework verwendet wird. Weiterhin beinhaltet sind *Intra-Thesaurus Verlinkungen* zwischen Habitaten und Species sowie DMEER und EARTh und eine Menge von *Inter-Thesaurus Verlinkungen* zwischen den EUNIS Species und den offiziell von der Europäischen Umweltagentur vorgegebenen EUNIS Species sowie von EARTh zu **GEMET**. Weitere Inter-Thesaurus Äquivalenzen wurden jüngst hinzugefügt [1]. GEMET's ausgehende Links zu **AGROVOC**, **EUROVOC**, **DBpedia** und **UMTHES** wurden nach EARTh durch GEMET's *skos:exactMatch* Beziehungen realisiert. Naturgemäß betreffen die durch diese Prozedur erstellten Verbindungen nur die Teilmenge von Konzepten, die sowohl in EARTh als auch in GEMET enthalten sind. Um diese Konzeptmenge zu ergänzen und eine vollständigere Verbindung zwischen den in EARTh und in GEMET enthaltenen Linked-Data Mengen zu erhalten, wurde ein zweistufiger Prozess durchgeführt: Zunächst wurde das Werkzeug SILK[8] angewandt, um neue Verbindungen zu entdecken; dann wurden die SILK-Resultate durch Experten validiert, um die Genauigkeit der Links zu bewerten und die geeignetsten Typen für die Interlinking Property (*skos:exactMatch* bzw. *skos:closeMatch*) zu identifizieren. Die gemeinsame Verwendung des transitiven Abschlusses von *skos:exactMatch* und der manuell validierten, von SILK vor-

[5] http://www.w3.org/2004/02/skos/
[6] http://www.w3.org/wiki/LinkedData
[7] http://linkeddata.ge.imati.cnr.it:2020/
[8] http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/

geschlagenen Links hat die Anzahl der ausgehenden Verbindungen, bezogen auf die vorherigen Versionen von EARTh, fast verdreifacht. So wurden 7171 Verbindungen mithilfe des transitiven Abschlusses entdeckt und weitere 465 durch SILK erzeugt. Die so entstandene neue Version von EARTh bereitet den Weg für die gemeinsame Nutzung des eENVplus Thesaurus Frameworks mit GEMET, AGROVOC, EUROVOC, DBpedia und UMTHES (etwa 33% der EARTh Konzepte besitzen eine Verbindung zu einem anderen Thesaurus), so dass die Anwender deren jeweilige Stärken und Komplementaritäten ausnutzen können.

3.2 Architektur der TF-Nutzungsdienste

Um die Verwendung des im TF zugreifbaren Wissens effektiv, nachhaltig und erweiterbar gestalten zu können, wurde die in Abbildung 8 gezeigte Architektur gewählt: Für den Zugriff von außen durch die Clients von Dritt-Software wird vom Thesaurus Exploitation Server als http Server eine **REST-Schnittstelle** angeboten. Diese erlaubt den Zugriff auf den **Service Manager**; der wiederum auf eine nichtleere Menge von **Service-Modulen** zugreift. Im Wesentlichen leitet der Service-Manager die Anfragen an geeignete Service-Module weiter (basierend auf deren veröffentlichten *Capabilities*), sammelt die von den Service-Modulen gelieferten *Callables* (ausführbare Code-Einheiten) ein, orchestriert deren Ausführung, führt die Ergebnisse zusammen und gibt sie zurück an den HTTP-Server.

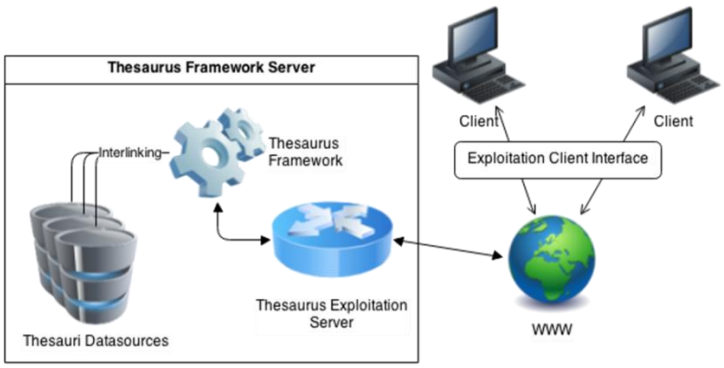


Abbildung 8: Nutzung verlinkter Thesauri über
eENVplus Thesaurus-Framework und Auswertungsdienste

Die Service-Module realisieren die Kernfunktionen des Systems. Sie implementieren Ausführungslogiken zur Umsetzung der TF-Nutzungsdienste. Konkret bedeutet das in

der Regel, dass sie Anfragen des Endanwenders in SPARQL-Queries umsetzen, die vom Virtuoso-Server im TF Backend beantwortet werden können. Die Service-Module sind die einzigen Berührungspunkte zwischen TF Backend und dem Rest der TF-Nutzungsschicht. Sie können modular ein- oder ausgeklinkt werden, um das System um gewissen Funktionalitäten zu ergänzen oder neue Implementierungen (effizienter, raffinierter, komplexer, domänenspezifisch, …) für bestehende Funktionalitäten einzubauen. Die Architektur stellt sicher, dass die verschiedenen Service-Module vollkommen unabhängig voneinander sind.

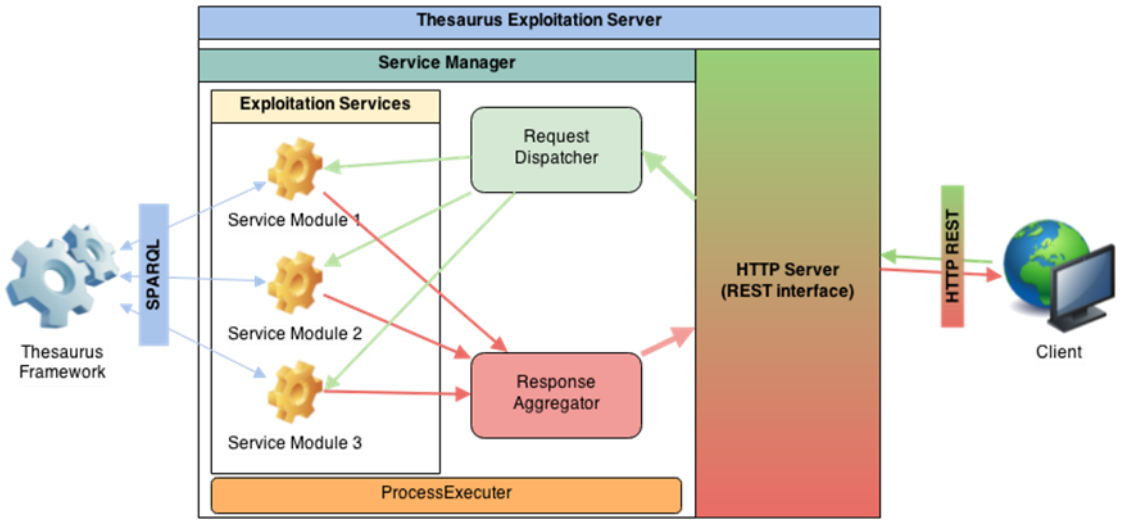


Abbildung 9: Interne Umsetzung des TF Exploitation Server

3.3 REST Schnittstelle und TFES Service-Module

Alle Service-Module sowie der Service-Manager implementieren eine *GetCapabilities* Operation, so dass jeder Client sich über die Fähigkeiten und Ressourcen des Systems informieren kann. Es gibt zwei Operationen (*DescribeConcept, ResolveThesaurus*), die durch genau ein Service-Modul ausgeführt werden sollen (s.w.u.). Für alle weiteren Operationen kann es mehrere Service-Module geben, die sie implementieren:

- Die Operation *GetSuggestions* schlägt passende Thesauruskonzepte vor sowie sie ein Schlüsselwort oder ein Teil davon als Eingabe erhält. Die Matching-Logik und die Kriterien zur Auswahl vorzuschlagender Begriffe hängen dabei vom konkreten Service-Modul ab.

- Die Operation *GetSynonyms* fragt nach allen Konzepten, die als synonym zum Eingabe-Schlüsselwort betrachtet werden. Die Gleichheitslogik und die exakten Kriterien zum Finden solcher gleicher Konzepte hängen von den aktiven Service-Modulen ab. Beispielsweise könnte man sich hier nur innerhalb eines einzigen Thesaurus bewegen, aber auch durch Cross-Walking verschiedene Thesauri über Inter-Thesaurus Links verknüpfen, unter Beachtung linguistisch synonymer Begriffe *(skos:altLabel)* oder auch nur der Normbegriffe *(skos:prefLabel)*.
- Die Operation *GetRelatives* sucht zu einem Schlüsselwort die verwandten Thesaurus-Konzepte. Auch hier hängen die Verwandtschaftslogik (Intra- oder Inter-Thesaurus-Betrachtung, Beachtung welcher Beziehungstypen, genutzte Tiefe/Entfernung von verwandten Konzepten, …) und die Auswahl zum Finden von Konzepten von den konkreten Service-Modulen ab.

Die folgende Tabelle zeigt die Liste der in Entwicklung befindlichen bzw. geplanten Service-Module mit ihren jeweiligen *Capabilities*.

Service-Module **Capabilities**	**Beschreibung der Ausführungslogik**
KeywordCompletion *GetSuggestions*	Zu einer gegebenen Zeichenfolge wird eine Liste von Thesaurus-konzepten geliefert, die mögliche Vervollständigungen darstellen. Auswahllogik: basiert auf *prefLabel* und *altLabel*.
KeywordExplanation *DescribeConcept*	Zu einem gegebenen Konzept-URI wird der entsprechende vollständige SKOS-Datensatz zurückgegeben.
KeywordTranslation *GetSynonyms*	Zu einem gegebenen Konzept-URI werden alle alternativen Label in einer beliebigen natürlichen Sprache im TF geliefert, ggf. auch unter Nutzung von *exactMatch* Links. Dies umfasst auch Cross-walking über Thesaurusgrenzen hinweg.
ThesaurusResolution *ResolveThesaurus*	Zu einem gegebenen Konzept-URI wird der Thesaurus zurückgegeben, aus dem das Konzept stammt (erforderlich für Anforderung 16 der „INSPIRE Metadata Implementing Rules - Technical Guidelines“).
AnnotationNormalization *GetSuggestions*	Zu einem gegebenen Konzept-URI warden die “am besten passenden” Normbegriffe geliefert, basierend auf Textähnlichkeit zwischen der Eingabe und den Alternativbegriffen von im TF enthaltenen Konzepten.
SemanticExplorative Search	Zu einem gegebenen Konzept-URI werden alle verwandten Begriffe und erforderlichen Zusatzinformationen geliefert, um die

GetRelatives *GetSynonyms* *DescribeConcept*	Visualisierung der semantischen explorativen Suche zu realisieren.

Hier ist zu beachten: (1) Die Operationen der Service-Module unterstützen in der Regel eine Reihe von Parametern, die wir hier aus Platzgründen auslassen (z.B. maximale Größe der Antwortmenge, Einschränkung auf bestimmte Thesauri oder Sprachen, Ein-/Ausschaltend es Cross-Walking, …). (2) In Anfragen und in der Ergebnismenge werden Konzepte normalerweise durch sogenannte *KeywordObjects* dargestellt, also nicht nur durch lexikalische Terme (Strings), sondern Objekte, die – falls verfügbar – auch den URI enthalten.

3.4 Anwendungsfälle für TF-Nutzungsdienste

Grob lassen sich die folgenden drei Anwendungsfälle unterscheiden.

(1) Metadaten-Erzeugung. Wenn ein Anwender Metadaten erzeugt, aktualisiert oder zusammenführt, könnten verschiedene TF-Nutzungsdienste zur Verwendung kommen: **KeywordCompletion** (Vervollständigung von Schlüsselworten) kann offensichtlich immer Zeit sparen und Fehlermöglichkeiten reduzieren. Falls mehrere Sachbearbeiter Metadaten eingeben, kann sie außerdem die Homogenität zwischen den Eingaben der unterschiedlichen Personen erhöhen. Die **KeywordExplanation** und **KeywordTranslation** (Erklärung und Übersetzung von Schlüsselworten) erleichtern das Verständnis und die korrekte Verwendung der angebotenen Schlüsselworte – insbesondere wenn der Anwender nicht in seiner Muttersprache oder nicht auf seinem angestammten Fachgebiet arbeitet. Die Betrachtung allgemeinerer, speziellerer oder verwandter Begriffe, auch aus unterschiedlichen Fachthesauri (geliefert von den Diensten **KeywordExplanation** und **QueryReformulation**) kann hier weitere Hilfestellung bieten. Ein Modul zur **KeywordValidation** könnte sogar (halb-)automatisch eine gewisse „Normalform" für eine gegebene Menge von Metadateneinträgen erzeugen, die z.B. nur bevorzugte Schlüsselworte (Normbegriffe) enthalten. Eine solche Validierungsprozedur könnte bspw. auch im Batch-Betrieb auf eine Datenbasis existierenden Metadaten angewandt werden.

(2) Metadaten-Suche. Abbildung 10 illustriert, wie die TF-Nutzungsdienste die Suche nach Metadaten unterstützen können: Angenommen, ein Anwender beginnt ein Schlüsselwort in der Suchmaske seiner CSW einzugeben, könnte das eENVplus TF zunächst mögliche Vervollständigungen vorschlagen (**KeywordCompletion**) (1). Darunter könnten sogar welche enthalten sein, die gar nicht lexikalisch passen, aber konzeptionell (der Anwender beginnt mit dem Tippen des Wortes „Eis“ und erhält als Vorschlag auch das Wort „Gletscher“, weil das TF ja auf Synonyme, Übersetzungen und verwandte Begriffe zugreifen kann).

Mithilfe der **KeywordExplanation** kann auch weiteres Hintergrundwissen zu den Begriffen angeboten werden, z.B. die Definitionsfelder aus dem Thesaurus als Tooltip (2). Hat sich der Anwender für einen Suchbegriff entschieden (3) und sendet die Query ab, kann die **QueryReformulation** eine erweiterte Liste von Suchbegriffen liefern, die z.B. Synonyme, Übersetzungen und verwandte Begriffe enthält, die in die Suchanfrage eingebaut werden können (4), was schließlich zu einer vollständigeren Ergebnismenge führt (5).

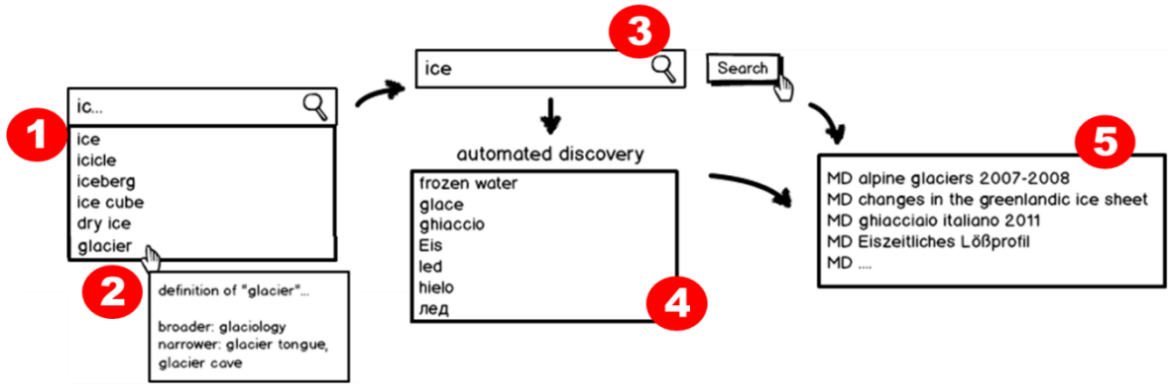


Abbildung 10: Mock-Up zum Anwendungsfall Metadaten-Suche

(3) Semantische Explorative Suche. Der dritte Anwendungsfall kann sowohl mit der Metadaten-Eingabe als auch mit der Suche verknüpft werden. Das Modul zur sogenannten **Semantischen Explorativen Suche** ermittelt die „konzeptuelle Umgebung” eines gegebenen Begriffs, um daraus eine interaktive, navigierbare Visualisierung zu erzeugen, mit deren Hilfe ein in der gegebenen Themenwelt noch unerfahrener Anwender einen Begriff mit all seinen Facetten und Zusammenhängen im Kontext besser

verstehen kann. Abbildung 11 illustriert die Idee in einem Mock-Up. Hier wird das Schlüsselwort „Soil" mit seinen *Broader Terms* (grau) und *Narrower Terms* (gelb) aus dem GEMET Thesaurus gezeigt, außerdem können per Klick Definition und Übersetzungen des Begriffs angezeigt werden. Verwandte Begriffe *innerhalb* des GEMET können über die Schaltfläche „Rel" angesteuert werden, während „Match" die *Inter-Thesaurus Verbindungen* anzeigt (*skos:exactMatch),* im Beispiel zum entsprechenden Konzept im Thesauru EarTH. Hier sehen wir komplett andere Ober- und Unterbegriffe (grau, gelb) als in GEMET, weil EarTH eine unterschiedliche Sichtweise auf das Themenfeld repräsentiert.

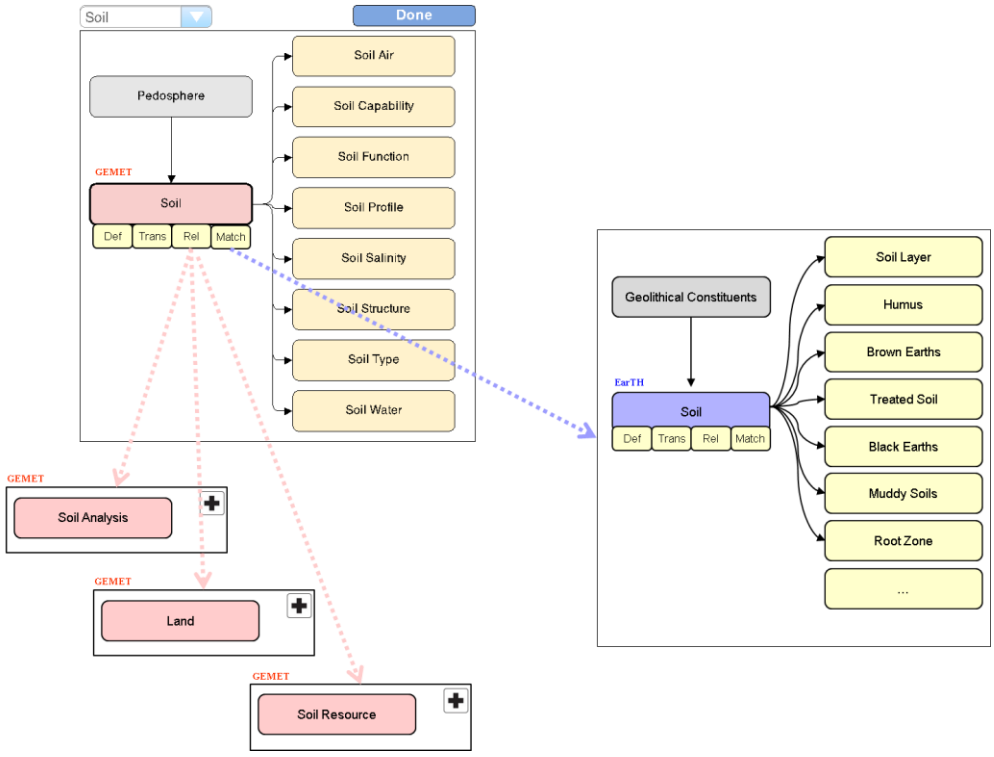


Abbildung 11: Mock-Up zur Semantischen Explorativen Suche

4 Zusammenfassung

Wir haben den Stand der Technik beschrieben sowie einen Ausblick auf mögliche zukünftige Entwicklungen im Metadatenmanagement in Geodateninfrastrukturen gegeben. Zum Stand der Technik wurde das kommerzielle Werkzeug Preludio vorgestellt, das großen Wert auf Konfigurierbarkeit, Benutzerfreundlichkeit, Qualitätssicherung und Standard-Konformität im Metadatenmanagement legt. Mit Blick auf zukünftige Entwicklungen wurde ein Zwischenstand der Arbeiten im EU-Projekt

eENVplus gegeben. Hier wird ein Multi-Thesaurus-Framework entwickelt, mit dem verschiedene Umweltthesauri und Fachvokabulare als Linked Data veröffentlicht und miteinander verknüpft werden können. Die Thesaurus-Nutzungsdienste greifen dann auf die vernetzten Ressourcen wie auf einen einzigen mehrsprachigen, domänen- und persepktivenübergreifenden Fachthesaurus zu. Die Nutzungsdienste abstrahieren vom Backend und können in verschiedene existierende Anwendungen eingebunden werden, wie Metadateneditoren, Verzeichnisdienste oder Geoinformationsportale. So können bspw. auch die Suchverfahren von Preludio durch Zugriff auf die TF-Nutzungsdienste verbessert werden.

5 Literatur

[1] Albertoni R., De Martino M., Di Franco S., De Santis V., Plini P., "EARTh: An Environmental Application Reference Thesaurus in the Linked Open Data Cloud," *Semantic Web*, 2014.

[2] Albertoni R., Gómez-Pérez A., "Assessing Linkset Quality for Complementing Third-Party Datasets," EDBT/ICDT Workshops 2013, S. 52-59.

[3] Berners-Lee T., "Design Issues: Linked Data," URL: http://www.w3.org/DesignIssues/LinkedData.html, 2006.

[4] Heath T., Bizer C., "Linked Data: Evolving the Web into a Global Data Space (1st edition)," Synthesis Lectures on the Semantic Web: Theory and Technology, S. 1-136, Morgan & Claypool, 2011.

[5] De Martino M., Albertoni R., "A Multilingual/Multicultural Semantic-based Approach to Improve Data Sharing in a SDI for Nature Conservation," *International Journal of Spatial Data Infrastructures Research,* vol.6, ISSN 1725-0463, S. 206-233, 2011.

[6] Abecker A., Heidmann C., Kazakos W., Nedkov R., Wössner R., „The Preludio Metadata Management System for Comprehensive Support of INSPIRE Compliant Metadata," in: INSPIRE Conference 2014, Aalborg, Dänemark.

[7] Abecker A., Albertoni R., Cipolloni C., de Martino M., Moradiafkan Y., Wössner R., “Software Services Exploiting the eENVplus Framework of Interlinked Thesauri for Metadata Management,” in: INSPIRE Conference 2014, Aalborg, Dänemark.

[8] Grimm S., Abecker A., Völker J., Studer R., “Ontologies and the Semantic Web,“ in: J. Domingue, D. Fensel und J.A. Hendler (Hrsg.): *Handbook of Semantic Web Technologies*, Berlin Heidelberg: Springer-Verlag, 2011, S. 507-580.

Danksagung: Die vorgestellten Ergebnisse wurden teilweise mit finanzieller Unterstützung der EU im Projekt eENVplus (http://www.eenvplus.eu/, FKZ 325232) im Rahmen des Programms CIP-ICT-PSP realisiert.

Kartenservice Umgebungslärm Schleswig-Holstein

Friedhelm Hosenfeld
Institut für Digitale Systemanalyse & Landschaftsdiagnose (DigSyLand),
hosenfeld@digsyland.de

Ludger Gliesmann,
Landesamt für Landwirtschaft, Umwelt und ländliche Räume
des Landes Schleswig-Holstein (LLUR-SH)
Ludger.Gliesmann@llur.landsh.de

Andreas Rinker,
Institut für Digitale Systemanalyse & Landschaftsdiagnose (DigSyLand),
rinker@digsyland.de

Abstract

This contribution presents the "mapping service environmental noise Schleswig-Holstein": The main component of this web platform consists of a WebGIS with strategic noise maps ("Noise Atlas"). The system was developed in order to support the demands for the first period of the EC Environmental Noise Directive in 2007. The Noise Atlas was since then enhanced continuously, the additions include among others download options for action plans and integration of evaluation platform Disy Cadenza.

Kurzzusammenfassung

Vorgestellt wird der Kartenservice Umgebungslärm Schleswig-Holstein: Eine Web-Plattform, deren Hauptkomponente die mit einem WebGIS realisierten Lärmkarten bilden („Lärm-Atlas“). Der zur 1. Stufe der EG-Umgebungslärmrichtlinie aufgebaute Lärm-Atlas wurde seitdem um Zusatz-Module zur Präsentation von Gemeinde-bezogenen Lärm-Daten, zur Datenpflege und Bereitstellung von Lärm-Aktionsplänen sowie um eine Anbindung an die Auswertungsplattform Disy Cadenza ergänzt.

1 Überblick

Mit der Richtlinie 2002/49/EG des Europäischen Parlaments und des Rates vom 25. Juni 2002 über die Bewertung und Bekämpfung von Umgebungslärm („Umgebungslärmrichtlinie“) hat die Europäische Union ein Konzept vorgegeben, Lärmauswirkungen zu erfassen und ihnen entgegen zu wirken. Die wesentlichen Ziele sind:

- die Ermittlung der Belastung durch strategische Lärmkarten,
- Bewertung und soweit erforderlich Vermeidung oder Verminderung von Belastungen durch Aktionspläne.

Mit der ersten Stufe der Umgebungslärmrichtlinie musste 2007 erstmals die Ausarbeitung und Bereitstellung der Lärmkarten stattfinden, 2012 wurden aktualisierte und erweiterte Daten für die zweite Stufe bereitgestellt.

Generell ist die Überprüfung und bei Bedarf die Überarbeitung von Lärmkarten und Lärmaktionsplänen mindestens alle 5 Jahre vorgesehen.

Die Information der Öffentlichkeit über vorhandene Lärmbelastungen und die Mitwirkung der Öffentlichkeit bei der Aktionsplanung sind dabei von zentraler Bedeutung. Das Landesamt für Landwirtschaft, Umwelt und ländliche Räume (LLUR) ist in Schleswig-Holstein verantwortlich für die Berichterstattung zur Umsetzung der Umgebungslärmrichtlinie gegenüber Bund und EU und unterstützt die Gemeinden bei der Ausarbeitung der Lärmkarten und der Aufstellung der Aktionspläne.

Daher wurde ein Werkzeug benötigt, das das Landesamt bei der Erfüllung dieser Aufgaben geeignet unterstützt. Dies führte zur Entwicklung des in diesem Beitrag beschriebenen Werkzeugs. Der gesetzliche Auftrag zur Veröffentlichung der Lärmkarten und Lärmaktionspläne wird mit dem vorgestellten Lärm-Atlas erfüllt.

2 Zeitliche Entwicklung

2.1 Entwicklungen für die erste Stufe der Umgebungslärmrichtlinie

Bereits 2007 wurde in Schleswig-Holstein erkannt, dass die Anforderungen zur Informationen und Mitwirkung der Öffentlichkeit am geeignetsten mit Hilfe einer allgemein verfügbaren WebGIS-Anwendung umzusetzen sind.

So wurde 2007 die Firma DigSyLand beauftragt, den „Lärm-Atlas“ als spezielle Variante des bereits erfolgreich eingesetzten Landwirtschafts- und Umweltatlas Schleswig-Holstein [Görtzen et al. 2004, 2009, Rammert & Hosenfeld 2003] zu entwickeln. Neben der Präsentation der Lärmkarten (s. Kap. 4.1) konnten im Lärm-Atlas auch Gemeindeinformationen und detaillierte Zahlen der in den jeweiligen Gemeinden von Lärm Betroffenen abgerufen werden (s. Kap. 4.3).

Zusätzlich wurde jeder betroffenen Gemeinde eine Zugangskennung bereitgestellt, mit der die Gemeinde eigene Informationen im Lärm-Atlas bereitstellen kann und Zugang zu Download-Paketen mit Lärmmodelldaten zu ihrer Gemeinde bekommt.

2.2 Erweiterungen für die zweite Stufe

Für die zweite Stufe der Umgebungslärmrichtlinie wurden ab 2011 weitere Funktionen im Lärm-Atlas ergänzt. Dazu zählten Prüf- und Korrekturfunktion der Eingangsdaten für die Modellierung der Lärmbelastung, die die Gemeinden ebenfalls mit Hilfe des WebGIS nutzen konnten (s. Kap. 4.2). Weitere Ergänzungen betrafen die Hochlade- und Bereitstellungsmöglichkeiten der Lärmaktionspläne durch die Gemeinden (s. Kap. 4.4).

Bei ca. 500 betroffen Gemeinden in SH war es für die Sachbearbeitung erforderlich alle Gemeinde-bezogenen Adress- und Kontaktdaten für Berichtspflichten (Serienbriefe), Infobriefe und die Betreuung der Gemeinden durch das Landesamt konsequent in die zentrale Datenbank zu verlagern.

Die im Umweltbereich weitverbreitete Auswertungsplattform Disy Cadenza wurde als Auswertungswekzeug eingebunden, um die Adressdaten z.B. zur Zusammenstellung von Verteilerlisten zu nutzen (s. Kap. 4.5 und 4.6). Insbesondere konnte die für die Berichterstattung an die EU aufwendige manuelle Erfassung der von der EU geforderten

Kenndaten mit Einscannen der Lärmaktionspläne auf ein automatisiertes Verfahren (Hochladen der PDF-Datieien und Eintragen der Kenndaten durch die Gemeinden) mit der Berichterstellung aus Cadenza ersetzt werden.

Ergänzt wird der Kartenservice Umgebungslärm durch ein umfangreiches Informationsangebot im zentralen Internet-Portal der Landesregierung.

3 Technische Grundlagen

Die Web-Anwendung zum Lärm-Atlas wurde mit PHP entwickelt, wobei die WebGIS-Funktionalitäten auf der Basis des UMN Mapservers umgesetzt wurden. Die Datenhaltung findet im Oracle Datenbanksystem (derzeit Oracle 11.2) statt. Die Eingangsdaten der Lärm-Modellierung werden als Geometrie-Informationen ebenfalls in Oracle gehalten, während die Lärmbelastungskarten im Shape-Dateiformat eingebunden werden. Zu den Eingangsdaten zählen z.B. Angaben zu Geschwindigkeitsbegrenzungen, Straßenoberflächen, Gebäudenutzungen und Lärmschutzwände.

Für die Auswertung der Daten kommt Disy Cadenza Professional zum Einsatz, das direkt auf die Produktivdatenbank zugreift.

4 Umsetzung einzelner Module

4.1 Darstellung der Lärmkarten

Im Lärm-Atlas wird die Darstellung der von der EU vorgegebenen Lärmkarten (z.B. Straßenlärm 24h oder Straßenlärm nachts) zur Auswahl angeboten, wobei die Ergebnisse der 1. Stufe 2007 mit denen der 2. Stufe 2012 verglichen werden können. Über Standard-WebGIS-Funktionen wie beliebiges Zoomen, Maßstabseingabe und eine Druckfunktion (siehe Abb. 1) können die modellierten Lärmbelastungen nach Bedarf dargestellt werden. Mit Hilfe einer Gemeinde-Suchfunktion kann gezielt die Lärmbelastung einer Gemeinde visualisiert werden.

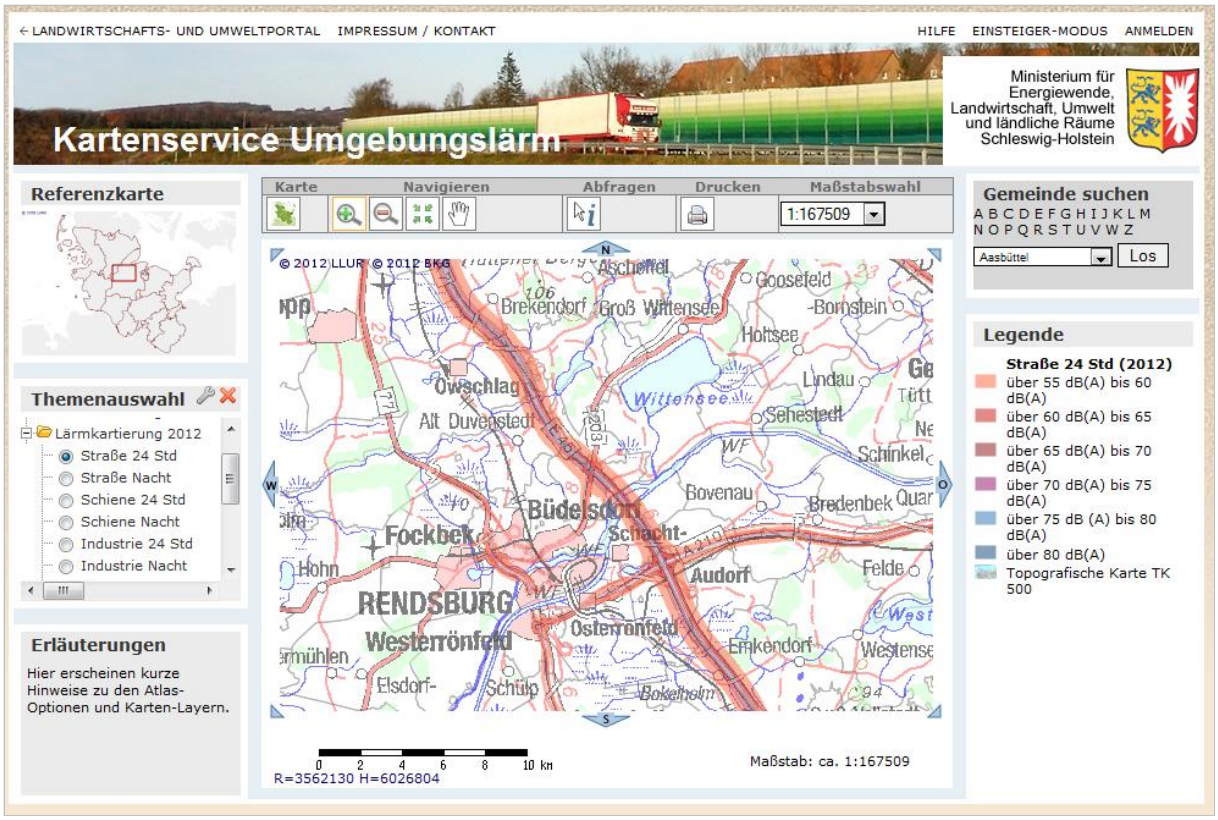


Abbildung 1: Präsentation der Lärmbelastung im Lärm-Atlas

4.2 Karten der Eingangsdaten mit Korrekturmöglichkeiten

Mit der 2. Stufe der Lärmrichtlinie wurde den Gemeinden die Möglichkeit gegeben, im Lärm-Atlas die Eingangsdaten, die die Grundlagen der Modellierung der Lärmbelastung bilden, auf ihrem Gemeindegebiet zu überprüfen und direkt online zu korrigieren. Nach dem Anwählen eines Eingangsdaten-Themas wie z.B. Geschwindigkeit, Verkehrsstärke oder Gebäudenutzung, konnte per Klick auf den entsprechenden Straßenabschnitt ein Eingabefenster aufgerufen werden, in dem die Eingangsdaten korrigiert und mit Bemerkungen versehen werden konnten (siehe Abb. 2).

Die Eingangsdaten werden mit ihren Geometrien in der Oracle-Datenbank verwaltet, so dass nach dem Abspeichern der korrigierten Daten die Ergebnisse sofort auf der digitalen Karte sichtbar sind.

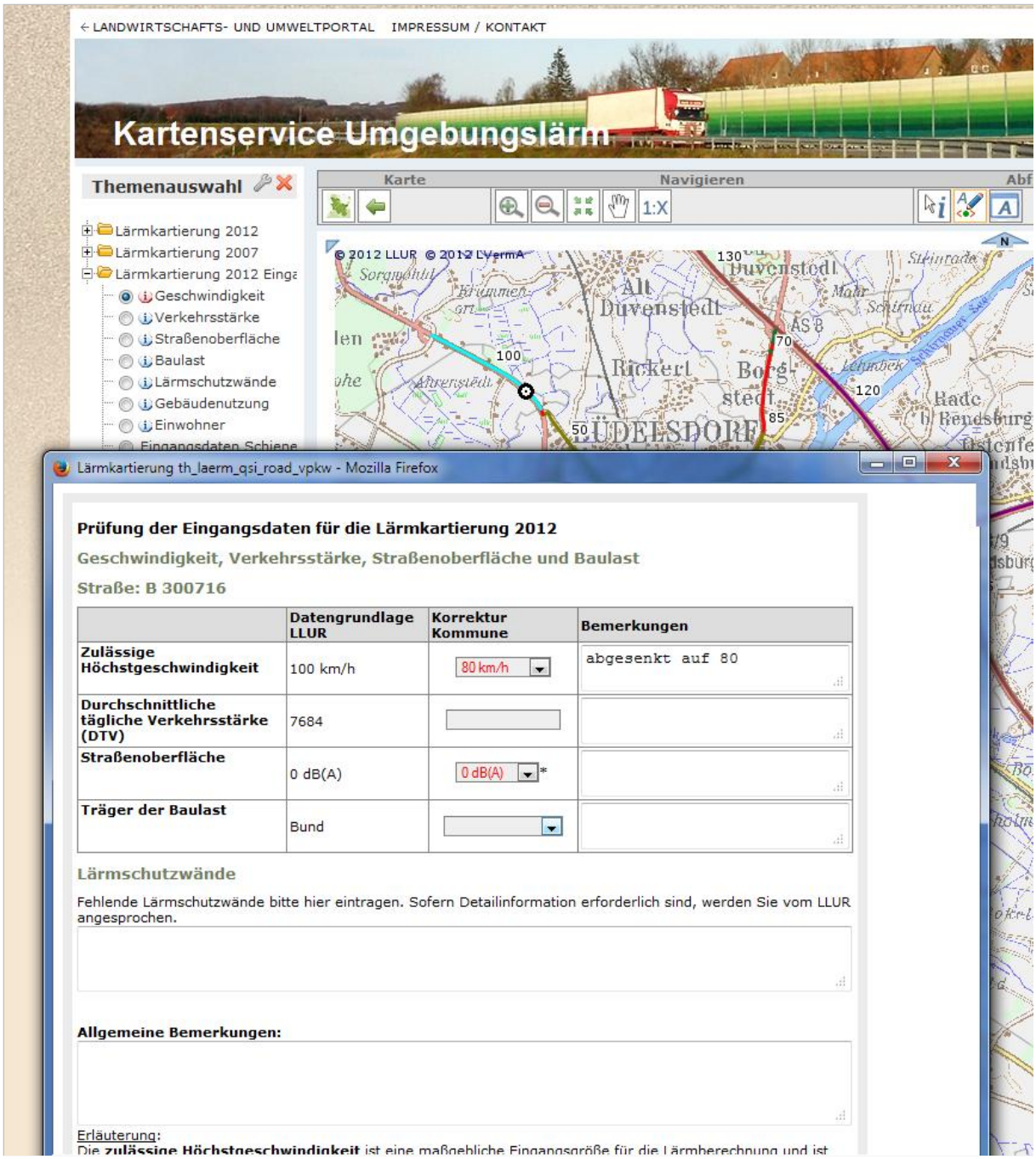


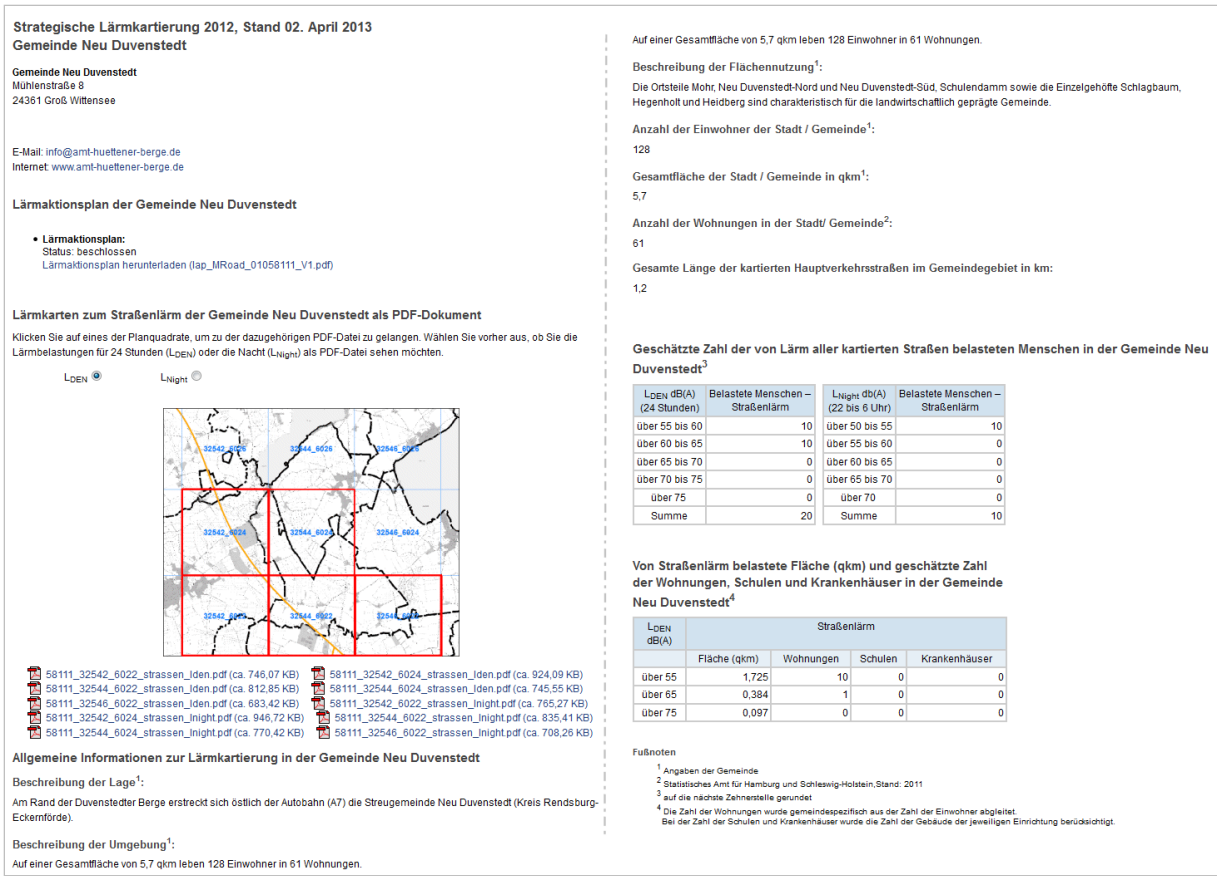
Abbildung 2: Bearbeitung der Eingangsdaten für die Lärmmodellierung (Muster)

4.3 Darstellung von Gemeindeinformationen und Belastungszahlen

Ergänzend zu den Lärmkarten können zu jeder Gemeinde, die von der Lärmrichtlinie betroffen ist, weitere Informationen abgerufen werden, die gesetzliche Bestandteile der Lärmkarten sind. Dazu zählen grundlegende Gemeinde-Angaben aber auch Tabellen mit den Belasteten-Zahlen (siehe Abb. 3).

Zusätzlich werden die von dem Büro Lärmkontor GmbH, Hamburg erstellten Lärmkarten als PDF-Dateien speziell für das Gemeindegebiet zum Download auf der Basis von DTK5-Blattschnitten (durch clickable maps) angeboten.

Alle benötigten Angaben werden in der zentralen Oracle-Datenbank verwaltet, so dass Aktualisierungen unkompliziert und zeitnah vorgenommen werden können. Die Gemeinden können über ihre Zugangskennung ihre Stammdaten kontrollieren, bearbeiten und zur Veröffentlichung freigeben.



Strategische Lärmkartierung 2012, Stand 02. April 2013
Gemeinde Neu Duvenstedt

Gemeinde Neu Duvenstedt
Mühlenstraße 8
24361 Groß Wittensee

E-Mail: info@amt-huettener-berge.de
Internet: www.amt-huettener-berge.de

Lärmaktionsplan der Gemeinde Neu Duvenstedt

- **Lärmaktionsplan:**
  Status: beschlossen
  Lärmaktionsplan herunterladen (lap_MRoad_01058111_V1.pdf)

Lärmkarten zum Straßenlärm der Gemeinde Neu Duvenstedt als PDF-Dokument

Klicken Sie auf eines der Planquadrate, um zu der dazugehörigen PDF-Datei zu gelangen. Wählen Sie vorher aus, ob Sie die Lärmbelastungen für 24 Stunden (L_{DEN}) oder die Nacht (L_{Night}) als PDF-Datei sehen möchten.

L_{DEN} L_{Night}

58111_32542_6022_strassen_lden.pdf (ca. 746,07 KB)
58111_32544_6022_strassen_lden.pdf (ca. 812,85 KB)
58111_32546_6022_strassen_lden.pdf (ca. 683,42 KB)
58111_32542_6024_strassen_lnight.pdf (ca. 946,72 KB)
58111_32544_6024_strassen_lnight.pdf (ca. 770,42 KB)
58111_32542_6024_strassen_lden.pdf (ca. 924,09 KB)
58111_32544_6024_strassen_lden.pdf (ca. 745,55 KB)
58111_32542_6022_strassen_lnight.pdf (ca. 765,27 KB)
58111_32544_6022_strassen_lnight.pdf (ca. 835,41 KB)
58111_32546_6022_strassen_lnight.pdf (ca. 708,26 KB)

Allgemeine Informationen zur Lärmkartierung in der Gemeinde Neu Duvenstedt

Beschreibung der Lage[1]:
Am Rand der Duvenstedter Berge erstreckt sich östlich der Autobahn (A7) die Streugemeinde Neu Duvenstedt (Kreis Rendsburg-Eckernförde).

Beschreibung der Umgebung[1]:
Auf einer Gesamtfläche von 5,7 qkm leben 128 Einwohner in 61 Wohnungen.

Auf einer Gesamtfläche von 5,7 qkm leben 128 Einwohner in 61 Wohnungen.

Beschreibung der Flächennutzung[1]:
Die Ortsteile Mohr, Neu Duvenstedt-Nord und Neu Duvenstedt-Süd, Schulendamm sowie die Einzelgehöfte Schlagbaum, Hegenholt und Heidberg sind charakteristisch für die landwirtschaftlich geprägte Gemeinde.

Anzahl der Einwohner der Stadt / Gemeinde[1]:
128

Gesamtfläche der Stadt / Gemeinde in qkm[1]:
5,7

Anzahl der Wohnungen in der Stadt/ Gemeinde[2]:
61

Gesamte Länge der kartierten Hauptverkehrsstraßen im Gemeindegebiet in km:
1,2

Geschätzte Zahl der von Lärm aller kartierten Straßen belasteten Menschen in der Gemeinde Neu Duvenstedt[3]

L_{DEN} dB(A) (24 Stunden)	Belastete Menschen – Straßenlärm	L_{Night} db(A) (22 bis 6 Uhr)	Belastete Menschen – Straßenlärm
über 55 bis 60	10	über 50 bis 55	10
über 60 bis 65	10	über 55 bis 60	0
über 65 bis 70	0	über 60 bis 65	0
über 70 bis 75	0	über 65 bis 70	0
über 75	0	über 70	0
Summe	20	Summe	10

Von Straßenlärm belastete Fläche (qkm) und geschätzte Zahl der Wohnungen, Schulen und Krankenhäuser in der Gemeinde Neu Duvenstedt[4]

L_{DEN} dB(A)	Straßenlärm			
	Fläche (qkm)	Wohnungen	Schulen	Krankenhäuser
über 55	1,725	10	0	0
über 65	0,384	1	0	0
über 75	0,097	0	0	0

Fußnoten
[1] Angaben der Gemeinde
[2] Statistisches Amt für Hamburg und Schleswig-Holstein,Stand: 2011
[3] auf die nächste Zehnerstelle gerundet
[4] Die Zahl der Wohnungen wurde gemeindespezifisch aus der Zahl der Einwohner abgeleitet.
Bei der Zahl der Schulen und Krankenhäuser wurde die Zahl der Gebäude der jeweiligen Einrichtung berücksichtigt.

Abbildung 3: Gemeindeinformationen, Belastetetenzahlen und PDF-Download
(Ausschnitte aus Platzgründen nebeneinander montiert)

4.4 Dokumentation, Upload und Bereitstellung der Lärmaktionspläne

Auf der Basis der Lärmbelastungszahlen stellen die Gemeinden Lärmaktionspläne auf, um geplante Maßnahmen zur Lärmminderung zu dokumentieren. Im Lärm-Atlas werden den Gemeinden Funktionen angeboten, die von ihnen erstellten Lärmaktionspläne in diese Plattform hochzuladen, zu kommentieren und sie wiederum zum Download auf der Gemeindeinformationsseite im Lärm-Atlas bereitzustellen (siehe Abb. 3). Dies ist

der per Erlass eingeführte verbindliche Weg zur Bericherstattung an die EU. Die Dateien werden auf dem Server datenbankgestützt verwaltet, versioniert und mit einem Zugriffsschutz versehen, so dass nur freigegebene Lärmaktionspläne allgemein heruntergeladen werden können.

4.5 Pflege von Adress-Daten in einem Administrationsmodul

Um die Pflege aller Gemeinde-bezogenen Daten durch das LLUR zu erleichtern, wurde der Lärm-Atlas 2013 um ein Modul zur Bearbeitung aller für die Umgebungslärmrichtlinie relevanten Angaben zu Gemeinden, Ämter und Kreisen ergänzt. Für jede Stufe der Umgebungslärmrichtlinie können beispielsweise Einwohnerzahlen, Streckenlängen von Bahn und Straße, beauftragte Lärm-Kartier-Büros und ähnliches zentral verwaltet werden.

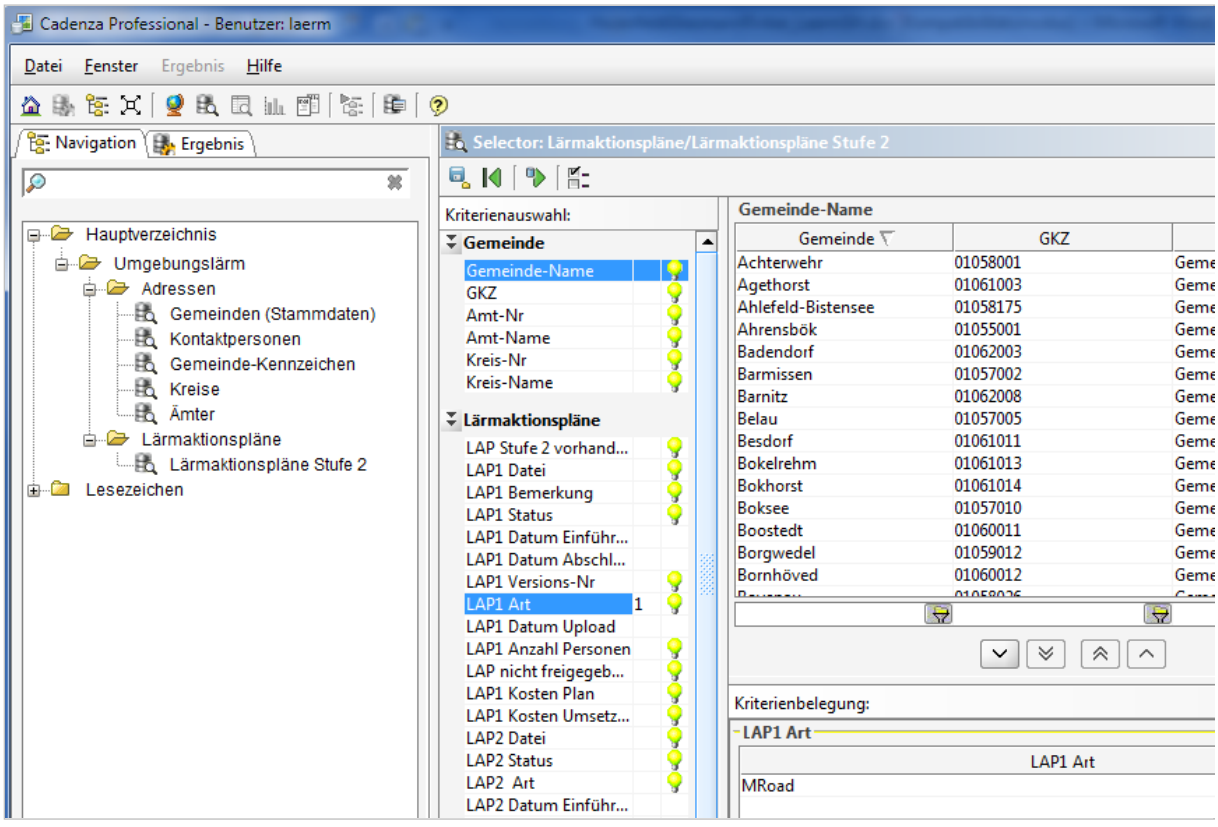


Abbildung 4: Auswertung der Umgebungslärmdaten mit Cadenza Professional

4.6 Auswertung der Lärmaktionspläne und Adress-Daten mit Disy Cadenza

Zur flexiblen Auswertung aller in den vorgenannten Abschnitten genannten Daten der Lärmrichtlinie und zur Berichterstattung kommt die Auswertungsplattform Disy Cadenza zum Einsatz. Dazu wird die im LLUR bereits für andere Abteilungen eingerich-

tete Cadenza-Installation genutzt, aber direkt auf die Internet-Datenbank zugegriffen, so dass auf eine Daten-Replikation in das LLUR-Intranet verzichtet werden kann.

Mit Cadenza Professional können Adressverteiler nach allen relevanten Merkmalen zusammengestellt werden, dazu gehören beispielsweise Infobriefe des LLUR an die Gemeinden mit aktuellen Informationen zur Umsetzung der Lärmrichtlinie.

Außerdem können die durch die Gemeinden gelieferten Lärmaktionspläne ausgewertet und kontrolliert werden. Per Link auf den Lärm-Atlas können die Lärmaktionspläne als PDF direkt aus Cadenza heraus geöffnet werden.

5 Zusammenfassung und Ausblick

Seit der ersten Realisierung 2007 für die Umsetzung der 1. Stufe der EG-Umgebungslärmrichtlinie in Schleswig-Holstein wurde der Kartenservice Schritt für Schritt erweitert und an aktuelle Anforderungen angepasst. Während der Lärm-Atlas bereits für die 1. Stufe ein wichtiges Werkzeug zur Unterstützung der Gemeinden und Landesbehörden bildete, konnte die Applikation für die 2. Stufe der Richtlinie ab 2012 noch einmal optimiert und ausgebaut werden. So werden derzeit alle relevanten Daten konsequent in der zentralen Oracle-Datenbank verwaltet, während in der Vergangenheit noch zusätzlich Microsoft Access und Excel zur Datenhaltung dienten. So kann einerseits den Gemeinden ein allgemeiner dezentraler Web-Zugang zur Kontrolle ihrer Daten ermöglicht werden, andererseits können durch das LLUR übergreifende Auswertungen mit Disy Cadenza durchgeführt werden.

Die Ergebnisse der Lärmmodellierungen für die 2. Stufe der Lärmrichtlichtlinie wurden von den Gemeinden sehr viel besser akzeptiert. Durch die aktive Einbeziehung der Gemeinden bei der Zusammenstellung der Eingangsdaten musste das LLUR deutlich weniger Anfragen bearbeiten als noch für die 1. Stufe. Dadurch konnten nicht zuletzt Kosten für Koordinierungs- und Datenaufbereitungstätigkeiten eingespart werden.

Gesetzlichen Aufgaben, wie den Pflichten zur Veröffentlichung und zur Berichterstattung kann automatisiert und sehr effizeint nachgekommen werden.

Zukünftig sollen Alt-Daten der 1. Stufe sowie die Zahlen der von Umgebungslärm Betroffenen ebenfalls mit Cadenza erschlossen werden, um das Auswertungspotenzal noch besser auszunutzen.

Einige der Lärmkarten sind prototypisch bereits als WMS-Dienst realisiert worden, so dass künftig die Einbindung der Geodaten seitens der Gemeinden und anderer Interessierte auch in andere GIS-Systeme möglich sein wird.

6 Literaturverzeichnis

[Görtzen et al. 2004]

Görtzen, D.; Schneberger, S. & Rammert, U. (2004): Schleswig-Holsteins Environmental Atlas for the Public and for Special Interest Groups. In: Proceedings of the 18th Conference Informatics for Environmental Protection. October 21-23, 2004. CERN, Geneva, pp. 634-640.

[Görtzen et al. 2009]

Görtzen, D.; Rammert, U. & Bornhöft, U. (2009): Geo Data and Infrastructure provided by the Environmental Administration of Schleswig-Holstein. In: Wohlgemuth, V.; Page, B. & Voigt, K. (eds.): Proceedings of the 23rd on Environmental Informatics and Industrial Environmental Protection: Concepts, Methods and Tools. Berlin, 2009, pp. 369-376.

[Rammert & Hosenfeld 2003]

Rammert, U. & Hosenfeld, F. (2003): Dynamic and Interactive Presentation of Environmental Information. In: Gnauck, A. & Heinrich, R. (eds.): The Information Society and Enlargement of the European Union, 17th International Symposium Informatics for Environmental Protection, Cottbus 2003, pp. 517-524.

Links:

Umgebungslärmrichtlinie in Schleswig-Holstein:

http://www.laerm.schleswig-holstein.de/

Kartenservice Umgebungslärm:

http://www.umweltdaten.landsh.de/laermatlas/

Erfahrungen mit MongoDB bei der Verwaltung meteorologischer Massendaten

Richard Lutz, Karlsruher Institut für Technologie (KIT), Institut für Angewandte Informatik (IAI), Richard.Lutz@kit.edu

Abstract

The remote sensing of atmospheric trace gases investigates dynamic, microphysical and chemical processes in the Earth's atmosphere, with the goal to understand, quantify and predict its natural variability and long-term changes. Accurate measurements of atmospheric trace gases from various observational platforms (ground-based stations, air craft, balloons, satellites) provide the data that are required for the modelling of atmospheric processes. The instrument GLORIA (Gimballed Limb Observer for Radiance Imaging of the Atmosphere), developed by KIT/IMK and FZ Jülich, an Infrared Spectrometer, which measures atmospheric emissions, was engaged in several measurement campaigns on board of HALO (High Altitude and Long Range Research Aircraft) and provided a large amount of data, which has to be managed efficiently for processing and visualisation. This paper describes the system background and the use of MongoDB for the provision of measured and processed mass data.

Einleitung

Die Datenfernerkundung mit Flugzeugen, Ballons und Satelliten ist seit vielen Jahren ein wichtiger Bestandteil der Atmosphärenforschung. Ziel der Datenauswertung ist die Untersuchung dynamischer, chemisch-physikalischer Prozesse in der Stratosphäre und oberen Troposphäre. Hierzu werden genaue Messungen atmosphärischer Spurengase durch unterschiedliche Beobachtungsplattformen wie Bodenstationen, Ballons,

Flugzeuge und Satelliten durchgeführt. Das Instrument GLORIA (Gimballed Limb Observer for Radiance Imaging of the Atmosphere), welches vom Karlsruher Institut für Technologie (KIT) und dem Forschungszentrum Jülich gemeinsam entwickelt wird, ist ein Infrarotspektrometer zur Messung atmosphärischer Emissionen. GLORIA wurde bereits bei mehreren Flugzeugkampagnen eingesetzt und lieferte dabei eine große Menge von Messdaten, welche für die Auswertung und Visualisierung effizient verwaltet werden müssen. Der vorliegende Beitrag beschreibt das Systemumfeld und die Verwendung der NoSQL-Datenbank MongoDB zur Speicherung der Mess-, Prozessierungs- und Ergebnisdaten

1 Meteorologische Massendaten – das Instrument GLORIA

Das Infrarotspektrometer GLORIA wird für den Einsatz auf Flugzeugen und Ballons entwickelt [GLORIA, 2014; GLORIA2, 2014]. Es wurde bereits mehrere Male auf Kampagnen des Forschungsflugzeugs HALO (High Altitude and Long Range Research Aircraft [HALO, 2014]) eingesetzt und lieferte bisher Daten in der Größenordnung von über 20 TB.

Für diese ersten Kampagnen wird zurzeit ein Verwaltungs- und Berechnungssystem entwickelt, welches die Daten zukünftiger Kampagnen ebenfalls aufnehmen soll.

1.1 Systemumfeld der GLORIA-Datenauswertung

Die GLORIA-Messdaten werden in der Large Data Facility (LSDF) des Steinbuch Centre for Computing (SCC, KIT) gespeichert. Zur effizienten Weiterverarbeitung der Daten durch das KIT-Institut für Meteorologie und Klimaforschung (IMK) und das Forschungszentrum Jülich (FZJ), die u.a. zur Optimierung der Berechnungsalgorithmen führen soll, werden die Messdaten nach einer Konvertierung in einer MongoDB-Datenbank gespeichert. Auf die Datenbank greifen mehrere Module lesend und schreibend zu (siehe Abbildung 1).

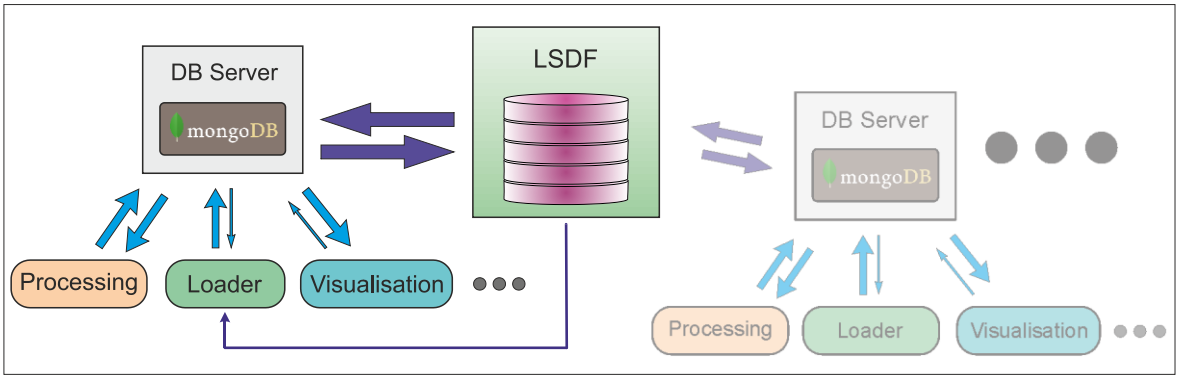


Abbildung 1: Das Systemumfeld der GLORIA-Datenauswertung

Neben den Messdaten werden Prozessierungs-, Konfigurations- und Ergebnisdaten gespeichert, was die Gesamtdatenmenge bis Ende aller Auswertungen beträchtlich ansteigen lassen wird.

1.2 Struktur der GLORIA-Messdaten

GLORIA ist ein neuartiges abbildendes Fourier-Spektrometer (Imaging FTS). Es besitzt als wesentliches Element zur Datengewinnung ein Detektorfeld der Größe 254x254 Pixel. Über das Detektorfeld wird kontinuierlich Infrarotstrahlung aus der Atmosphäre in Zeitabständen von 15 ns bis 300 ms aufgenommen. Die gewonnenen Pixel-Daten werden für vorgegebene Zeiträume zu so genannten Cub-Dateien transformiert und zusammengefasst. Je nach Messkonfiguration entsteht innerhalb von ca. 3 – 13 Sekunden eine Datei mit einer Größe von 100 MB bis 1 GB. Ein ausgewählter Flug dauerte ca. 10 Stunden, wobei ca. 11.000 Dateien mit insgesamt 2 TB Größe erzeugt wurden. Die Messungen wurden dabei teilweise durch Kalibrierungen und weitere Tests unterbrochen und es wurde – abhängig von wissenschaftlichen Zielsetzungen – ein Teilfeld des Detektors von der Größe 128x48 Pixeln verwendet.

Die räumliche Struktur des Inhalts einer Cub-Datei zeigt Abblidung 2. In der Datei werden die einzelnen „Aufnahmen“ des Detektorfelds (Slices) hintereinander abgespeichert. Alle Werte eines gegebenen Pixels werden zu Interferogrammen zusammengefasst.

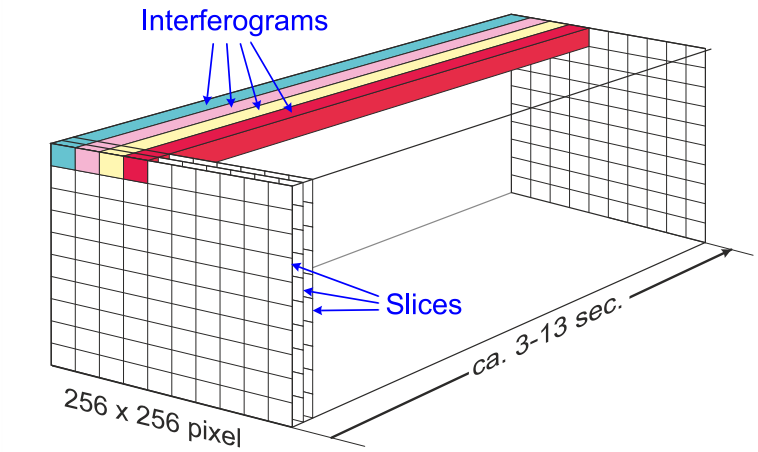


Abbildung 2: Struktur einer Cub-Datei

Die Wissenschaftler sind besonders an der Prozessierung von Interferogrammen interessiert. Da die Daten einer Cub-Datei Slice-orientiert abgelegt sind, müssen sie für eine interferogrammorientierte Speicherung in MongoDB vorher konvertiert werden.

2 Anforderungen an eine Datenbank zur Verwaltung der GLORIA-Daten

An die Verwaltung der Daten durch eine Datenbank wurden folgende Hauptanforderungen gestellt, auf die in den folgenden Abschnitten näher eingegangen wird:

- Zugriff auf Daten und Metadaten mit einer einzigen Abfrage
- Frei verfügbare Datenbank mit einfacher Verwaltung
- Die Daten müssen leicht erweiterbar sein für unterschiedliche Datentypen
- Schneller Zugriff auf große Datenmengen
- Verfügbarkeit leistungsfähiger Schnittstellen für Java, Python, C++ usw.
- Möglichkeit der Speicherung von Daten in Arrays

2.1 Zugriff auf Daten und Metadaten mit einer einzigen Abfrage

Es soll möglich sein, über eine einzige Abfrage sowohl Metadaten als auch die Messdaten zu erhalten, da auf die GLORIA-Daten über das Internet zugegriffen wird. In den meisten ähnlich gelagerten Fällen werden die Messdaten als Dateien im Dateisystem abgelegt und lediglich die Referenzen darauf in der Datenbank verwaltet. Das erfordert

allerdings mindestens einen zweiten Zugriff auf das System über eine andere Verbindung, um an die Dateien zu gelangen.

2.2 Frei verfügbare Datenbank mit einfacher Verwaltung

Um für die Entwicklung des Gesamtsystems viele Experimente und Tests auf unterschiedlichen Plattformen und Computern durchführen zu können, ist es erforderlich, dass das Datenbanksystem einfach und schnell installiert und ohne großen Aufwand administriert werden kann. Mit einer frei verfügbaren Datenbank müssen dabei keine Lizenzen beachtet werden.

2.3 Leichte Erweiterbarkeit der Daten für unterschiedliche Datentypen

Da sich die Anforderungen und Inhalte der zu speichernden Metadaten während der Entwicklung ständig ändern können, ist es erforderlich, Erweiterungen des Datenbankschemas mit geringstmöglichem Aufwand durchführen zu können. Schemafreie Datenbanken sind prädestiniert für diese Anforderungen.

2.4 Schneller Zugriff auf große Datenmengen

Die Auswertung der GLORIA-Daten beschränkt sich nicht auf eine einmalige Berechnung, sondern es werden die Ergebnisse dazu benutzt, die vorhandenen Algorithmen zu validieren und zu optimieren. In diesem Iterationsprozess finden häufig Berechnungen statt, welche Werte benötigen, die über einen gesamten Flug anfallen, d.h. es müssen ständig sehr große Datenmengen zur Verfügung gestellt bzw. geladen werden können.

2.5 Verfügbarkeit leistungsfähiger Schnittstellen für Java, Python, C++ usw.

Die Softwareumgebung von GLORIA verwendet mehrere unterschiedliche Ansätze, um die Messdaten zu prozessieren. Ergebnisse der Berechnungen sind Interferogramme, Spektren und Spurengasverteilungen. Um die Anforderungen zu erfüllen, werden unterschiedliche Softwarewerkzeuge mit unterschiedlichen Programmiersprachen verwendet, d.h. eine Datenbank muss leistungsfähige Programmierschnittstellen für die im GLORIA-Projekt besonders wichtigen Sprachen Java, Python und C++ anbieten.

2.6 Möglichkeit der Speicherung von Daten in Arrays

Die Mess- und Ergebnisdaten sowie zahlreiche Prozessierungsdaten liegen in Form von Arrays vor. Frühere Tests mit sehr großen Satellitendatenmengen zeigten, dass besonders Datenbanken geeignet sind, die Arrays als Datentypen anbieten. Nur wenige Datenbanken bieten Arrays an (z.B. Oracle™, Firebird, MongoDB), wobei mehrere leistungsfähige Schnittstellen für den Arrayzugriff vor allem von Oracle und MongoDB angeboten werden. Während z.B. Oracle für jeden (primitiven) Datentyp Arrays anbietet, ist das Angebot von MongoDB auf seine Standard-BSON-Typen begrenzt [MongoDB, 2014, BSON, 2014].

3 Die NoSQL-Datenbank MongoDB

MongoDB ist eine dokumentorientierte Open-Source-Datenbank, die für eine große Bandbreite von Applikationen eingesetzt wird. Sie gehört zur Gruppe der NoSQL-Datenbanken und ist hinsichtlich großer Datenmengen und verteilter Daten optimiert [MongoDB, 2014].

Ein besonderer Vorteil gegenüber relationalen Datenbanken ist die Schemafreiheit, d.h. die Daten sind nicht zwingend an eine vorgegebene Datenstruktur oder Reihenfolge gebunden und einzelne Datensätze können zu einem späteren Zeitpunkt auf einfache Weise mit beliebigen Datentypen erweitert werden. Das kommt der Entwicklung wissenschaftlicher Softwarewerkzeuge entgegen, bei der sich die Anforderungen während der Entwicklung ständig ändern und erweitern können.

4 Erfahrungen mit MongoDB im GLORIA-Projekt

Die Analyse der Anforderung hat ergeben, dass MongoDB prinzipiell sehr gut für die Verwaltung der GLORIA-Daten geeignet ist. Umfangreiche Tests sollten klären, wie sich die Datenbank hinsichtlich der einzelnen Anforderungen verhält.

Die meisten Tests wurden durchgeführt, um die Verwendbarkeit und Stabilität von MongoDB für die gegebenen Anforderungen zu ermitteln. Die größte Herausforderung

dabei ist das Testen mit sehr großen Datenmengen, wenn sie zuvor erst konvertiert und nicht nur kopiert und geladen werden müssen.

Das Konvertieren und Laden der gewählten HALO/GLORIA-Flugdaten dauerte ca. 34 Stunden für die gegebenen 11.500 Cub-Dateien (Gesamtgröße ca. 2 TB). Aus diesem Grund müssen einige wichtig konfigurationsabhängige und strukturelle Entscheidungen getroffen werden, bevor die Daten geladen werden. Die folgenden Abschnitte beschreiben einige ausgewählte Aspekte für die Einschätzung von MongoDB.

4.1 Testumgebung

Das Testumfeld bestand vorwiegend aus 64-bit-Linux-Rechnern (Server: Ubuntu (lokal) und CentOS 6 (remote), Clients: Ubuntu und Fedora), das LSDF (s. Abschnitt 1.1) ist über eine 10-Gigabit-Ethernet-Verbindung angeschlossen.

Für die Tests wurde ein ausgewählter HALO/GLORIA-Flug mit einem Gesamtumfang von 2 TB verwendet. Er umfasst ca. 11.500 Datensätze (Cub-Dateien) mit insgesamt 75 Mio. Interferogrammen, die zusammen mit ihrer ID (z.B. „20120926_060047“) sowie deren X- und Y-Position im Detektorfeld in jeweils einem eigenen Dokument abgelegt sind. Für die ID, der X- und Y-Position wurde ein kombinierter Index erzeugt. Die Größe der Indexdatei von MongoDB war ca. 9 GB groß.

4.2 Zugriff auf Daten und Metadaten mit einer einzigen Abfrage

Da die Interferogramme als größte Dateneinheiten als eigene Objekte (Byte-Arrays) in den MongoDB-Dokumenten, d.h. in der Datenbank selbst gespeichert sind, ist ein Zugriff über eine einzige Abfrage leicht möglich. Oracle™ bietet als eines der ganz wenigen Datenbanksysteme eine ähnliche Lösung mit den BFiles an. Das sind binäre Dateien, welche im Dateisystem abgelegt sind, auf die jedoch über SQL-Befehle zugegriffen werden kann.

4.3 Installation und Handhabung

Das Herunterladen und Installieren von MongoDB geht sehr einfach vonstatten, es existiert eine sehr umfangreiche und sehr gute Dokumentation im Internet [MongoDB,

2014]. Daneben gibt es inzwischen eine große Internet-Community, die sich mit MongoDB und deren Programmierschnittstellen beschäftigt.

Wie andere Datenbanken auch, bietet MongoDB eine Commandline-Shell an. Hierbei wird über Javascript-Befehle kommuniziert. Das Anlegen und Verwalten von Benutzern und deren Zugriffsrechten sowie von Datenbankinstanzen, Collections und Indizes ist sehr einfach und meist intuitiv.

MongoDB legt seine Daten in Dateien im Dateisystem ab, indem je nach aktuellem Datenbestand weitere Dateien mit der Größe von 2 GB erzeugt werden. Die maximale Größe eines Dokuments beträgt 16 MB (Interferogramm ≈ 150 kB). Größere Datensätze können in so genannten GridFS-Dateien gespeichert werden, einer Art BLOB-Datei (Binary Large Object) die durch einzelne Datenbankobjekte referenziert werden können.

4.4 BSON-Datentypen und Konvertierungen

MongoDB speichert seine Dokumente im BSON-Format (BSON, 2014). Der „kleinste“ BSON-Typ ist ein 32-bit-Integer. Die Slices bzw. Interferogramme sind jedoch in 16-bit-Short-Werten (Arrays) abgelegt. Wenn die Daten in 32-bit-Integerwerten gespeichert werden, so wird das Datenvolumen unnötigerweise verdoppelt, weshalb die Short-Werte vor dem Speichern in Byte-Arrays konvertiert werden.

Zu Beginn des Projekts bestand der Wunsch, dass aus den interferogrammbasierten Daten direkt die zugehörigen Slices an einer gegebenen Stelle des Interferogramms zurückgeliefert werden sollen. Eine einfache Möglichkeit zur Lösung der Anforderung stellt das Speichern in Arrays dar. MongoDB bietet hierzu komfortable Zugriffs-möglichkeiten auf Arrays an, deren Elemente allerdings BSON-Typen sein müssen. Ein BSON-Typ, der das o.g. Problem der 16-bit-Short-Werte zusätzlich löst, ist der Typ Binary, ein Objekt, das Byte-Arrays hält. Nun ist es möglich, ein Array aus Binary-Elementen zu erstellen, dessen Elemente jeweils 2 Byte groß sind und die Short-Werte abbildet. Die Tests zeigten jedoch, dass Binary-Arrays erstaunlich langsam zu laden sind und dass die Zugriffe auf deren einzelne Elemente ebenfalls sehr zeitraubend ist.

Deshalb wurde entschieden, die Interferogramme als einzelne Byte-Arrays zu speichern und auf den datenbankseitigen Zugriff auf Array-Elemente zu verzichten.

4.5 Indizierung und Datenzugriff

Für eine sehr schnelle Suche in großen Datenbeständen bietet MongoDB einen sehr leistungsfähigen Index- und Cachingmechanismus an. Die fixe Tabellenstruktur von relationalen Datenbanken erleichtert die Bestimmung der Datentypen der Spaltenelemente. Dieser Vorteil, der sich durch eine akzeptable Geschwindigkeit bei der Suche nach nicht indizierten Spaltenelementen äußert, kann bei MongoDB nur dadurch kompensiert werden, dass die Datensätze ohne Restrukturierung erweitert werden können. Dieser Vorteil der Schemafreiheit wird allerding dadurch erkauft, dass bei sehr großen Datenmengen für alle interessierenden Elemente Indizes erzeugt werden müssen. Eine Suche auf nicht indizierte Elemente dauert recht lange, eine Suche auf indizierte Werte ist jedoch auch bei sehr großen Datenmengen (ca. 57 Mio. Dokumente) sehr schnell.

So dauerte die Suche nach allen Inteferogrammen einer Cub-Datei durchschnittlich weniger als eine Sekunde, um die Liste der Dokumente zu erhalten. Das Herunterladen auf einen Client-Rechner ist dabei allerdings nicht inbegriffen, sondern ausschließlich die Suche in der Index-Datei.

Beliebige Abfragen jeglicher Komplexität können über Javascript-Befehle erfolgen, entweder über die Command-Line-Shell oder über die Programmierschnittstelle. Im Gegensatz zu SQL-Abfragen müssen die Dokument-Elemente, welche nicht angezeigt werden sollen (z.B. Byte-Arrays), bei der Abfrage explizit ausgeschlossen werden.

4.6 Laden und Lesen von Daten

Die Tests haben ebenfalls ergeben, dass das Lesen und Herunterladen von Daten nicht nur vom sehr leistungsfähigen Caching-Mechanismus von MongoDB abhängt sondern auch von der Reihenfolge der Speicherung. So werden sämtliche Interferogramme einer Cub-Datei sehr viel schneller geladen als alle Interferogramme für ein gegebenes Pixel des Detektorarrays, obwohl sich beide Datenmengen nur um den Faktor 2 unterscheiden.

5 Zusammenfassung

Die Speicherung der GLORIA-Daten für die gegebenen Anforderungen stellt für jedes Datenbanksystem eine große Herausforderung dar. Das Testszenario beinhaltete einen HALO/GLORIA-Flug mit ungefähr 11.500 Cub-Dateien und 57 Mio. Interferogrammen und einer Gesamtgröße von 2,2 TB. Mit MongoDB wurde ein gangbarer Weg gefunden, um die Daten zu verwalten. Folgende Vorteile sprechen für einen Einsatz von MongoDB:

- Freie Open-Source-Software, einfache und schnelle Installation sowie einfaches Management der Benutzerkonten und Inhalte.
- Schneller Zugriff auf die Daten. Sowohl Attribute als auch der komplette Datensatz können mit einer einzigen Abfrage ermittelt und heruntergeladen werden.
- Es sind alle benötigten Programmierschnittstellen für die Integration in Applikationen verfügbar.
- Sehr gute Indizierungs- und Caching-Mechanismen

Als Nachteil für das Testszenario stellte sich das Fehlen primitiver BSON-Typen wie short, float, int usw. heraus.

Zurzeit sind die Basistests abgeschlossen. Als nächster Schritt folgt das Laden der Daten der weiteren HALO/GLORIA-Flüge, was zu einer Gesamtgröße der reinen Messdaten von ca. 30 TB führen wird. Weitere Tests werden untersuchen, ob es Unterschiede im Suchverhalten gibt, wenn über eine einzige große Datenbank oder über mehrere Datenbanken für die einzelnen Flüge gesucht wird.

6 Literaturverzeichnis

Sämtliche aufgeführten Internetseiten wurden zuletzt im Mai 2014 besucht.

[BSON, 2014]

BSON: Internetseite der BSON-Spezifikation; http://bsonspec.org/, 2014

[Friedl-Vallon, F. et al, 2014]

Friedl-Vallon, F. et al: Instrument concept of the imaging Fourier transform spectrometer GLORIA, Atmos. Meas. Tech. Discuss., 7, 2301-2337, doi:10.5194/amtd-7-2301-2014, 2014

[Garcia, A. O. et al, 2011]

Garcia, A. O. et al: Data-intensive analysis for scientific experiments at the Large Scale Data Facility, Proceedings on Symposium on Large Data Analysis and Visualization (LDAV 2011), IEEE Computer Society Press, pp. 125-126, DOI: 10.1109/LDAV.2011.6092331, 2011

[GLORIA, 2014]

GLORIA: Internetseite der Helmholtzgesellschaft über das Instrument GLORIA: http://gloria.helmholtz.de/,2014.

[GLORIA2, 2014]

GLORIA2: GLORIA-Internetseite des Earth Observation Portal: https://directory.eoportal.org/web/eoportal/airborne-sensors/gloria, 2014.

[HALO, 2014]

HALO: Internetseite der DLR (HALO): http://www.halo.dlr.de/, 2014.

[Kleinert, A. et al, 2014]

Kleinert, A. et al: Level 0 to 1 processing of the imaging Fourier transform spectrometer GLORIA: Generation of radiometrically and spectrally calibrated spectra, Atmos. Meas. Tech. Discuss., 7, 2827-2878, doi:10.5194/amtd-7-2827-2014, 2014

[MongoDB, 2014]

MongoDB: Internetseite von MongoDB; http://www.mongodb.org/, 2014.

Aus Klein mach Groß: Komposition von OGC Web Services mit dem RichWPS ModelBuilder

Felix Bensmann, Hochschule Osnabrück, f.bensmann@hs-osnabrueck.de

Rainer Roosmann, Hochschule Osnabrück, r.roosmann@hs-osnabrueck.de

Jörn Kohlus, Landesbetrieb für Küstenschutz, Nationalpark und Meeresschutz Schleswig-Holstein, joern.kohlus@lkn.landsh.de

Abstract

Distributed geospatial data and services can be used to form Spatial Data Infrastructures (SDI). Current technological developments enable SDI to process large amounts of geospatial data. In this context, Web Processing Services (WPS) can be used as an open interface standard for accessing and executing geospatial processes. Since many SDI are based on Service-Oriented Architectures (SOA), the orchestration of services is enabled by composing existing services. The results are higher services, e.g. complex geospatial applications. The RichWPS research project focuses on analysing and providing practical approaches of web service orchestration. For this, software components are developed to build a modular orchestration environment. The main client-side application is the ModelBuilder.

Einleitung

Der Web Processing Service (WPS) stellt die Basis für eine einheitliche Nutzung verteilter Verarbeitungsdienste in der Geoinformatik dar. Im Mittelpunkt des WPS-Standards stehen Prozesse, die über einen WPS abgefragt und aufgerufen werden können. Der Begriff „Prozess“ umfasst hierbei Alles, was im weitesten Sinne als Algorithmus verstanden werden kann. Eine Einschränkung auf ausschließlich raumbezogene Prozesse und / oder Daten ist nicht sinnvoll und nicht gefordert. Zur Durch-

führung komplexer Arbeitsprozesse lassen sich vorhandene, verteilte Prozesse kombinieren. Dies entspricht der Orchestrierung von Web-Services, wie aus den service-orientierten Architekturen (SOA) bekannt. Obwohl es ein wesentliches Konzept verteilter Applikationen darstellt, findet die Orchestrierung in aktuellen Geodateninfrastruktur (GDI) aktuell noch selten Anwendung [Zhao, 2012]. Die RichWPS-Orchestrierungsumgebung erweitert GDIs um Orchestrierungsfunktionalität für WPS-Prozesse. Die zentrale Applikation auf der Client-Seite ist der ModelBuilder.

1 WPS: von kleinen und großen Prozessen

Aufgrund der INSPIRE-Richtlinie sind Behörden in der Europäischen Union aufgerufen, am Ausbau einer EU-weiten GDI mitzuwirken. Zur Umsetzung geeignet sind die Web-Service-Standards des Open Geospatial Consortiums (OGC). Diese sollen die Interoperabiliät zwischen einzelnen Diensten ermöglichen [Schäffer, 2009]. Spezifische, als Teil einer GDI bereitzustellende Dienste, wie Download- oder View-Services [INSPIRE, 2008] lassen sich eindeutig vohandenen OGC Web-Services (OWS) wie dem Web Feature Service (WFS) bzw. der Web Map Service (WMS) zuordnen. Daneben gibt es den WPS [OGC, 2007], der speziell für die (Geodaten-) Prozessierung konzipiert ist [Lanig, 2010]. Mit seinem flexiblen Interface kann ein großer Bereich möglicher Anwendungen abgedeckt werden.

Ein WPS-Server bietet einen WPS an und ermöglicht die serverseitige Prozessierung definierter Algorithmen. Grundsätzlich werden vorhandene Prozesse verwaltet, angeboten und letztendlich in einer definierten Umgebung ausgeführt.

Während aktuell in vielen Projekten die Bereitstellung feingranularer Prozesse über einen WPS im Fokus steht, stellt die Orchestrierung den nächsten logischen Schritt dar, um auf der Basis vorhandener Dienste ganze Anwendungsfälle oder gar Arbeitsprozesse zu komponieren. Der Schlüssel hierzu ist die Orchestrierung vorhandener Dienste, also die automatisierte Abfolge von Aufrufen einzelner WPS-Prozesse nach einer formalen Ablaufbeschreibung durch eine Orchestration Engine [Foerster, 2011]. Dabei lassen sich auch herkömmliche Web Services mit einbeziehen. Der so

komponierte Arbeitsprozess kann wiederum wie ein normaler WPS-Prozess auf einem WPS-Server bereitgestellt werden.

Abseits des Umstands, dass durch INSPIRE eine verteilte Architektur vorgeschrieben wird, kann neben dem Einsatz von Bereitstellungsdiensten das verteilte Angebot von Verarbeitungsdiensten ebenfalls einen Mehrwert gegenüber lokalisierter Datenverarbeitung haben. Die Durchführung von geographischen Analysen findet zur Zeit mehrheitlich auf lokal installierten Geoinformationssystemen (GIS) statt. Dabei belasten langläufige Berechnungen und große Datenmengen den Arbeitsplatzrechner unnötig. Ein weiteres Problem ist, dass lokale Funktionalität nicht einheitlich und auch nicht von mehreren Anwendern genutzt werden kann. Die gemeinsame Nutzung von Geodaten ist bereits weiter verbreitet. Sie finded in GDIs z.B. mit dem WFS, WCS und dem WMS statt. Die Ausweitung auf Prozessierungsdienste, wie den WPS, ist der nächste Schritt. Durch den Einsatz von Orchestrierung können diese Dienste erst effektiv genutzt werden.

Orchestrierung ermöglicht beispielsweise, dass ein von Dritten bereitgestellter inkompatibler Bereitstellungsdienst oder Prozessierungsdienst an eigene Bedürfnisse angepasst wird. Dies kann geschehen, indem er mit einem Veredlungsdienst zusammengefasst wird, der die benötigten Anpassungen vornimmt. Darüber hinaus ermöglicht die Orchestrierung Anwendern, Prozesse aus Kompositionen wieder zu verwenden oder Prozesse auszutauschen. Dabei wird zum einen Implementierungsaufwand gespart und zum anderen die gemeinsame Nutzung aufrechterhalten.

In einem größeren Zusammenhang wird hiermit die Grundlage gelegt, geographiebezogene Dienste in nicht-spezialisierte Web-Service-Architekturen zu integrieren. Beispielsweise können im Berichtswesen über Servergrenzen und damit über Behördengrenzen hinweg Berichte individuell parametrisiert, generiert und zusammengeführt werden.

2 Modellierung mit dem ModelBuilder

Das RichWPS-Projekt realisiert eine Orchestrierungsumgebung für OGC Web Services. Zusammen mit der Disy Informationssysteme GmbH entwickelt die Hochschule Osnabrück die benötigten Softwarekomponenten. Eine zentrale Applikation ist dabei der RichWPS-ModelBuilder. Hierbei handelt es sich um ein Werkzeug zur graphischen Modellierung von Arbeitsprozessen (Workflows) aus vorhandenen Prozessen. Weitere Komponenten sind der RichWPS-SemanticProxy, ein Verzeichnisdienst für verfügbare WPS-Dienste und -Prozesse im Zielnetzwerk sowie der RichWPS-Server, ein WPS-Server mit integrierter Orchestrierungsengine.

Die Administration von Servern und Services zählt nicht zu den Aufgaben typischer Anwender wie z.B. Geographen. Der Einsatz von Orchestrierung insbesondere in Verbindung mit dem WPS erfordert jedoch genau das. Der ModellBuilder legt den Fokus auf den Abbau der Hindernisse. Über den ModelBuilder können ein Arbeitsprozess modelliert, diese Beschreibung an die Orchestraton Engine übertragen und der Lifecycle des Prozesses verwaltet werden. Des Weiteren wird ein Debugging, Testing und Profiling bereitgestellter Arbeitsprozesse unterstützt.

Die Oberflächengestaltung des ModelBuilders folgt den Prinzipien etablierter Modellierungswerkzeuge. Abbildung zeigt den Aufbau der Anwendung.

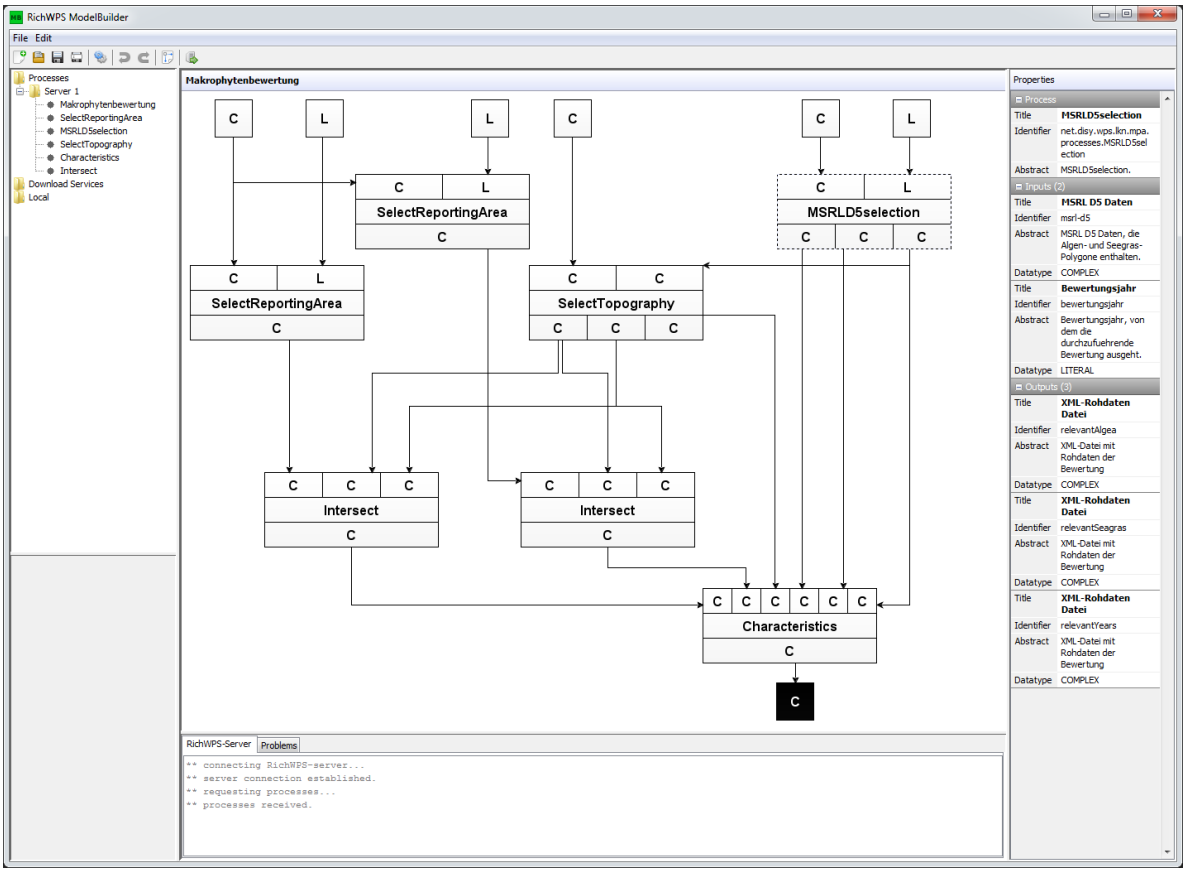


Abbildung 1 Bildschirmaufnahme des RichWPS ModelBuilders

Die Komplexität eines Arbeitsprozesses ist praktisch nicht begrenzt. Durch Abstraktion der Verteilung der WPS und WPS-Prozesse von ihren physikalischen Orten wird jedoch die fachliche Sicht in den Vordergrund gerückt. Für die Modellierung eines Arbeitsprozesses verwendet der ModelBuilder ein speziell entwickeltes Daten- / Kontrollflussdiagramm unter Verwendung einer neuen graphischen domänenspezifischen Sprache (DSL). Abbildung 2 stellt einen beispielhaft erstellten Arbeitsablauf dar.

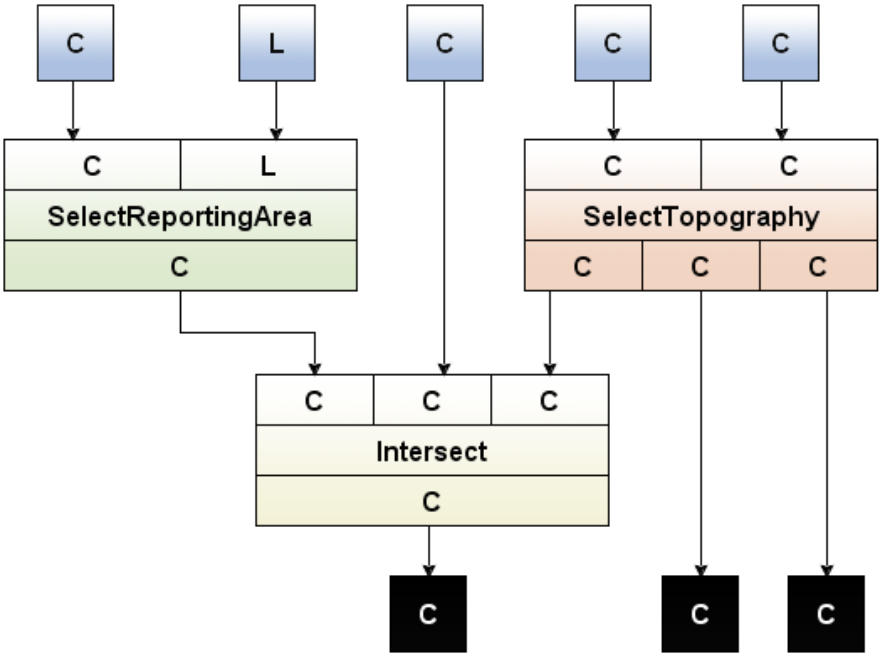


Abbildung 2 Beispielmodell eines Arbeitsablaufes

Das Modell folgt dem Prinzip eines Datenflussdiagramms, für diesen Fall wurde die Notation etwas angepasst. Die Arbeitsrichtung verläuft von oben nach unten. Es gibt Input- und Output-Elemente für den Gesamtprozess, Inputs werden durch weiße Quadrate mit schwarzem Text dargestellt und Outputs durch schwarze Quadrate mit weißem Text. Prozesse werden durch Rechtecke mit kleineren Rechtecken an der oberen und unteren Seite für Prozess-Inputs und -Outputs dargestellt. Inputs und Outputs werden ferner mit Datentypen assoziiert. Ein „C" weist dabei auf den in WPS spezifizieten Datentyp *Complex* hin, ein „B" auf *BoundingBox* und ein „L" auf *Literal*.

Die globalen Inputs stellen Datenspeicher dar, aus denen nur gelesen werden kann, in die prozessgebundenen Inputs kann nur geschrieben werden. Anders herum stellen die globalen Outputs Speicher dar, in die nur geschrieben werden kann und die prozessgebundenen Outputs Speicher, aus denen nur gelesen werden kann.

Der Datenfluss wird durch Pfeile zwischen den verschiedenen Inputs und Outputs dargestellt. Die Steuerung des Kontrollflusses wird später ergänzt.

Der ModelBuilder ermöglicht dem Anwender, sein Modell auf formale Korrektheit zu prüfen. Ist eine Verbindung nicht erlaubt, verhindert die Anwendung, dass sie hergestellt werden kann. Außerdem wird der Anwender durch zahlreiche Kontext-

informationen insbesondere zu den verwendeten Prozessen unterstützt. Diese Art der Modellierung hat sich in bekannten Modellierungswerkzeugen wie der Taverna Workbench bewährt.

Für die Modellierung werden Informationen über vorhandene Prozesse und Dienste benötigt. Diese Aufgabe ist nicht trivial. Der ModelBuilder greift hierzu auf einen Verzeichnisdienst, den SemanticProxy, zu. Dieser stellt die erforderlichen Informationen bereit und bietet semantik-gestütze Suchfunktionen an. Der SemanticProxy ist auf Erweiterbarkeit ausgelegt und soll mit dem Resource Description Framework und der Umsetzung der Linked-Data-Prinzipien mit anderen Daten verlinkt werden können.

Das graphische Modell wird dann zur Beschreibung des Ablaufes eines Arbeitsprozesses in eine textuelle Notation überführt. Das textuelle Modell legt das Laufzeitverhalten des Prozesses auf dem WPS-Server fest, wird an die serversseitige Orchestration Engine übertragen, dort interpretiert, über einen WPS angeboten und letztendlich ausgeführt.

Über den ModelBuilder werden noch weitere Funktionen, insbesonere Testing und Debugging realisiert. Ein Problem, das mit der Orchestrierung auftritt, ist, dass ein Anwender kaum die Möglichkeiten hat, einen Arbeitsablauf auf fachliche Korrektheit zu testen. Diesem Problem soll mit diesen Funktionen begegnet werden. Ein Debugging kann nur auf Orchestrierungsebene stattfinden. Wie bei herkömmlichen Debugverfahren kann der Anwender Zwischenergebnisse einsehen, jedoch wird zunächst nur ein einfacher Request realisiert, der den Arbeitsprozess ausführt und die Zwischenergebnisse in gesammelter Form zurückliefert. Ein detailliertes Debugging mit Break Points und Schritt-für-Schritt-Ausführung ist zunächst nicht geplant.

3 Modellierung in der Praxis

Aufgaben der Umsetzung der Meeresstrategie-Rahmenrichtlinien (MSRL) und der Wasserrahmenrichtlinie (WRRL) gehören zum Aufgabenbereich des Landesbetriebes für Küstenschutz, Nationalpark und Meeresschutz Schleswig-Holstein (LKN). Auf der einen Seite sind geeigente Paramter zur Überwachung des Umweltzustandes zu erheben, auf

der anderen ist eine Bewertung hinsichtlich der Umweltziele der Richtlinien durchzuführen. Bei der Entwicklung der europäischen Umweltrahmenrichtlinien seit der Vogelschutzrichtlinie von 1979 ist ein stetiger Trend zu beobachten, nicht nur die Vergleichbarkeit sondern auch die technische Kommunizierbarkeit der Daten und Berichtsinformationen standardisiert zu verankern [Kohlus, 2009]. Bezüglich der Angaben der Mitgliedsstaaten ist zunehmende Transparenz, Nachvollziehbarkeit und Vergleichbarkeit gefordert. Insbesondere die Nachvollziehbarkeit und Vergleichbarkeit – z. B. mit ähnlichen oder benachbarten Regionen - solcher Bewertungen kann gefördert werden indem der Bewertungsvorgang algorithmisch definiert wird. Nach einer solchen Festlegung kann eine Umsetzung oft weitgehend automatisiert implementieren werden.

Typischerweise beruhen Bewertungen für die genannten Richtlinien auf der Kombination einiger weniger Verfahren, die hier nur kurz skizziert werden:

- Aussagen werden für einen Referenzraum vorgenommen
- Aussagen zu Teilparametern können sich auf unterschiedliche, überlappende Räume beziehen, sie werden mittels einer geometrischen Verschneidung aufeinander bezogen
- Aussagen über Räume werden binär (Ja/Nein), quanitativ oder graduell - oft in Relation zur Flächengröße vorgenommen.
- Über- oder Unterschreitungen von Werten kennzeichnen Qualitätsveränderungen
- häufig gehen mehrere gemessene Faktoren in das Verfahren ein, um sie miteinander in Verbindung zu setzen werden Relationen gebildet und auf ein Skalenniveau transformiert (z.B: ecological quality ration EQR, u.a. [van de Bund 2007], [Romero et al. 2007]).
- zur Bestimmung von Bewertungsstufen wird eine Klassenbildung vorgenommen

Makrophyten werden u.a. als Qualitätsfaktor bei der Bewertung des Europhierungszustandes bei der WRRL genutzt. Dabei gelten Makroalgen – insbesondere Darmtange – als Eutrophierungszeiger und umgekehrt der Bestand mit Seegräsern der Arten

Zostera marina und *Z. noltii* als Hinweis auf eine geringe Belastung [Reise 2008]. Ein von der WRRL-Expertengruppe Makrophyten vorgeschlagenes Bewertungsverfahren wurde in ein elektronisch unterstütztes Verfahren transformiert [Rieger 2013], das in den Rahmen der Geodateninfrastruktur MDI-DE [Wosniok, 2012; Kohlus, 2011; Kohlus 2010] integriert ist. Das Verfahren wurde bereits von Wössner [Woessner, 2013] in vereinfachter Form als Web-Prozessing-Service (WPS) implementiert.

Beim Bewertungsverfahren wird die maximale Ausdehnung des Seegrasbewuchses eines Jahres in Relation zur Wattfläche gesetzt. Für die Berichtsgebiete wird geprüft, ob die Arten präsent sind. Gegenläufig wird die Ausdehnung der Bedeckung der Watten mit makrophytischen Grünalgen betrachtet. Die Grenzkriterien anhand von EQRs zeigt Abbildung 3 beispielhaft für das Berichtsgebiet Nordfriesiches Wattenmeer [Dolch, 2009]. Bei diesem auf Flugzeugbeobachtungen gestützten Verfahren werden die jährlichen Bewertungen über den Berichtszeitraum von sechs Jahren gemittelt angegeben.

	Bewertungsmatrix Nordfriesland Makrophytobenthos-Index							
	Qualitäts-kategorien	0	1	2	3	4	Gewichtung %	Norm-EQR gemäß Gewichtung für 6-Jahre-Interval
		Schlecht	Unbe-friedigend	Mäßig	Gut	Sehr gut		
	Norm-EQR	0 – 0,19	0,2 – 0,39	0,4 – 0,59	0,6 – 0,79	0,8 – 1,0		
Modul Seegras [6]	Eulitorale Fläche (%) [1]	< 2	2 - 4,9	5 - 9,9	10 - 19,9	20 - 100	50	Mittelwerte aller Parameter-EQRs über 6 Jahre
	Anteil ≥ 60 % Bedeckung (%) [2]	< 6	6 - 11,9	12 - 24,9	25 - 49,9	50 - 100	10	
	Präsenz beider Arten (%) [3]	< 20	20 - 39,9	40 - 59,9	60 - 79,9	80 - 100	10	
Modul Grünalgen [7]	Eulitorale Fläche (%) [4]	100 - 15	14,9 - 7	6,9 - 3	2,9 - 1	< 1	20	
	Anteil ≥ 60 % Bedeckung (%) [5]	100 - 50	49,9 - 25	24,9 - 12	11,9 - 6	< 6	10	

Abbildung 3 Grenzkriterien anhand von EQR für das Berichtsgebiet Nordfriesisches Wattenmeer

Dieses in der Implementierung gut untersuchte Verfahren ist hochspezifisch und nur für die explizite Fragestellung einsetzbar. Es enthält aber viele Prozessschritte entsprechend der oben gegebenen Liste, die auch im Kontext anderer Bewertungs-verfahren Verwendung finden. Im Vorhaben wird daher der Gesamtprozess in granuläre Komponenten zerlegt, die auf ihre Eignung zur Wiederverwendbarkeit geprüft werden.

Zu differenzieren sind eher technische Prozesse, die Eingangsdaten anhand bestimmter Parameter selektieren und transformieren oder Berechnungsergebnisse als Chart-Diagramme darstellen, die in einen Prozess zur Berichtsgenerierung eingehen. Deren Wiederverwendung wird über diesen konkreten Arbeitsprozess hinaus groß sein, wenn die Prozesse generisch und anpassbar sind. Gleiches gilt für geometrische und toplogische Analysen, also die verteilte Bereitstellung typischer GIS-Funktionalität. Diese Basisprozesse wiederum werden zu spezifischen orchestrierten Prozessen orchestriert, die einen fachspezifischen Anwendungsfall abbilden. Dessen Wiederverwendung wird eher gering sein und sich auf die Nordsee-Anrainer reduzieren, die eine gleiche Methodik einsetzen.

Die Orchestrierung soll unter Anwendung des ModelBuilders erfolgen. Wesentlich ist hier die einfache Definition eines Arbeitsprozesses aus vorhandenen Basis- und Anwendungsfall-Prozessen. Das erstellte Modell wird an einen RichWPS-Server übertragen, von diesem interpretiert und bereitgestellt. Der Anwendungsfall macht die Unterstützung für zusätzliche komplexe Datentypen notwendig. Hierzu wird der RichWPS-Server entsprechend erweitert.

Der oben erwähnte Vorgang zur Makrophytenbewertung dient der Konzeptvalidierung von RichWPS und soll Schwächen des Verfahrens aufdecken. Abbildung 4 beschreibt, wie der Gesamtvorgang in einzelne WPS-Prozesse aufgeteilt werden kann. Wie sich herausstellt, ist die Erstellung der benötigten Prozesse kein triviales Problem. Zwar ist die Erstellung von Basisprozessen aus grundlegenden geographischen Verfahren Gegenstand vieler Projekte, und diese stehen auch zur Verfügung, in diesem Fall ist die Wiederverwendung im Anwendungsfall nicht gegeben. Die im Beispiel verwendeten Prozesse wurden durch Zerlegung des von Wössner erstellten WPS-Prozesses in kleinere Prozesse realisiert.

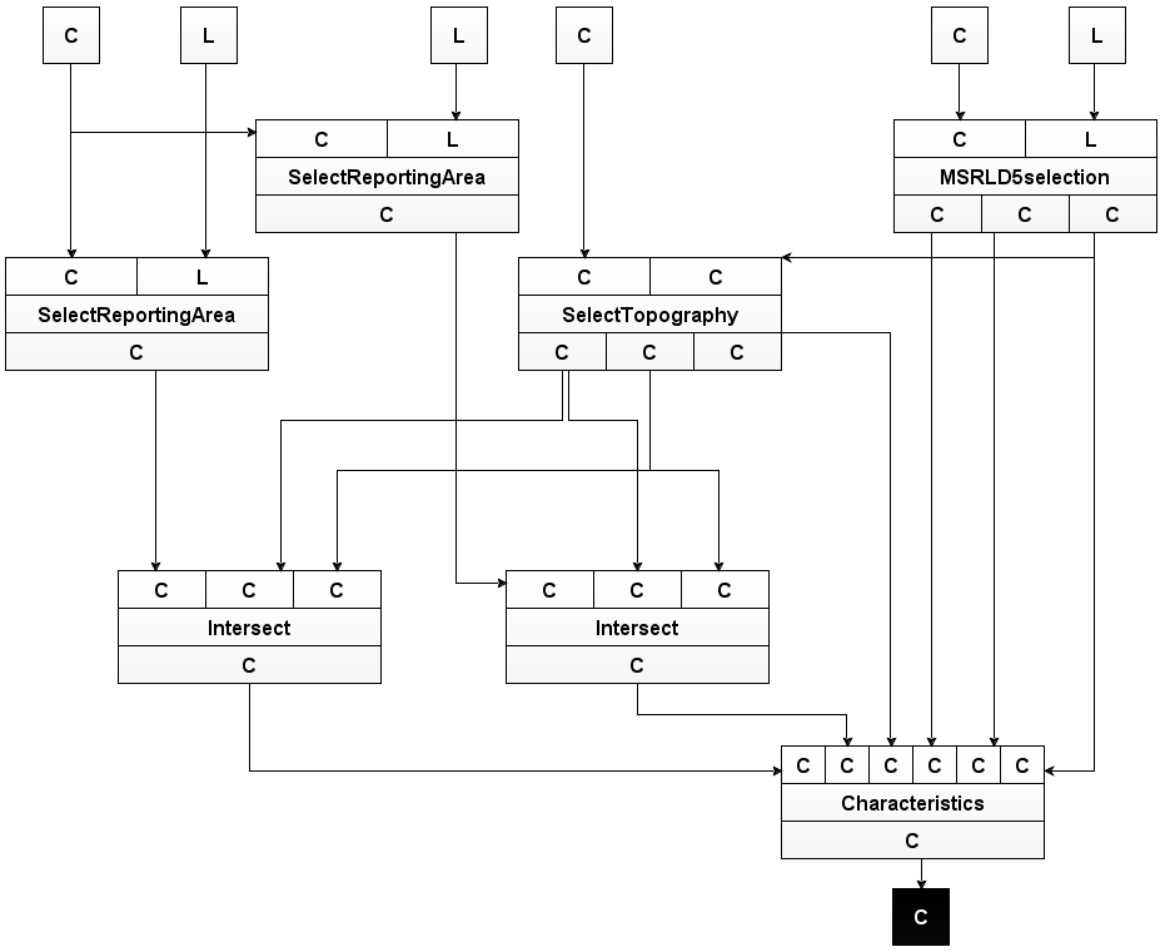


Abbildung 4 Dekomposition der Makrophytenbewertung

Für die praktische Anwendung kann die Ausarbeitung von domänenspezifischen Richtlinien für die Erstellung von WPS-Prozessen und Datentypen helfen, einen möglichst hohen Grad der Wiederverwendbarkeit zu erreichen. Die Wiederverwendbarkeit kann weiter durch Parametrisierung der Prozesse erhöht werden.

4 Fazit und Ausblick

Der Einsatz verteilter Dienste zur Förderung der behördlichen Kooperation und einheitlichen Resourcennutzung zahlt sich aus. Die wird nicht zuletzt durch INSPIRE begründet. Die im RichWPS-Projekt entworfenen Client- und Serverkomponenten in Kombination mit zwei DSLs dienen der Realisierung einer Orchestrierungsumgebung deren Ziel Einfachheit ist. Einfachheit bezieht sich hierbei sowohl auf Gebrauchstauglichkeit des ModelBuilders als auch auf den zu Grunde liegenden Ansatz mit einer DSL. Aus Anwendersicht steht jedoch das Bedienkonzept im Vordergrund, das durch den

ModelBuilder realisiert wird. Dies ist darauf ausgelegt, dass der Anwender ein Minimum an Informationen selbst eingeben bzw. modellieren muss und dabei mit allen benötigten Informationen unterstützt wird. Zum aktuellen Zeitpunkt wird diese Anforderung als erfüllt angesehen. Die Realisierung des Anwendungsfalls konnte durchgeführt werden, jedoch müssen initial benötigte Prozesse zuvor erstellt werden.

In den nächsten Schritten wird der ModelBuilder um Funktionalität zur Unterstützung weiterer OWS erweitert. Auch werden Schritte zur Umsetzung der Testfunktionalität unternommen. Ein weiterer wichtiger Punkt ist die Beschreibung von Prozessierungsergebnissen mit Metadaten, die den Entstehungsweg eines Ergebnisses dokumentieren. Die RichWPS Orchestrierungsumgebung ist ein geeignetes Umfeld, um diese Funktionen weiter zu untersuchen. Derzeit ist diese Aufgabe jedoch zu umfangreich, um in diesem Projekt noch Berücksichtigung zu finden.

5 Literaturverzeichnis

[Dolch, 2009] Dolch, T., Buschbaum, C., und Reise, K. (2009): Seegras-Monitoring im Schleswig-Holsteinischen Wattenmeer 2008 - Forschungsbericht zur Bodenkartierung ausgewählter Seegrasbestände http://www.schleswigholstein.de/UmweltLandwirtschaft/DE/WasserMeer/07_KuestengewMeere/02_Meeresmonitoring/01_BiologischesMonitoring/PDF/Seegras_Bericht_2008__blob=publicationFile.pdf

[Foerster, 2011] Foerster, T.; Schäffer, B.; Baranski, B. et al.: Geospatial Web Services for Distributed Processing – Application and Scenarios. Geospatial Web Services: Advances in Information Interoperability, 2011

[INSPIRE, 2008] Infrastructure for Spatial Information in the European Community: INSPIRE Network Services Architecture, Stand 19.7.2008. Online verfügbar: http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/D3_5_INSPIRE_NS_Architecture_v3-0.pdf

[Kohlus, 2011] Kohlus,J.; Reimers, H-C.; Diederichs, B.: Überlegungen zu einem prototypischen Knoten der Marinen Dateninfrastruktur Deutschland. In: Traub, K.-P., Kohlus, J. & T. Lüllwitz (Hrsg.): Geoinformationen für die Küstenzone. Band 3. Norden & Halmstad (S), S. 117-126, 2011.

[Kohlus, 2010] Kohlus, J.; Reimers, H-C.: Neue Herausforderungen im Datenmanagement für das europäische Meeresmonitoring - Das Projekt MDI-DE - Marine Daten-Infrastruktur in Deutschland. In: Schwarzer, K., Schrottke, K., Stattegger, K. (eds.). From Brazil to Thailand - New Results in Coastal Research. Coastline Reports (16), pp. 115-126. EUCC - Die Küsten Union Deutschland e.V., Rostock.

[Kohlus, 2009] Kohlus, J., Diederichs, B., Kazakos, W. & C. Heidmann: Von den Metadaten zum Bericht In: Traub, K-P., Kohlus, J. & T. Lüllwitz (Hrsg.): Geoinformationen für die Küstenzone - Band 2, Beiträge des 2. Hamburger Symposiums zur Küstenzone und Beiträge des 7. Strategie-Workshops zur Nutzung der Fernerkundung im Bereich der BfG/Wasser- und Schifffahrtsverwaltung; S.137 - 152, Norden, Halmstad (S), 2009.

[Lanig, 2010] Lanig, Sandra; Zipf, Alexander: Standardkonforme Geodatenverarbeitung und Diensteorchestrierung am Beispiel der Solarpotenzialanalyse mittels OGC Web Processing Service (WPS). AGIT 2010. Symposium für Angewandte Geoinformatik. Salzburg. Östereich, 2010

[OGC, 2007] Open Geospatial Consortium: OpenGIS Web Processing Service Version 1.0.0, Online verfügbar: http://www.opengeospatial.org/standards/wps, 2007

[Reise, 2008] Reise K. & J. Kohlus: Seagrass, an indicator goes astray. In: Colijn, F., Doerffer, R. & J. Beusekom (Eds.): Observing the Coastal Sea. An Atlas of Advanced Monitoring Techniques. Helmholtz

Centres - GKSS Research Centre and Alfred Wegner Institute for Polar and Marine Research. 1/2008, p. 64-67, Geesthacht.

[Rieger, 2013] Rieger, A.; Kohlus, J.; Traub, K-P: Automatisiertes webbasiertes Verfahren zur ökologischen Bewertung von Makrophyten im Schleswig-Holsteinischen Wattenmeer. In: Traub, K.-P., Kohlus, J. & T. Lüllwitz (Hrsg.): Geoinformationen für die Küstenzone. Band 4, Karlsruhe, 2013. In: Traub, K.-P., Kohlus, J. & T. Lüllwitz (Hrsg.): Geoinformationen für die Küstenzone. Band 4, Karlsruhe.

[Romero, 2007] Romero, Javier; Martınez-Crego, Begon; Alcoverro, Teresa; Pérez, Marta: A multivariate index based on the seagrass Posidonia oceanic (POMI) to assess ecological status of coastal waters under the water framework directive (WFD). Marine Pollution Bulletin 55 (2007) 196–204.

[Schäffer, 2009] Schäffer Bastian: Delegation in Geoprocessing Workflows. Geoinformatics, 2009 17th International Conference on, 2009.

[van de Bund, 2007] van de Bund, Wouter; Solimini, Angelo G.: Ecological Quality Ratios for Ecological Quality Assessment in Inland and Marine Waters. Directorate-General Joint Research Centre, Institute for Environment and Sustainability, Luxembourg: Office for Official Publications of the European Communities, 2007.

[Woessner, 2013] Woessner, R.: Untersuchungen zur praktischen Nutzbarkeit des OGC Web Processing Servide (WPS) Standards, Hochschule Karlsruhe Technik und Wirtschaft, Masterthesis, 2013

[Wosniok, 2012] Wosniok, Ch., Helbing, F., Kohlus, J. & R. Lehfeldt(2012): MDI-DE - Marine Dateninfrastruktur Deutschland: Die Komponenten des Netzwerks am Beispiel des Infrastrukturknoten Schleswig-Holsteins. In: Venzke, J.-F., Vött, A., Flitner, M., Mossig, I.,

Rosol, M. & B. Zolitschka (Hrsg.): Beiträge der 29. Jahrestagung des Arbeitskreises Geographie der Meere und Küsten 28. bis 30. April 2011 in Bremen. Bremer Beiträge zur Geographie u. Raumplanung, 44. Bremen, 2011, S. 144 - 154.

[Zhao, 2012] Zhao, P. et al.: The Geoprocessing Web. Computers & Geosciences, 2012

Aus Groß mach Klein:
Verarbeitung orchestrierter OGC Web Services mit RichWPS Server- und Client-Komponenten

Roman Wössner, Disy Informationssysteme GmbH, roman.woessner@disy.net
Andreas Abecker, Disy Informationssysteme GmbH, andreas.abecker@disy.net
Felix Bensmann, Hochschule Osnabrück, f.bensmann@hs-osnabrueck.de
Dorian Alcacer-Labrador, Hochschule Osnabrück, d.alcacer@hs-osnabrueck.de
Rainer Roosmann, Hoschschule Osnabück, r.roosmann@hs-osnabrueck.de

Abstract

RichWPS is a project aiming at more powerful and more user-friendly support for using the OGC Web Processing Services standard (WPS) for distributed geodata processing in the Web. While we have already discussed in other publications the RichWPS ModelBuilder for geoprocessing-workflow composition as well as the RichWPS approach to geoprocessing-workflow orchestration, this paper sketches the RichWPS Server software as well as some client-side developments facilitating the effective usage of WPS processes.

Zusammenfassung

Das Projekt RichWPS erarbeitet mächtigere und benutzerfreundlichere Werkzeugunterstützung für den OGC Web Processing Services Standard (WPS) für die verteilte Geodatenverarbeitung im Web. Während in anderen Publikationen der RichWPS-Ansatz zur Workflow-Komposition und -Orchestrierung betont wurde, beleuchten wir in diesem Artikel die Funktionalitäten des RichWPS-Servers sowie einige unterstützende clientseitige Entwicklungen.

1 WPS: von kleinen und großen Prozessen

Im Forschungsprojekt „RichWPS“ geht es darum, neue Konzepte zu entwerfen, Funktionserweiterungen umzusetzen und Referenzimplementierungen zu schaffen sowie beispielhafte Pilotanwendungen zu realisieren, die zeigen, wie man den Ansatz der Web Services für die Geodatenverabeitung, speziell auf Basis des OGC-Standards WPS (Web Processing Services)[9] benutzerfreundlicher, anwendungsnäher und praxistauglicher ausgestalten bzw. weiterentwickeln kann.

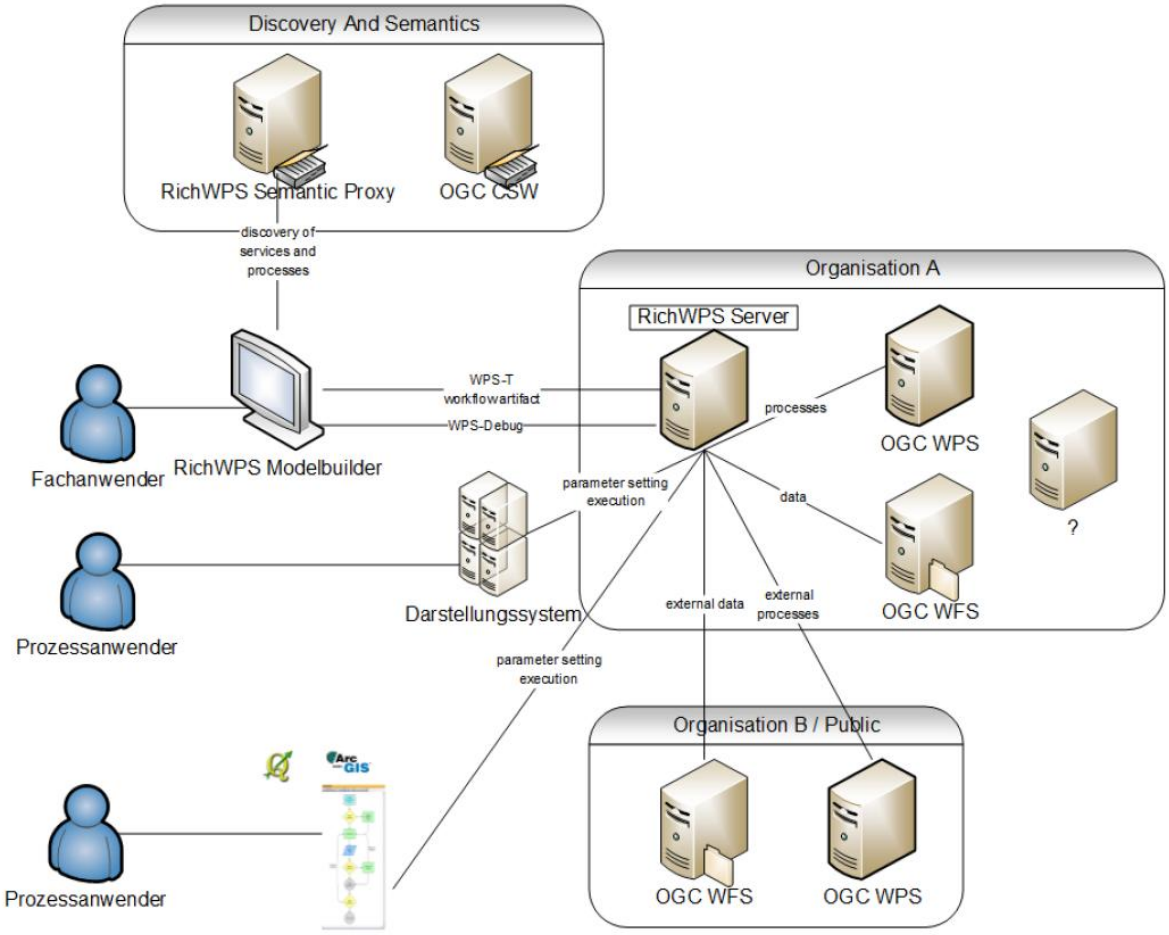


Abbildung 1: Überblick des RichWPS Gesamtsystems
mit Subsystemen und angrenzenden Systemen

Ein wesentlicher Bestandteil dieser Aufgabe ist es, mit einem möglichst einfach zu bedienenden Modellierungswerkzeug komplexe Verarbeitungsprozesse aus einfacheren zusammenzusetzen (vgl. [Bensmann et al, 2014]); diese können dann zusammen mit den erforderlichen Metadaten in einem Repository abgespeichert und veröffentlicht

[9] http://www.opengeospatial.org/standards/wps

werden, dann kann man man bei Bedarf den komplexen Prozess durch Interpretation des Prozessmodells und Rückführung der komplexen Dienste auf einfachere bis zu den Basisbausteinen und durch Ausführen dieser Basisprozesse ausführen (ggf. auf verteilten Rechnern, wenn die Basisdienste von unterschiedlichen Stellen angeboten werden). Während [Bensmann et al, 2014] sich primär dem Modellierungswerkzeug des Projekts widmet (dem RichWPS-ModelBuilder), sollen in diesem Artikel der RichWPS-Server und einige client-seitige Entwicklungen vorgestellt werden.

Abbildung (aus [Bensmann et al, 2013]) skizziert das Zusammenspiel der einzelnen Systeme und Subsysteme in RichWPS: Ein Fachanwender erstellt mithilfe des RichWPS-ModelBuilders komplexe Prozesse aus bestehenden Prozessbausteinen, die ihrerseits als WPS veröffentlicht wurden und über einen CSW sowie ergänzende Metadaten im RichWPS-SemanticProxy beschrieben und auffindbar sind. Der komponierte Prozess wird in Form eines sog. ROLA-Skripts [Bensmann et al, 2014b] auf dem RichWPS-Server ausführbar veröffentlicht. Prozessanwender können dann z.B. aus einem gängigen WPS-Client heraus den Prozess starten und ihm gg. Eingabe-Parameter übergeben. Der Prozess wird vom RichWPS-Server durch Aufrufen der (lokalen und/oder entfernten) Teilprozesse ausgeführt. Der Prozessanwender erhält das Ergebnis über seinen WPS-Client zurückgeliefert. Dabei müssen nicht alle genutzten Dienste Verarbeitungs-dienste (WPS) sein, sondern es können auch Datendienste (WFS) zur Datenbereit-stellung herangezogen werden. Deshalb reden wir allgemein von der Komposition komplexer WPS aus OGC Web Services (OWS).

2 Aufgaben des RichWPS-Servers

Abbildung 2 fasst die Anwendungsfälle des RichWPS Servers zusammen:

(S1) Im Rahmen des Lifecycle Management für zusammengesetzte Prozesse werden mithilfe des RichWPS Servers Prozesse veröffentlicht bzw. ihre Veröffentlichung wieder zurückgezogen. Dies realisiert die Funktionalitäten des *Transactional WPS* (WPS-T) aus [Schäffer, 2008].

(S2) Hauptaufgabe des Servers ist natürlich die Ausführung komplexer Prozesse (im Bild: „Arbeitsabläufe“) durch Interpretation des Prozessmodells, Aufruf lokaler und entfernter Basisdienste und Zusammensetzen des Ergebnisses.

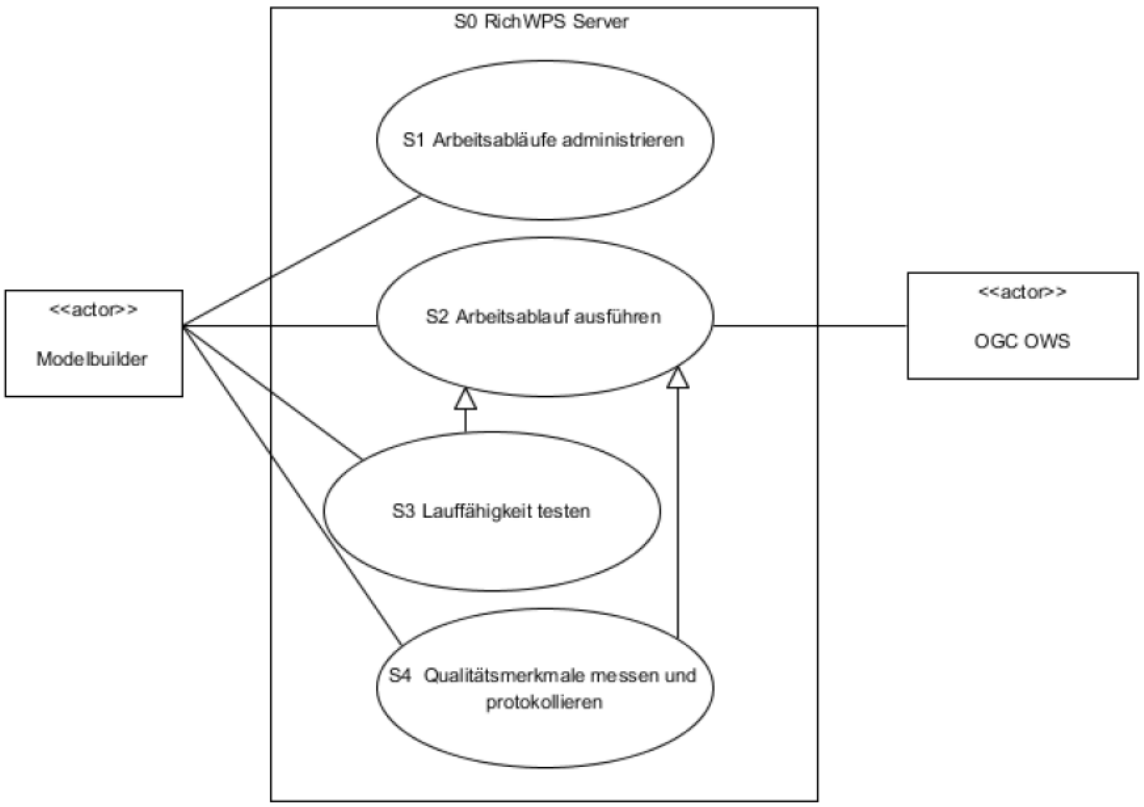


Abbildung 2: Anwendungsfälle des RichWPS-Servers

(S3) Ein Spezialfall der Prozessausführung ist die Ausführung zu Testzwcken während des Modellierungsvorgangs, um eventuelle Modellierungsfehler zu finden.

(S4) Ein weiterer Spezialfall ist die Prozessausführung zu Benchmarking-Zwecken während der Modellierung, um QoS-Aussagen treffen zu können oder Optimierungen der Modellierung zu unterstützen. Konkret umfasst dies zurzeit nur die Messung von Ausführungslaufzeiten für den Gesamtprozess und seine Teilprozesse.

Die Dienste-Schnittstelle des RichWPS-Servers bietet dafür die in Abbildung 3 aufgelisteten Operationen an:

- *GetCapabilities*, *DescribeProcess* und *Execute* sind die aus WPS1.0.0 bekannten Operationen.

- *DeployProcess* und *UndeployProcess* realisieren die WPS-T Funktionalität.
- *TestProcess* und *ProfileProcess* sind für die oben angesprochenen Anwendungsfälle (S3) und (S4) zuständig.
- *GetSupportedTypes* liefert die Liste der vom Server unterstützten Datentypen für Eingabe- und Ausgabedaten.

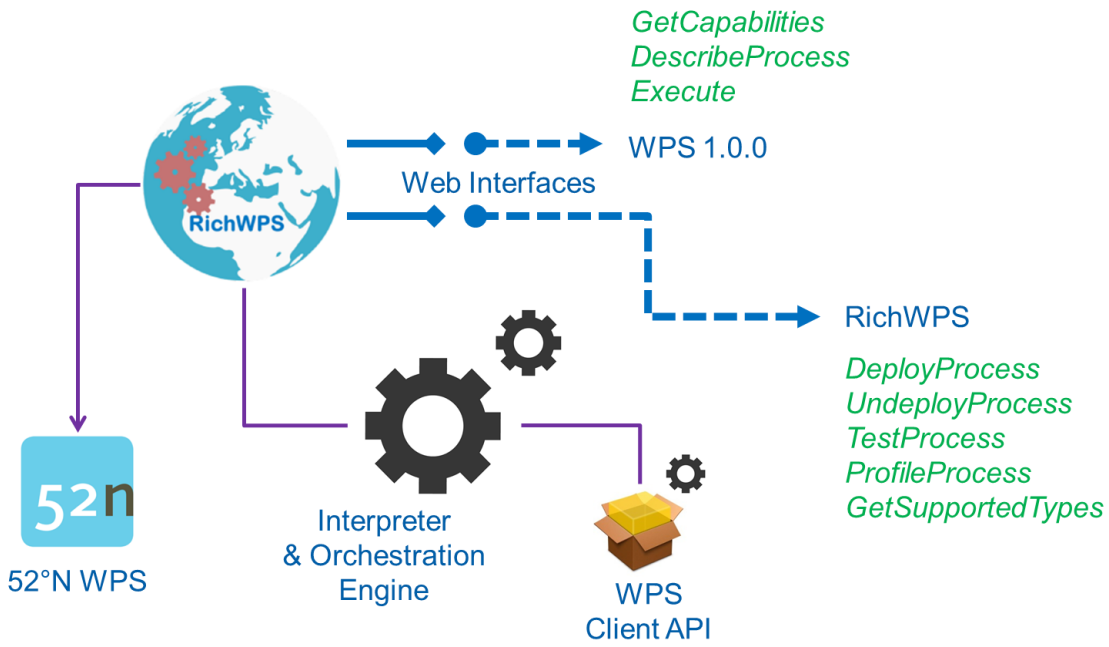


Abbildung 3: Schnittstelle des RichWPS Servers (in grün)

Zur Operation *GetSupportedTypes* noch einige Erläuterungen: den Autoren ist klar, dass bereits heute hervorragende WPS-Lösungen möglich sind, sofern Dienstanbieter und Dienstnutzer identisch sind bzw. zumindest die gleiche WPS-Implementierung verwenden, also z.B. deegree Server und Client, 52°North Server und Client, PyWPS Server und Client usw. Bleibt man innerhalb solcher geschlossener Systeme, gibt es bereits auch durchaus funktionale und benutzerfreundliche Prozessmodellierungs-Werkzeuge. Dies ist aber u.E. gerade ***nicht*** der primäre Sinn und die Zielsetzung dienstebasierte Software-Architekturen. Hier sollte es u.E. darum gehen, dass viele Diensteanbieter lokal mit ihren eigenen Werkzeugen Dienste erstellen und publizieren und dass sich Dritte mit ihren – ggf. anderen – Werkzeugen diese Dienste zunutze machen, auch innerhalb komplexer Workflows, deren Teile von mehreren unterschiedlichen Anbietern realisiert werden. Für den Anwender sollte es vollkommen transparent

sein, wer wo welchen Dienst ausführt und mit welchen Tools die Dienste erstellt wurden, sofern die Standardkonformität gegeben ist. Diese Zielsetzung der *„True Interoperability“* wollten wir (unter anderem) mithilfe von RichWPS etwas weitergehend umsetzen.

Eines der Hindernisse dabei ist die enorme Vielfalt an Datentypen, die in der Geoinformatik vorherrscht, zusammen mit der Tatsache dass die WPS-Spezifikation keinerlei Festlegungen dahingehend trifft, welche Datentypen in welchen Enkodierungen für Prozesseingaben und -ausgaben verwendet werden müssen oder können. Die verschiedenen WPS-Implementierungen unterscheiden sich daher beträchtlich, wenn es darum geht, welche I/O-Datenformate unterstützt werden. Daher wurde für den RichWPS-Server die Operation *GetSupportedTypes* implementiert, so dass immerhin der Server selber Auskunft darüber geben kann, welche Datentypen er unterstützt. Weiterhin wurde eine Adapterschicht für WPS-Clients zum Umgang mit verschiedenen Datentypen konzipiert und prototypisch implementiert, die wir weiter unten noch genauer erläutern.

3 Realisierung des RichWPS Servers

3.1 Kern-Implementierung

Der RichWPS-Server erlaubt das WPS-standardkonforme Anbieten und Ausführen von Arbeitsabläufen (WPS-Prozessmodellen) im Web. Er wurde realisiert als Modifikation des Open Source WPS Servers von 52°North[10]. Gegenüber einer regulären WPS 1.0 Implementierung verfügt der RichWPS-Server über eine transaktionale Komponente (WPS-T), um dem Modelbuilder eine Schnittstelle für die Veröffentlichung (Deployment) von Prozessen anzubieten. Hierzu wurde die Implementierung von [Schäffer, 2008] für die aktuelle Systemlandschaft wieder lauffähig gemacht.

Der Kern des RichWPS-Servers ist der ROLA-Interpreter, eine Laufzeitumgebung, die Arbeitsablaufbeschreibungen im Format ROLA (RichWPS Orchestration Language, siehe [Bensmann et al., 2014b]) zur Ausführung bringt. Dazu werden komplexe

[10] http://52north.org/communities/geoprocessing/wps/

Prozessbeschreibungen bis zu ihren ausführbaren Teilen heruntergebrochen, die – u.U. teilweise auf externen Servern liegenden – Basisdienste angestoßen, die Zwischenergebnisse von Teilschritt zu Teilschritt übergeben usw. Der Interpreter (Orchestration Engine) kann über eine WPS Client API programmatisch angesprochen werden. Im Rahmen des Forschungsprojekts wurde die ROLA ohne weiterführende Kontrollstrukturen realisiert, d.h. zurzeit betrachten wir nur Sequenzen von Arbeitsschritten. Komplexere Strukturen wie bedingte Ausführung, Parallelausführungen, Schleifen o.ä. würden aber keine prinzipiellen Probleme darstellen.

Vor der Veröffentlichung eines Prozessmodells kann der RichWPS-Server als Testplattform für die Funktionalität und die Einhaltung von Qualitätsmerkmalen (zurzeit nur Benchmarking von Prozessen und Teilprozessen) genutzt werden. Dafür wurde der Server dahingehend modifiziert, dass er ein „provisorisches Deploy“ zu Testzwecken umsetzt und bei Testläufen Zwischenergebnisse und Laufzeiten von Teilprozessen ermitteln und dem Modelbuilder übergeben kann.

Um mit der oben bereits angesprochenen Datenvielfalt in der Geoinformatik umzugehen, erforderte die Realisierung des RichWPS-Servers insbesondere auch die Implementierung diverser Konvertierungsfunktionen zwischen I/O-Datentypen und deren interner Repräsentation in Java für die Prozessausführung sowie die Implementierung eines Registers für Laufzeitvariablen (Datenübergabe zwischen Teilprozessen).

3.2 WPS-Präsentationsdirektiven

Die WPS 1.0.0 Spezifikation beschreibt das Auffinden und Ausführen von Prozessen, nicht jedoch konkrete Ergebnis-Datentypen oder deren clientseitige Interpretation und Darstellung. Es wird lediglich zwischen den Ein- und Ausgabearten 'Literal', 'BoundingBox' und 'Complex' unterschieden. Praktisch sind dem Inhalt von 'Complex'-Arten und somit Art und Umfang von Eingabe- und Ausgabedaten keine Grenzen gesetzt. Dieser generische Charakter der WPS-Spezifikation macht es für WPS-Clients unmöglich, auf alle Arten von WPS-I/O Parametern vorbereitet zu sein. An dieser Stelle setzt das WPS-PD Schema an [Hofmann et al, 2014]. WPS-PD liefert einen Beitrag zur *Konkretisierung von einigen wenigen Ausgabetypen*, ohne die mögliche Vielfalt der Datentypen zu

beeinträchtigen. Die im PD-Schema beschriebenen Elemente sind komplexe Ausgabeparameter, die von einem WPS-Client speziell interpretiert werden können, damit sie für den Endanwender – zusammen mit einer client-seitigen Karten-Anwendung – einen Mehrwert aus WPS-Prozessen darstellen. Der Anwender erhält nicht nur ein bestimmtes Prozessergebnis, sondern auch eine *angemessene Präsentation* desselben bzw. je nach PD-Element auch eine entsprechende Aktion.

Das PD-Schema beschreibt konkrete Ausgabe-Elemente für WPS-Prozesse sowie Richtlinien für die Präsentation und Handhabung dieser Elemente in einem WPS-Client. Die PD-Elemente repräsentieren sowohl eigenständig als auch in einer Gruppe bzw. in definierten Kombinationen von mehreren einzelnen Elementen sinnvolle Konkretisierungen von Prozessergebnissen. Das WPS-PD Schema ist in Form einer XML Schema Definition (XSD) beschrieben. Es basiert auf den Schema-Beschreibungen von XLink [11], OGC OWS 1.1.0[12] und OGC GML 3.1.1[13] und ist in seiner aktuellen Version unter http://schemas.disy.net/wps-pd/1.0.0/wps-pd.xsd verfügbar. In der Form, wie es in Disy Legato 1.4.0 (Juli 2012) integriert ist, beschreibt das Schema die folgenden Elemente, wobei mit WPS-Client ein Client gemeint ist, der das WPS-PD Schema unterstützt und interpretieren kann:

pd:Link
Ein pd:Link ist ein Element, welches eine URL zu einer Web-Ressource beinhaltet. Außerdem kann gemäß der XLink-Spezifikation für ein pd:Link Element optional festgelegt werden, in welcher Form die enthaltene URL geöffnet werden soll (z.B. in einem bestehenden oder einem neuen Browser-Fenster). Ebenso optional ist die Angabe eines MimeTypes der Web-Ressource (z. B. text/html für Webseiten, image/jpeg für JPEG-Bilder oder application/pdf für PDF-Dokumente). Erhält ein WPS-Client ein in das WPS-Response-Dokument eingebettetes Link-Element, öffnet er die Webressource und zeigt sie auf die entsprechende Art in einem Browserfenster an.
pd:Message

[11] http://www.edition-w3.de/TR/2001/REC-xlink-20010627
[12] http://schemas.opengis.net/ows/1.1.0
[13] http://schemas.opengis.net/gml/3.1.1/base

Eine pd:Message ist ein Element zum Übermitteln und Anzeigen von Nachrichten im WPS-Client. Diese Nachrichten werden üblicherweise in einem Popup-Fenster dargestellt. Optional kann die Art der pd:Message festgelegt werden, was zum Beispiel Auswirkungen auf die Darstellung der pd:Message am Client hat. Die Art der pd:Message kann zum Beispiel eine Info oder eine Warnung sein. Der Inhalt der Nachricht wird entweder in einem pd:Content Element oder einem pd:Link Element bereitgestellt. Eine pd:Message, die einen pd:Link enthält, ermöglicht es z. B., dass eine Nachricht aus einem referenzierten Bild oder einer ganzen HTML-Seite besteht.

pd:Viewport

Ein pd:Viewport ist ein Element, welches eine Bounding Box enthält, die durch zwei Koordinaten repräsentiert wird. Diese Bounding Box wird konkret anhand eines gml:Envelope-Elements realisiert. Innerhalb dieses Elements sind die beiden Koordinaten und das zugehörige Koordinatenreferenzsystem angegeben. Ein Client interpretiert ein pd:Viewport als Aktion, die das aktuelle Kartenfenster auf den im pd:Viewport angegebenen gml:Envelope zentriert.

pd:Marker

Ein pd:Marker ist ein Element, welches eine Markierung auf einer Karte repräsentiert. Es enthält immer das Element pd:Position, in dem die Koordinatenwerte und das Koordinatenreferenzsystem für den Marker angegeben sind. Optional enthält ein pd:Marker ein pd:Message-Element, welches der Markierung eine Nachricht verleiht. Erhält ein Client ein pd:Marker-Element aus einem WPS-Prozess, erzeugt er eine neue punktförmige Markierung auf der Karte und versieht diese gegebenenfalls bei der Selektion mit einer Nachricht in Form eines kleinen Popup-Fensters.

pd:Geometry

Eine pd:Geometry repräsentiert ein einfaches Geometrie"=Element. Eine einfache Geometrie kann entweder ein gml:Point, ein gml:LineString oder ein gml:Polygon sein. Der Aufbau dieser Geometrie-Elemente wird im GML-Schema definiert. Ein Client, der ein pd:Geometry-Element aus einem WPS-Prozess erhält, erzeugt die enthaltene Vektorgeometrie und stellt sie auf einem Layer seiner Karte dar.

pd:Group

Ein pd:Group-Element stellt die Kombination bzw. Ansammlung von beliebigen anderen PD-Elementen dar. Zum Beispiel kann eine pd:Group ein pd:Link, ein pd:Viewport und ein pd:Marker Element gleichzeitig enthalten. Vor allem aus Kombinationen mit einem

pd:Viewport-Element kann ein großer praktischer Nutzen in der Kartenanwendung des Clients gezogen werden: Ein Ergebnis vom Typ pd:Geometry wird nicht nur dargestellt, sondern anhand des pd:Viewports auch gleich im Kartenfenster zentriert. Ein Client interpretiert die einzelnen Elemente einer pd:Group nacheinander.
pd:StyledFeatureCollection
Eine pd:StyledFeatureCollection ist ein Element zur Darstellung von GML-FeatureCollections anhand einer integrierten Stilbeschreibung. Ein pd:StyledFeatureCollection-Element muss zwingend eine FeatureType-Beschreibung in Form eines XSD-Dokuments enthalten. Passend zur FeatureType-Beschreibung enthält ein pd:StyledFeautureCollection-Element ein pd:Features-Element, welches für die darin enthaltene gml:FeatureCollection das Koordinaten-referenzsystem angibt. Optional enthält die pd:StyledFeatureCollection ein se:FeatureTypeStyle, welches durch den OGC „Symbology Encoding“-Standard definiert wird und Regeln zur visuellen Darstellung von Features enthält. Ebenso optional ist die Angabe eines pd:Hints-Elements, welches zusätzliche Informationen zur Visualisierung der pd:StyledFeatureCollection enthält. Beispielsweise können verkürzte Attributbezeichnungen, die im se:FeatureTypeStyle Element zur Beschriftung von Feature verwendet werden, durch lesbare Ausdrücke ersetzt werden. Ein Client, der ein pd:StyledFeatureCollection-Element aus einem WPS-Prozess erhält, erzeugt die darin enthaltene FeatureCollection auf einem Layer seiner Kartendarstellung und visualisiert diese anhand der enthaltenen Stilvorschriften.

Der RichWPS-Server kann aus WPS-Prozessen beliebige WPS-PD Elemente erzeugen. Hierzu wurden WPS-PD-spezifische IData-Elemente zur Integration in die 52°North WPS Dateninfrastruktur realisiert und WPS-PD spezifische Generatoren implementiert. Abbildung 4 zeigt das Prozessergebnis eines Beispielprozesses, der einen pd:Marker erzeugt. Obwohl die Direktiven einfach sind und sich dadurch ohne großen Aufwand implementieren lassen, können aufgrund der Kombinationsmöglichkeiten mächtige Präsentationsszenarien damit umgesetzt werden.

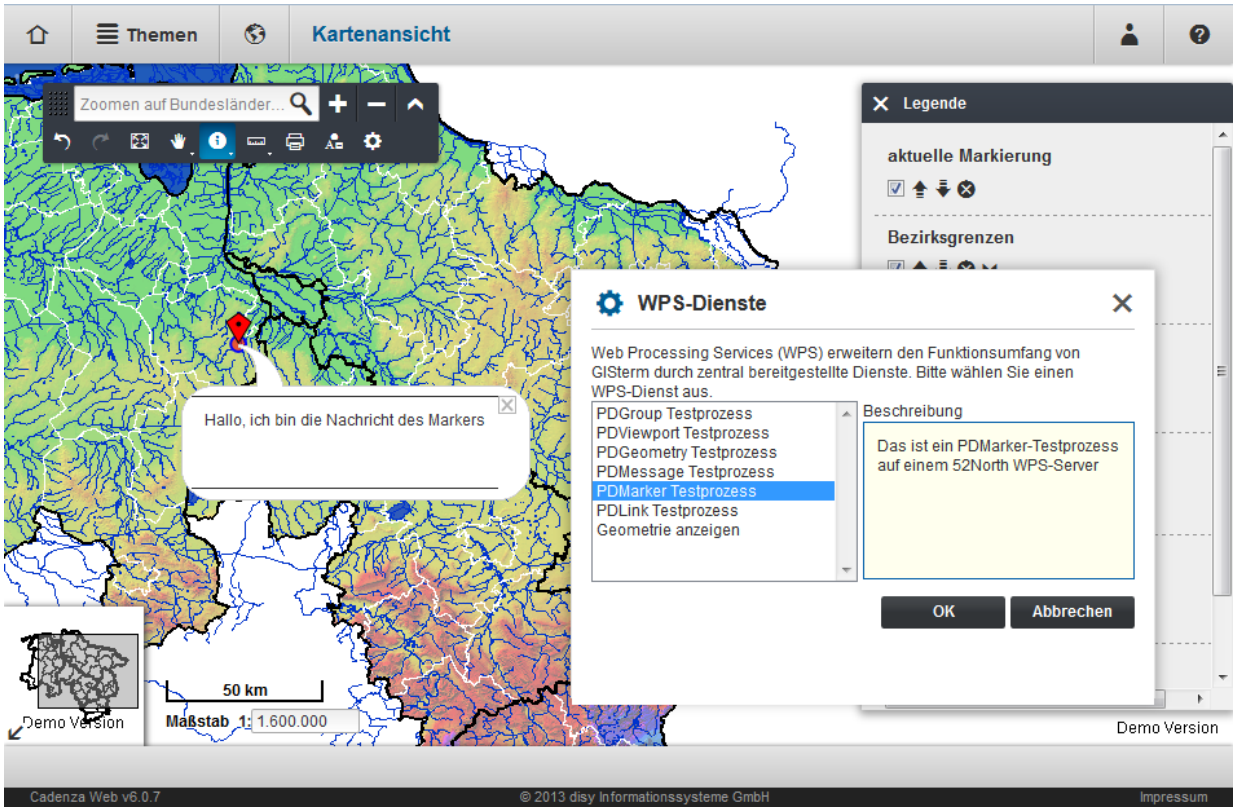


Abbildung 4: Cadenza Web als RichWPS-Client zeigt das Ergebnis eines WPS-PD Testprozesses auf dem 52°North RichWPS-Server an

Die entsprechenden Entwicklungen wurden unter der GPL v2 Lizenz auf Github veröffentlicht.[14]

4 Client-seitige Entwicklungen

Neben der Weiterentwicklung des Disy WPS-Clients, so dass er auch die aktuelle Version der WPS-PD Ergebnisse adäquat darstellen kann, ist eine der wichtigeren client-seitigen Entwicklungen in RichWPS eine Adapterschicht, die auch auf Seiten des Clients helfen soll, mit der bereits angesprochenen Datenvielfalt umzugehen (vgl. Abbildung 5, [Scholtyssek, 2015]).

[14] https://github.com/romanwoessner/wps-pd-52n

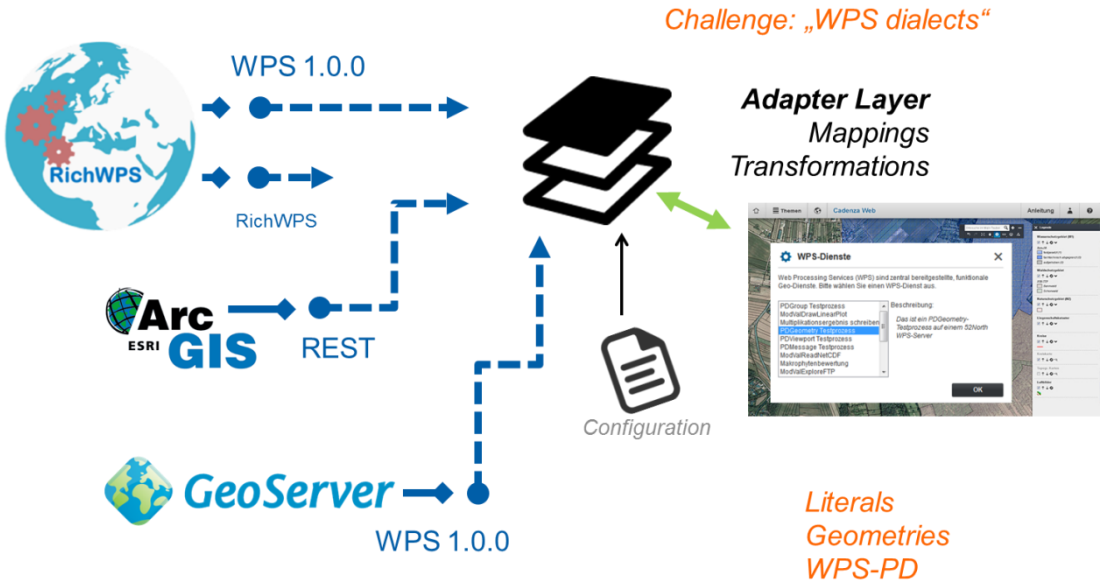


Abbildung 5: Illustration zur Client-seitigen Adapterschicht

Diese konfigurierbare und erweiterbare Software übersetzt zwischen den Datenformaten des RichWPS-Servers, des GeoServers und der ArcGIS Geoprocessing REST Schnittstelle, um client-seitig transparent zwischen WPS-Diensten dieser unterschiedlichen Anbieter wechseln zu können. Die Adapterschicht ist so konzipiert, dass weitere Datentypen bzw. WPS-Implementierungen verhältnismäßig einfach hinzugefügt werden können.

Weitere laufende Arbeiten befassen sich damit, für mobile Clients die Unwägbarkeiten unzuverlässiger Netzverbindungen durch einen intelligenten Proxy auszugleichen [Barth, 2015].

5 Zusammenfassung

In diesem Artikel haben wir den RichWPS-Server mit seinen Funktionen und einigen Implementierungsdetails erläutert. Der RichWPS-Server erweitert den 52°North Open Source Server um transaktionales WPS, Auskunftsfähigkeit und Konvertierungsfunktionen hinsichtlich verwendeter Datentypen sowie Unterstützung beim Debugging und Profiling komponierter Prozesse. Den Kern des Servers stellt der Interpreter für komponierte Workflows in Form von ROLA-Skripten dar.

Wir haben das Prinzip der WPS-Präsentationsdirektiven (WPS-PD) erklärt, die helfen, WPS-Ergebnisse verschiedenen auf der GUI adäquat darzustellen. WPS-PD kann als Ausgabeformat vom RichWPS-Server erzeugt und vom Disy WPS-Client verarbeitet werden. Außerdem haben wir das Problem der Datenvielfalt in WPS angesprochen, das Anlass war, eine erweiterbare client-seitige Adapterschicht für WPS-Daten verschiedener Werkzeug-Suiten zu bauen.

Es sei ausdrücklich darauf hingewiesen, dass RichWPS, nicht *„yet another WPS workbench“* oder *„yet another WPS orchestration approach“* sein soll – den Autoren ist klar, dass z.B. mit deegree und anderen bereits hervoragende Werkzeuge und akademische Prototypen zur Verfügung stehen. Dennoch zeigt die u.E. noch erstaunlich geringe Verbreitung von WPS in der Praxis (WPS 1.0.0 wurde bereits 2007 veröffentlicht), dass es organisatorische oder technische Barrieren zur Nutzung geben muss. Organisatorische Hürden (z.B. die Frage der Leistungsverrechnung, wenn Behörde A Dienste von Behörde B nutzt) lassen sich im Rahmen eines IT-Forschungsprojekts nur am Rande behandeln. Aber auch technisch haben wir klare Verbesserungspotenziale identifiziert, die wir im Projekt teilweise angehen konnten. Die Einfachheit und Benutzerfreundlichkeit waren zentrale Ziele beim Orchestrierungs- und Modellierungsansatz. Der Umgang mit der unbegrenzten Datenvielfalt war eines der zentralen Themen bei Server- und Client-Software. Viele der Projektentwicklungen (WPS-PD, Adapterschicht) sollen die existierenden Lösungen nicht ersetzen, sondern können sie ergänzen.

Die meisten RichWPS-Entwicklungen stehen als Open Source auf GITHUB zur Verfügung.[15]

Die Werkzeuge in RichWPS wurden in enger Zusammenarbeit mit Pilotanwendern des Landesbetriebs für Küstenschutz, Nationalpark und Meeresschutz Schleswig-Holstein (LKN, Tönning) und der Bundesanstalt für Wasserbau (BAW, Hamburg) erstellt und erprobt (vgl. auch [Ziegenhagen et al., 2014]). Die ursprüngliche Zielsetzung im Projekt, eine vollständig fachanwender-taugliche Werkzeugumgebung zu schaffen, die

[15] https://richwps.github.io/

eine umfassende WPS-Verwendung ohne vertiefte IT-Kenntnisse ermöglicht, war wahrscheinlich unrealistisch. Trotzdem gehen wir davon aus, dass der RichWPS-Ansatz zu einer deutlichen Verminderung des erforderlichen Programmieraufwands und der erforderlichen IT-Kenntnisse zum Aufsetzen und Nutzen WPS-basierter, verteilter Geodaten-Prozessierungsdienste führen kann und dass er es erheblich vereinfacht, Code-Wiederverwendung und Dienste-Kooperation in komplexen GDI-Szenarien zu erreichen.

Zum Zeitpunkt der Veröffentlichung dieses Artikels steht WPS2.0 „vor der Tür“. Nach unseren Informationen wird die neue Version des Standards einige neue und bessere Konzepte bringen, aber die in RichWPS angegangenen Probleme u.E. nicht tangieren.

Danksagung: *Die vorgestellten Ergebnisse wurden teilweise mit finanzieller Unterstützung des Bundesministeriums für Bildung und Forschung (BMBF) im KMU-innovativ Projekt RichWPS („Eine Software-Umgebung für Fachanwender zur effizienteren Nutzung von Geodaten mit Web Processing Services“, FKZ 01IS11029) realisiert. Projektergebnisse wurden im Abschluss-Workshop ausgiebig diskutiert: http://www.disy.net/nc/aktuelles/newsartikel/artikel/2942.html*

6 Literaturverzeichnis

[Barth, 2015]

Barth, Sebastian: Konzeption und Implementierung eines mobilen Clients für RichWPS-Server, Master Thesis, Hochschule Karlsruhe, 2015.

[Bensmann et al, 2013]

Bensmann, Felix; Alcacer-Labrador, Dorian; Wössner, Roman: Grobspezifikation - RichWPS Projektergebnis E1.1. Interne Dokumentation, RichWPS-Projektkonsortium, Dezember 2013.

[Bensmann et al, 2014]

Bensmann, Felix; Roosmann, Rainer; Kohlus, Jörn: Aus Klein mach Groß: Komposition von OGC Web Services mit dem RichWPS ModelBuilder, Workshop AK UIS, Karlsruhe, 2014 . *In diesem Band.*

[Bensmann et al, 2014b]

Bensmann, Felix; Alcacer-Labrador, Dorian; Ziegenhagen, Dennis; Roosmann, Rainer: The RichWPS Environment for Orchestration. *ISPRS International Journal of Geo-Information,* 2014.

[Hofmann et al, 2014]

Hofmann, Claus et al: GDI-Dienste UIS BW - WPS-Dienste im Umweltinformationssystem Baden-Württemberg für die Geodateninfrastruktur Baden-Württemberg, in: R. Mayer-Föll, Roland; Ebel, Renate, Geiger, Werner (Hrsg.): F+E-Vorhaben KEWA – Kooperative Entwicklung wirtschaftlicher Anwendungen für Umwelt, Verkehr und benachbarte Bereiche in neuen Verwaltungsstrukturen Phase V 2009/10, Karlsruhe: KIT Scientific Reports No. 7544, 2014.

[Schäffer, 2008]

Schäffer, Bastian: Towards a Transactional Web Processing Service (WPS-T). In: Pebesma, E.; Bishr, M.; Bartoschek, T. (Hrsg.): GI-Days 2008 – Proc. 6th Geographic Information Days. IfGIprints Nr. 32, Institut für Geoinformatik, Münster, 2008.

[Scholtyssek, 2015]

Scholtyssek, Felix: Generalisierung eines integrierten WPS-Clients hin zu einer größeren Interoperabilität mit verschiedenen Serverkomponenten zur Geoprozessierung," Master Thesis, Hochschule Karlsruhe, April 2015.

[Ziegenhagen et al, 2014]

Ziegenhagen, Dennis; Kohlus, Jörn; Roosmann, Rainer: Orchestration of Geospatial Processes with RichWPS – A Practical Demonstration. In: EnviroInfo 2014: Proc. 28th Int. Conf. on Informatics for Environmental Protection, S. 525-532, 2014.

Cloud Computing für die Kalibrierung von Hochwassersimulationen

Marco Brettschneider, HTW Berlin, Marco.Brettschneider@htw-berlin.de
Bernd Pfützner, BAH Berlin, Bernd.Pfuetzner@bah-berlin.de
Frank Fuchs-Kittowski, HTW Berlin, Frank.Fuchs-Kittowski@htw-berlin.de

Abstract

Well calibrated rainfall runoff models of sufficient complexity are a necessary basis for accurate flood water simulations in order to prevent losses. Using the hydrologic modeling system ArcEGMO in a small catchment as an example, IaaS cloud computing has be proven fast and cost-efficient for short-term usage. On the other hand, IaaS cloud instances cannot replace local computers, since long-term usage is still expensive and the performance of the offered architectures can slow down a calibration process on the long run. Ultimately, cloud computing has the potential to be a useful extension of local computer capacities if state of the art computer hardware with a fair pricing is offered by IaaS providers.

1 Einleitung

Überschwemmungen und Sturzfluten führen immer wieder und in zunehmender Anzahl zu großen Schäden in der Nähe von Fließgewässern [Bates u.a. 2008]. Im deutschsprachigen Raum haben beispielsweise die Elbe-Hochwasser von 2002 und 2013 verheerdende Schäden in Millionenhöhe angerichtet [GDV 2013]. Der Simulation von Hochwasser, einmal zur Ermittlung von Bemessungsgrößen für die Planung von Hochwasserschutzmaßnahmen, aber auch zur Vorhersage während eines Hochwasserereignisses, kommt dementsprechend eine hohe Bedeutung in der Abschätzung möglicher Risiken im Vorfeld und zum Zeitpunkt von Hochwasserwellen zu, um rechtzeitig

Maßnahmen zur Vorsorge ergreifen zu können [Bornebusch 2006]. Die Simulation selbst lässt sich allgemein in einen hydrologischen Anteil zur Abbildung der Abflussprozesse im Gebiet und einen hydraulischen Anteil zur Abbildung der eigentlichen Hochwasserausprägung untergliedern, in urbanen Regionen kommt häufig noch ein Anteil zur Stadtentwässerung hinzu [Maßmann u.a. 2010]. In dieser Studie steht der hydrologische Aspekt von Hochwasser im Vordergrund, wofür generell Niederschlag-Abfluss-Modelle (NA-Modelle) verwendet werden.

Je nach Modellkomplexität kann die Kalibrierung eines NA-Modells sehr aufwändig sein. Robuste Werte für die Modellparameter sind häufig nur anhand einer Vielzahl von Modellläufen zu ermitteln, und gerade bei großflächigen, hochaufgelösten Modellen stößt man damit schnell an die Grenzen der Rechenkapazität eines Computers. Möchte man dennoch mehrere Parametrisierungen testen, ist man gezwungen, über alternative Wege der Kalibrierung nachzudenken.

Die zur Verfügung stehende Rechentechnik ändert sich ständig, und es gibt mittlerweile verschiedene Wege, mehr Rechenkapazitäten für eine noch detailliertere Beschreibung hydrologischer Prozesse einzubeziehen. Zu den neuen Technologien gehören unter anderem die Multicore- und Manycore-Architekturen aktueller Prozessoren und Grafikchips [z.B. Rouholahnejad u.a. 2012] und nicht-lokale Rechenkapazitäten, welche die lokalen erweitern können. Zu den letzteren zählt das Cloud Computing. Hierbei lagert man seine Rechnungen auf virtuelle Computer aus, welche man bei Bedarf von Serverfarmen eines Cloud-Dienstleisters generieren kann. Das besondere Potenzial der Cloud liegt in der Skalierung der Ressourcen: für jeden Anwendungsfall kann eine entsprechende Konfiguration von Rechenleistung, Arbeitsspeicher, Festplattenspeicher und Betriebssystem gefunden werden, auch eine Kopplung mehrere virtueller Computer ist möglich.

Cloud Computing befindet sich aber noch in einer frühen Phase. Abgesicherte Kenntnisse zum Einsatz von Cloud Computing in der Hydrologie und insbesondere zur Simulation von Hochwasser liegen bisher nicht vor. Es existieren nur wenige Berichte zur Verwendung von Cloud Computing in der Hydrologie, wobei auch andere Aspekte

wie beispielsweise Unsicherheitsanalysen im Vordergrund stehen und bereits fertig kalibrierte NA-Modelle genutzt werden [z.B. Moya Quiroga u.a. 2012].

Weil Cloud-Projekte schnell und kostengünstig umgesetzt werden können, bietet es sich an, in kleinen Pilotprojekten eigene Erfahrungen zu sammeln [BITKOM 2012]. In diesem Beitrag soll die Auslagerung von Modellrechnungen zur Kalibrierung eines NA-Modells zur Hochwassersimulation auf virtuelle Computer in der Cloud diskutiert werden. Hierbei steht die erzielbare Rechenleistung und die Skalierbarkeit der Ressourcen im Vordergrund. Ein Unternehmen, welches die Modelle erstellt und pflegt, ist auch immer bestrebt, die dabei entstehenden Kosten mit dem Nutzen abzugleichen, weshalb auch der Kostenaspekt in die Betrachtungen mit einfließen soll.

Der Beitrag gliedert sich wie folgt: In Kapitel 2 werden das hydrologische Modellsystem ArcEGMO als Basis für die Hochwassersimulationen vorgestellt, die Potenziale des Cloud Computing kurz skizziert, um darauf aufbauend die Anforderungen an eine Cloud-Simulation zu definieren. Kapitel 3 stellt die Methodik und das Setup der Untersuchung anhand der verwendeten Cloud-Dienste und des Beispielszenarios vor. In Kapitel 4 werden die Ergebnisse der Beispielanwendung des Modells auf virtuellen Rechnern präsentiert und damit gezeigt, inwieweit Cloud Computing den gestellten Anforderungen gerecht werden kann. Im abschließenden Kapitel 5 werden die Potenziale des Cloud Computing in Bezug auf Rechenleistung und Wirtschaftlichkeit im Rahmen komplexer hydrologischer Modellierung bewertet und ein Ausblick auf die weitere Entwicklung gegeben.

2 Anforderungen an eine Cloud Unterstützung für die Simulation von Hochwasser mit dem Modellsystem ArcEGMO

2.1 Niederschlags-Abfluss-Modellierung mit dem Modellsystem ArcEGMO

Das hydrologische Modellierungssystem ArcEGMO dient der zeitlich hochaufgelösten, physikalisch fundierten Simulation aller maßgeblichen Prozesse des Wasserhaushaltes und des Abflussregimes in Einzugsgebieten verschiedener Ausprägungen und Maßstabsbereiche [Pfützner 2002, Pfützner u.a. 2013]. Sein modularer Aufbau, die

variablen Möglichkeiten zur Flächengliederung und eine Reihe von Schnittstellen zur Kopplung mit anderen Modellen erlauben die Bearbeitung einer Vielzahl praxisrelevanter und wissenschaftlicher Fragestellungen (**Abb. 1**).

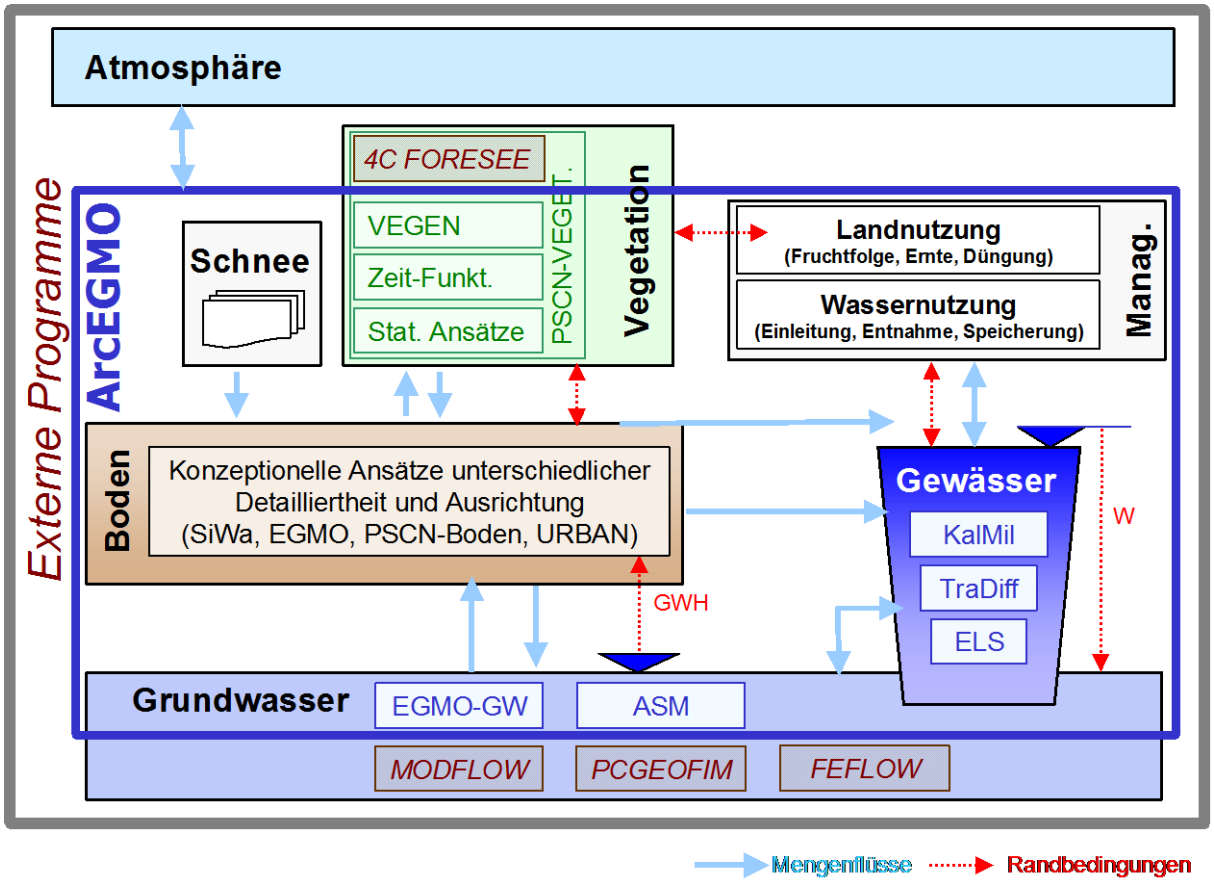


Abbildung 1: Domänen und Module von ArEGMO

ArcEGMO wurde bereits bei einer Vielzahl von Untersuchungen zur Hochwasserentstehung in unterschiedlichen Landschaftstypen und Gebietsgrößen erfolgreich eingesetzt [z.B. Pfützner & Hesse 2011, Büttner u.a. 2009]. Dabei wurde deutlich, dass der Parametrisierung des Modells eine entscheidende Bedeutung für die Abbildung der realen Verhältnisse und Dynamiken zukommt. Da die Kalibrierung des Modells anhand realer Messwerte ein iterativer Prozess ist, sind sehr viele Modellläufe notwendig, um die Wirkung verschiedener Parameterwerte für die Qualität des Modells beurteilen zu können. Gerade bei komplexen Aufgabenstellungen kann die Rechenzeit schnell zum begrenzenden Faktor für die Suche nach optimalen Parametern werden. Nachfolgend soll daher abgeschätzt werden, inwieweit Cloud Computing konkret als Möglichkeit zur

schnellen Akquirierung zusätzlicher Rechenkapazitäten hilfreich sein kann, um lokal begrenzte Rechenkapazitäten zu erweitern.

2.2 Potenziale des Cloud Computing

Cloud Computing ist eine Form der bedarfsgerechten und flexiblen Bereitstellung und Nutzung von IT-Ressourcen und IT-Leistungen. Dabei kann es sich um IT-Infrastruktur (Netzwerk, Server, Speicher), Plattformen oder Anwendungen aller Art handeln. Diese werden mit minimalen Verwaltungs-Aufwand in Echtzeit als Service über das Internet bereitgestellt und nach Verbrauch abgerechnet [BITKOM 2012].

Cloud Services werden allgemein in die drei Ebenen „Infrastructure as a Service (IaaS)“, „Platform as a Service (PaaS)“ sowie „Software as a Service (SaaS)“ eingeteilt. IaaS bietet ein Minimum an Infrastruktur zum Rechnen, zum Speichern von Daten und zur Netzwerkanbindung. Die Trennung zu PaaS ist unscharf, jedoch wird letzteres meist mit der Entwicklung und Bereitstellung von Webdiensten assoziiert. Mit SaaS wird komplette Software direkt zur online-Nutzung zur Verfügung gestellt [Tonninger, 2011]. Aufgrund des Fokus auf reine Rechenleistung wird in dieser Studie IaaS genutzt und nachfolgend kurz erläutert.

Bei IaaS werden IT-Ressourcen als virtuelle Instanzen einer kompletten Hardware- und Betriebssystemplattform bereitgestellt. Aus der Nutzerperspektive sieht IaaS wie physische Hardware aus: Die Dienst-Nutzer können das gesamte Betriebssystem sowie die darauf laufende Software kontrollieren, und es gibt keine Einschränkungen im Bezug auf Anwendungen, die von den Systemen gehostet werden können. Zur Verwaltung der Infrastruktur werden spezielle Webservices zur Verfügung gestellt, mit welchen man innerhalb nur weniger Minuten beliebig viele Instanzen (virtuelle Server) starten und wieder entfernen kann. Die Nachteile von IaaS liegen darin, dass eine automatische Skalierung der Recheninstanzen nur schwer umgesetzt werden kann, bei Nutzung über einen längeren Zeitraum hinweg die Kosten schnell ansteigen können und zumindest theoretisch die Möglichkeit eines Ausfalls einer Instanz besteht [Walterbusch & Teuteberg 2014].

Unternehmen investieren in Cloud Computing vor allem wegen des Potenzials zur Kostensenkung. Weitere Nutzen werden in der Verlagerung von Fixkosten (Investitionen etc.) zu variablen Kosten, in der Erhöhung der Flexibilität und Skalierbarkeit der IT-Ressourcen sowie in der verbrauchsabhängigen Abrechnung der IT-Leistungen gesehen [BITKOM 2012, S. 10].

2.3 Anforderungen an Cloud Computing für die Kalibrierung des Modellsystems ArcEGMO

Für die Simulation von Hochwasser zur Vorsorge von Schäden ist ein gut kalibriertes Niederschlags-Abfluss-Modell die Basis für verlässliche Vorhersagen. Der dabei notwendige Aufwand mit mehreren Modellläufen zur iterativen Suche robuster Parameterwerte kann durch Cloud Computing aufgefangen werden, wenn die Rechnungzeiten mit denen lokaler Computer vergleichbar oder kürzer sind, und die sich dabei ergebenden Kosten gering bleiben. Die Kosten bei Bezahlmodellen ohne langfristige Buchung hängen zum einen von der Dauer zwischen Generieren und Entfernen einer Instanz ab sowie von zusätzlichen Gebühren für Datenspeicherung, Datensicherung und Kommunikation mit oder zwischen den Instanzen [Deelman u.a. 2008, Berendes u.a. 2013]. Die eben aufgeführten zusätzlichen Dienste werden in dieser Studie außer acht gelassen, da die zu transferierenden Ein- und Ausgabedaten und die Modellsoftware von ihrer Größe und somit von den Kosten her vernachlässigbar sind, ebenso wie der Informationsverlust und die zu erwartetenden zusätzlichen Kosten durch plötzlichen Ausfall einer Instanz im Rahmen einer Kalibrierung. Die Rechenzeiten sind daher in dieser Studie das bestimmende Kriterium zur Beurteilung des Nutzens von Cloud Computing zur Kalibrierung von NA-Modellen. Der Zeitaufwand durch Verzögerungen beim Datentransfer spielt ebenfalls eine Rolle, diese sind jedoch meist klein gegenüber den Rechenzeiten, besonders bei einer Kalibrierung mit vielen Raumelementen.

Die Cloudsparten von Amazon und Microsoft bieten Konfigurationen, welche vergleichbar zu normalen Desktopcomputern sind, jedoch auch Instanztypen, welche bereits modernen Serverrechnern entsprechen [Amazon 2014a, Amazon 2014b, Microsoft 2014]. Die Dienste Amazon EC2 und Microsoft Windows Azure werden daher

für die weiteren Betrachtungen verwendet. Neben den Rechenzeiten und den Kosten wird auch der Frage nachgegangen, inwieweit sich die Rechenleistung je Anbieter für die Modellierung mit ArcEGMO steigern lässt, und wie sich die Kosten in diesem Fall entwickeln.

3 Konfiguration der IaaS-Dienste zur Kalibrierung von Hochwassersimulationen

3.1 Beispielszenario

Um allgemein Kosten und Nutzen von Cloud Computing einander gegenüberzustellen, stehen bereits viele verschiedene Ansätze zur Verfügung, welche meist auf die Prüfung der Wirtschaftlichkeit im Maßstab größerer Unternehmen und staatlicher Einrichtungen ausgerichtet sind, und beispielsweise auf heuristischen Methoden oder Break-Even-Analysen beruhen [z.B. Berendes u.a. 2013, Armbrust u.a. 2010, Matros u.a. 2009]. Der hier eingeschlagene Weg ist ebenfalls heuristisch, jedoch steht die Leistungsfähigkeit der verschiedenen Instanztypen dabei im Vordergrund. Im Folgenden soll anhand eines realistischen Szenarios gezeigt werden, wie leistungsfähig die virtuellen Computer im Hinblick auf NA-Modellrechnungen mit ArcEGMO sind, bei gleichzeitiger Betrachtung der jeweils entstehenden Kosten. Die Beschreibung des Szenarios im Hinblick auf die Konfiguration des Modells und der IaaS-Dienste sind die Themen dieses Kapitels.

3.2 Setup des Modellsystems ArcEGMO

Gerechnet wird ein reales Beispielgebiet, welches auch zu Schulungszwecken genutzt wird [Pfützner et al. 2013]. Dabei handelt es sich um ein Einzugsgebiet im Tiefland mit 59km² Fläche aufgeteilt in 362 Raumelemente (Hydrotope). Gerechnet wird ein Zeitraum von 5,5 Jahren in Tagesauflösung. Modelleingangsdaten sind verschiedene hydrologische und meteorologische Zeitreihen für Abfluss, Niederschlag und Klima-Größen. Erstere werden bei der Kalibrierung als Zielgröße genutzt.

Nachdem ArcEGMO anhand seiner Steuerdateien vorbereitet wurde, kann eine Simulation gestartet werden. Die Initialisierungsphase eines Modelllaufs wird von

Dateilese- und –schreibprozessen auf der Festplatte dominiert, während die Simulation durch Zugriffe auf den Arbeitspeicher gekennzeichnet ist. Aus diesem Grund wird bei den nachfolgenden Betrachtungen eine Aufteilung bezüglich der verbrauchten Rechenzeit vorgenommen.

ArcEGMO arbeitet seriell, d.h. je Simulation wird nur ein Prozessor-Kern beansprucht. Die Größe des genutzten Arbeitsspeichers richtet sich nach der Anzahl der Raumelemente, welche im Rahmen der Simulation bearbeitet werden müssen. In diesem Beispielszenario ist die Anzahl gering, weswegen auch die Rechenzeit relativ gering ist. Aufgrund der seriellen Arbeitsweise kann hier nur die Skalierbarkeit anhand der einzelnen Prozessor- und Festplattenarchitekturen und der dabei erzielbare Leistungsgewinn in Betracht gezogen werden. Da aktuell nur Amazon mehrere Rechen- und Speicherarchitekturen zur Verfügung stellt, bezieht sich die Betrachtung zur Skalierbarkeit nur auf EC2.

3.3 Vorstellung der genutzten IaaS-Dienste

Windows Azure besitzt Eigenschaften sowohl von PaaS und IaaS und deckt damit eine hohe Bandbreite an Möglichkeiten zur Entwicklung und Betreibung von eigener Software ab, mit Fokus auf Web-Anwendungen. Die Amazon Web Services (AWS) konzentrieren sich auf IaaS und setzen den Fokus auf Flexibilität und Kombinierbarkeit der bereitgestellten Ressourcen für jegliche Art von Gebrauch. Dazu gibt es ein großes Angebot an Betriebsystemen, welche bei Bedarf auch schon weitere, benötigte Software und Entwicklungsumgebungen vorinstalliert haben.

Innerhalb der AWS ist EC2 der Dienst zur Bereitstellung virtueller Rechenkapazitäten. Die Abrechnung erfolgt stundengenau, zusätzliche Festplattenkapazitäten können per Aufpreis dazugebucht werden [Amazon, 2014a]. Unter Windows Azure werden virtuelle Rechner über die Azure Benutzeroberfläche gestartet und überwacht. Die gewählte Leistung der Instanz und die Nutzung von Zusatzdiensten oder zusätzlichen Speicherplatzkapazitäten bestimmen auch hier den finalen Stundenpreis [Microsoft, 2014].

3.4 Wahl der Instanzkonfigurationen

Für beide betrachteten Cloud-Anbieter wurden Konfigurationen gesucht, die der Hardware eines lokalen Vergleichsrechners entsprachen. Eine exakte Übereinstimmung und Vergleichbarkeit zwischen allen drei Konfigurationen ist nicht vollständig herzustellen, da allein schon die beiden Cloud-Dienste für ihre virtuellen Rechner auf unterschiedliche Hardware-Anbieter aufsetzen. Der Vergleich kann daher grundsätzlich nur näherungsweise durchgeführt werden.

Der lokale Computer besitzt einen E7400-DualCore-Prozessor von Intel mit 64bit-Architektur und 2,8GHz Taktfrequenz je Kern, 3 GB Arbeitsspeicher, und HDD-Festplatte. Bei Amazon wurde die m1.large-Konfiguration als entsprechende Instanz gewählt, mit 2 Kernen eines Intel Xeon-Prozessors (ohne Angabe der Taktfrequenz) und 7,5 GB Arbeitsspeicher [Amazon 2014b], bei Microsoft kam unter den aktuellen Wahlmöglichkeiten die A2-Konfiguration mit AMD Opteron 2377 EE Prozessor (2 x 1,6-GHz-Kerne) und insgesamt 3,5 GB Arbeitsspeicher den lokalen Gegebenheiten am Nächsten [Microsoft 2014, Hochstätter 2011]. **Tabelle 1** fasst alle Konfigurationen nochmal zusammen.

Tabelle 1: Lokaler Rechner und gewählte Cloud-Instanztypen im Vergleich

Amazon EC2 m1.large	Windows Azure A2	Lokal
-Intel Xeon Prozessor, 2 Kernen gebucht, 64bit	- AMD Opteron Prozessor, 2 Kerne (jeweils 1,6 GHz) gebucht, 64bit	- Intel E7400 DualCore mit 2 Kernen à 2,8 GHz, 64bit
- 7 GB RAM	- 3,5 GB RAM	- 3 GB RAM
- Windows Server 2012	- Windows Server 2012	- Windows 7

Tabelle 2 zeigt einen Preisvergleich der Anbieter für diejenigen Preismodelle, welche keine Reservierung bzw. Abonnements voraussetzen. Die Kosten pro Stunde sind die Basiskosten und aufgrund der stundengenauen Abrechnung fix, die Kosten pro 30 Tage beziehen sich auf eine durchgehend aktive Instanz. Werden noch Dienste wie zusätzlicher Speicher, Datenbanken, Netzwerkinfrastruktur etc. hinzu gebucht, entstehen

weitere Kosten. Ein lokaler Rechner wurde in dieser Tabelle nicht aufgeführt, da hier variable und fixe Kosten gemeinsam betrachtet werden müssen, und ein Vergleich somit schwierig ist.

Tabelle 2: Kosten für die Cloud-Computing-Dienste von Amazon und Microsoft im Vergleich für die Preismodelle ‚on demand' bzw. ‚pay-as-you-go' (Quelle: Amazon 2014a, Microsoft 2014)

	Amazon EC2 m1.large on demand	Windows Azure A2 pay-as-you-go
Kosten [Euro pro Stunde]	0.26	0.135
Kosten [Euro pro 30 Tage]	190.66	97.20

In dieser Studie wurden nur die Preismodelle „on demand" bzw. „pay-as-you-go" beachtet. Diese beziehen sich auf eine spontane Bereitstellung von Instanzen, ohne diese vorher reserviert zu haben.

Um die mögliche Leistungsteigerung mit verbesserten Cloud-Ressourcen zu untersuchen, wurde bei Amazon die neuste Generation an Instanzen zum Vergleich hinzugezogen. Die m3-Instanzen nutzen Intel Xeon-E5-2670-Sandy-Bridge-Prozessoren und stellen SSD-Festplatten bereit, die c3-Instanzen ebenfalls, jedoch arbeiten diese mit Xeon-E5-2680v2-Ivy-Bridge-Prozessoren [Amazon 2014b]. Für beide Typen wurden ebenfalls wieder Architekturen mit 2 Kernen akquiriert.

Bei Microsoft gibt es derzeit nur eine Generation an Instanzen, wodurch sich die genutzten Prozessoren und Festplattenkapazitäten von ihrer Leistung her nicht unterscheiden sollten, was in Rahmen dieser Studie durchgeführte, jedoch hier nicht weiter ausgewertete Tests belegt haben.

4 Ergebnisse der Untersuchung von IaaS-Diensten zur Kalibrierung von Hochwassersimulationen

4.1 Vergleich aller genutzten Konfigurationen

Bei Microsoft und Amazon lassen sich Hardware-Konfigurationen finden, welche denen lokaler Computer ähnlich sind, auch wenn ein direkter Vergleich nicht möglich ist

aufgrund der Unterschiede bei der zugrundeliegenden Hardware zwischen beiden Cloud Computing Diensten. Amazon bietet darüber hinaus aktuelle Hardware, welche sich auf dem Niveau von Servercomputern der neuesten Generation bewegt. Von den Kosten her ist dennoch die Anschaffung eines lokalen Geräts günstiger. Aktuell kostet ein Computer, wie er hier genutzt und vorgestellt wurde, ca. 600 Euro [INNOVA 2014]. Auch bei Beachtung weiterer variabler Kosten, wie z.B. Strom lässt sich mit Blick auf die monatlichen Kosten bei voller Auslastung von ca. 200 Euro bei Amazon und 100 Euro bei Microsoft (**Tab. 2**) bereits sagen, dass der Nutzen von Cloud Computing bisher nur in Fällen kurzen und kurzfristigen Bedarfs zum Tragen kommt. Wie man leicht einsehen kann, wäre der lokale Rechner nach 3 Monaten bzw. 6 Monaten voller Auslastung von der Anschaffung her refinanziert. Auf Preismodelle mit Reservierung bzw. Abonnement wird hier nicht weiter eigegangen, da in diesem Fall die Anmietung von Serverkapazitäten günstiger wäre [z.B. Hetzner 2014].

4.2 Ergebnisse für die Modellierung mit ArcEGMO

Abb. 2 zeigt die erzielten Rechenzeiten je Anbieter im Überblick und im Vergleich zum lokalen Computer. Dabei fällt auf, dass die Gesamtrechenzeiten der Virtuellen Rechner über der des lokalen Geräts liegen und diese Unterschiede zum Großteil während der Simulation entstehen. Daraus lässt sich schließen, dass die reine Rechenleistung von technisch vergleichbaren Cloud-Instanzen geringer ist als jene eines entsprechenden lokalen Computers. Die Phase der Lese- und Schreibprozesse auf der Festplatte während der Modellinitialisierung ist gekennzeichnet durch einen sehr ähnlichen Zeitaufwand.

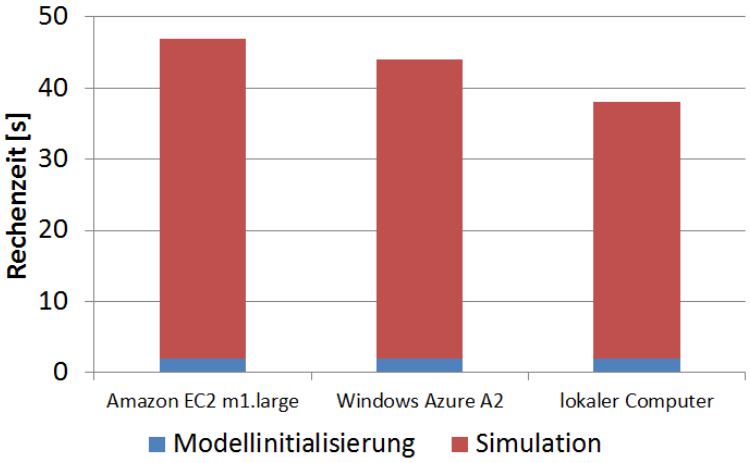


Abbildung 2: Rechenzeiten von ArcEGMO auf den einzelnen Plattformen

Werden jedoch die Instanzen der neuesten Generation genutzt, wie sie von Amazon bereitgestellt werden, so lässt sich eine beachtliche Leistungssteigerung (hier bis zu 15%) gegenüber dem lokalen Computer erzielen (**Abb. 3**).

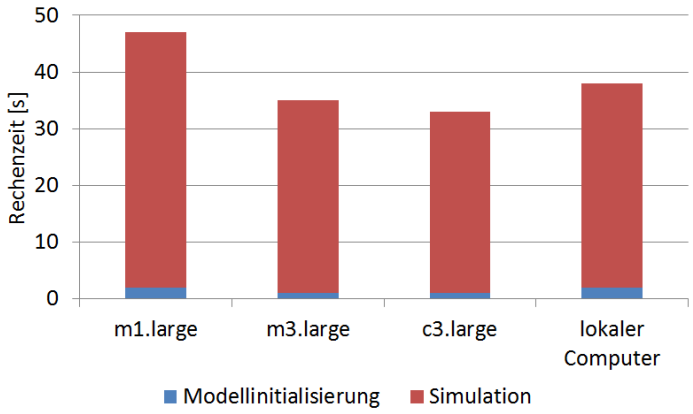


Abbildung 3: Unterschiede zwischen den einzelnen Amazon-EC2-Instanztypen und dem lokalen Rechner in der benötigten Rechenzeit

Wird die Anzahl der Kerne je neuem Instanztyp erhöht, so lässt sich die Rechengeschwindigkeit erwartungsgemäß nicht mehr steigern, da diese zusätzlichen Resourcen aufgrund seiner sequentiellen Programmierung nicht durch das Modell genutzt werden (Untersuchungsergebnisse hier nicht weiter ausgeführt). Eine Überraschung stellt dabei das entsprechende Preismodell von Amazon dar, da die momentane Preise für die neuen Generationen unter denen der alten Generation liegen (Amazon, 2014a; **Abb. 4**).

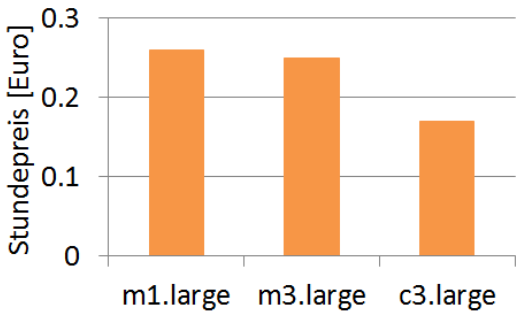


Abbildung 4: Stundenpreise je Amazon-EC2-Instanztyp

Aufgrund der besseren Rechenleistung und den geringeren Preisen ist eine Entscheidung für die Amazon-Instanzen der aktuellsten Generation durchaus vertretbar. Wenn man die on-demand-Preise ohne Reservierung auf Monatsbasis betrachtet (180 Euro für m3.large bzw. 125 Euro für c3.large bei voller Auslastung) ist der Preis dennoch zu hoch für eine dauerhafte Nutzung. Die Kosten, die dabei entstehen, halten sich für kleine Gebiete, in denen die Rechenzeit gering ist, in Grenzen, und liegt bei Amazon bei unter 20 Cent für eine Instanz mit aktuellster und rechenstärkster Hardware. Sollen jedoch große, räumlich hoch aufgelöste Modelle kalibriert werden, so dass die Rechenzeit mehrere Tage drastisch steigt, werden zum einen Instanzen notwendig, die mehr Arbeitsspeicher aufweisen, und zum anderen mehr Rechenzeit benötigen, was zusammen die Kosten stark erhöht. Ein einzelner Modelllauf kann somit schnell mehrere Euro kosten. Nutzt man dann Instanzen der aktuellsten Generation von Amazon, kann man dennoch eine gute Performance des Modellbetriebs im Vergleich zu einem lokalen Gerät erwarten.

5 Fazit und Ausblick

Hochwassersimulationen zur Schadensvorsorge müssen verlässlich sein. Sie benötigen dafür ein ausreichend komplexes und gut kalibriertes Niederschlags-Abfluss-Modell. Am Beispiel des bewährten, hydrologischen Modellsystems ArcEGMO ließ sich für ein kleines Modellgebiet zeigen, dass Modellrechnungen , wie sie zur Kalibrierung von NA-Modellen zur Hochwassersimulation notwendig sind, auch auf virtuellen Computern in IaaS-Clouds gerechnet werden können. Die Rechenleistung liegt jedoch leicht unter der eines vergleichbaren lokalen Computers mit ähnlicher Hardware. Nur

wenn aktuelle Technologien auch von den IaaS-Anbietern zur Verfügung gestellt werden, ist eine Auslagerung von Modellrechnungen sinnvoll. Die Kosten, die dabei entstehen, sind vor allem bei Modellen mit verhältnismäßig wenig Komplexität gering. Es ist jedoch davon auszugehen, dass diese bei einem Modell mit sehr vielen Raumelementen aufgrund der benötigten Rechenzeit schnell ansteigen und mehrere Euro pro einzelnem Modelllauf betragen.

Durch den sequentiellen Aufbau des Modellssystems ArcEGMO können zusätzliche Kapazitäten in Form zusätzlicher Prozessorkerne momentan noch nicht ausgenutzt werden, weswegen auch nur eine verhältnismäßig geringe Leistungssteigerung durch verbesserte Prozessortechnologie in der Cloud erzielt werden kann. Aus diesem Grund ist Cloud Computing momentan nur eine Alternative unter den besonderen Umständen eines kurzfristigen, erhöhten Rechenbedarfs.

Die zunehmende Entwicklung der vergangenen Jahre hin zu Technologien zur Parallelisierung von Software wird ebenfalls einen Einfluss auf die Weiterentwickung von Modellen haben. Von den Autoren wird dazu bereits die Umstellung des Modellsystems ArcEGMO hin zur Einbindung mehrerer Prozessorkerne während einer Rechnung getestet. Der Vorteil dabei ist, dass Rechnungen auf mehrere Prozessorkerne verteilt werden können und je nach Anteil des parallelisierbaren Quellcodes eine Verringerung der Rechenzeit je Simulation zu erwarten ist. Nachteile sind der größere Programmieraufwand und der höhere Bedarf an technischer Kompetenz. Bestehender Code ist auch nicht unbegrenzt parallelisierbar, daher wird Cloud Computing auf lange Sicht interessant bleiben, zumal Parallelisierung aufgrund der Skalierung der zu nutztenden Hardware sehr gut unterstützt wird und bei den Anbietern auch weiterhin von einer positiven Preisentwicklung auszugehen ist.

Grundsätzlich existieren noch weitere Möglichkeiten, den Zeitaufwand für eine Modellkalibrierung zu verringern. Die bei einer automatisierten Modellkalibrierung genutzten Optimierungsverfahren können beispielsweise dazu angewiesen werden, die einzelnen Modellläufe während einer Optimierung auf mehrere Prozessorkerne zu verteilen und die Verwaltung der einzelnen Modellläufe selbst zu übernehmen, soweit

dies von den Verfahren unterstützt wird. Entsprechende Studien durch die Autoren, in welchen auch Clustercomputer zur Anwendung kommen sollen, befinden sich in der Vorbereitung.

6 Danksagung

Dieser Beitrag ist im Rahmen des Projekts „EEnVEB“ entstanden. Das Vorhaben wird gefördert aus Mitteln der Europäischen Union (Europäischer Sozialfonds).

7 Literaturverzeichnis

[Amazon 2014a]

Amazon Web Services: Amazon EC2 – Preise. aws.amazon.com/de/ec2/pricing, 11.3.2014.

[Amazon 2014b]

Amazon Web Services: Amazon EC2-Instances. aws.amazon.com/de/ec2/instance-types/, 11.3.2014.

[Armbrust u.a. 2010]

Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: A View of Cloud Computing. Communications of the ACM 53 (4), 2010.

[Bates et al. 2008]

Bates, Bryson; Kundzewicz, Zbigniew W.; Wu, Shaohong; Palutikof, Jean: Climate Change and Water. IPCC Technical Paper VI, Intergovernmental Panel on Climate Change (IPCC, Geneva), www.ipcc.ch, 2008.

[Berendes u.a. 2013]

Berendes, C.I., Ertel, M., Röder, T., Sachs, T., Süptitz, T., Eymann, T.: Cloud Computing lohnt sich (noch) nicht. 11th International Conference on Wirtschaftsinformatik, Leipzig, 2013.

[BITKOM 2012]

BITKOM: Cloud Computing - Evolution in der Technik, Revolution im Business. BITKOM-Leitfaden, Berlin, 2012.

[Bornebusch 2006]

Bornebusch, M.: Hydraulische und Hydrologische Modellsimulationen als Planungswerkzeug für Hochwasser-Schutz-Maßnahmen. Aachen, 2006.

[Bräuninger u.a. 2012]

Bräuninger, Michael; Haucap, Justus; Stepping, Katharina; Stühmeier, Torben: Cloud Computing als Instrument für effiziente IT-Lösungen. Hamburgisches WeltWirtschafts-Institut (HWWI), Hamburg, 2012.

[Büttner u.a. 2009]

Büttner, U., Thieming V, Lechtaler, E., Pfützner, B.: Kap. 2 Untersuchungsgebiet. In: Schumann, A. (Hrsg.): Entwicklungen integrativer Lösungen für das operationelle Hochwassermanagement am Beispiel der Mulde. Schriftenreihe Hydrologie / Wasserwirtschaft Ruhr-Universität Bochum, 2009.

[Deelman u.a. 2008]

Deelman, E., Singh, G., Livny, M., Berriman, B., Good, J.: The Cost of Doing Science on the Cloud: The Montage Example. In: Proceedings of the 2008 ACM/IEEE conference on Supercomputing, article No. 50. IEEE Press, Piscataway, 2008.

[GDV 2013]

Gesamtverband der Deutschen Versicherungswirtschaft: Erste Schadensbilanz: Hochwasser 2013 verursacht 180000 versicherte Schäden in Höhe von fast 2 Milliarden Euro. Pressemitteilung. Berlin, 2013.

[Hetzner 2014]

Hetzner Online AG: www.hetzner.de/hosting/produktmatrix/rootserver-produktmatrix/, 2014 (13.3.2014).

[Hochstätter 2011]

Hochstätter, C.H.: Microsofts Sicht auf die Cloud: Windows Azure im Praxistest. www.zdnet.de/41554705/microsofts-sicht-auf-die-cloud-windows-azure-im-praxistest/, 2011 (geprüft 11.3.2014).

[Humphrey u.a. 2012]

Humphrey, M., Beekwilder, N., Goodall, J.L., Ercan, M.B.: Calibration of watershed models using cloud computing. E-Science (e-Science), IEEE 8th International Conference, Chicago, 2012.

[Innova 2013]

Innova Handelshaus AG: www.innova24.biz/item/computer-und-navigation/pc-systeme/acer-desktop-veriton-s670g-intel-core2duo-e7600-2gb-intel-g-793994.htm, 2013 (geprüft 13.3.2014).

[Maßmann u.a. 2010]

Maßmann, S., Jakobs, F., Sellerhoff, F., Feldmann, J., Sieker, H., Lange, C., Om, Y. & Hinkelmann, R.: Hyd³Flow - Integrierte hydrologische und hydro-numerische Modellsysteme für eine verbesserte Hochwasservorhersage. Forum für Hydrologie und Wasserbewirtschaftung, "Tag der Hydrologie 2010", Braunschweig, 2010.

[Matros u.a. 2009]

Matros, R., Stute, P., von Zuydtwyck, N.H., Eymann, T.: Make-or-Buy im Cloud-Computing – Ein entscheidungsorientiertes Modell für den Bezug von Amazon Web Services. Bayreuther Arbeitspapiere zur Wirtschaftsinformatik, Bayreuth, 2009.

[Microsoft, 2014]

Microsoft: Windows Azure: Virtuelle Computer – Preisdetails. www.windowsazure.com/de-de/pricing/details/virtual-machines/, 11.3.2014.

[Moya Quiroga u.a. 2012]

Moya Quiroga, V., Popescu, I., Solomatine, D.P., Bociort, L.: Cloud and cluster computing in uncertainty analysis of integrated flood models. Journal of Hydroinformatics 15(1), 2012.

[Pfützner 2002]

Pfützner, B.: Description of ArcEGMO; Official homepage of the modelling system ArcEGMO. www.arcegmo.de, ISBN 3-00-022290-5, 2002

[Pfützner u.a. 2013]

Pfützner, B., Hesse, P., Mey, S., Klöcking, B.: N-A-Modellierung mit ArcEGMO. In: Dietrich, J. & Schöninger, M. (Hrsg.): HydroSkript, http://www.hydroskript.de, 2013.

[Pfützner & Hesse 2011]

Pfützner, B., Hesse, P.: Anwendung hydrologischer Modelle für die Hochwassserbemessung - Erfahrungen aus Sachsen-Anhalt. In: Schulte, A., Reinhardt, C., Dittrich, A., Jüpner, R., Lüderitz, V. (Hrsg.): Hochwasserdynamik und Risikomanagement - neue Ansätze für bekannte Probleme? Beiträge zum Gemeinsamen Kolloquium am 24.11.2011 in Berlin. Berichte aus der Geowissenschaft Freie Universität Berlin, Aachen, 2011.

[Rouholahnejad u.a. 2012]

Rouholahnejad, E., Abbaspour, K.C., Vejdani, M., Srinivasan R., Schullin, R., Lehmann, A.: A parallelization framework for calibration of hydrological models. Environmental Modelling & Software 31 (May), 2012.

[Tonninger 2011]

Tonninger, W.: Die Cloud-Gretchen-Frage: IaaS oder PaaS? http://businessreadyblog.wordpress.com/2011/02/25/die-cloud-gretchen-frage-iaas-oder-paas/, 2011 (geprüft 11.3.2014).

[Walterbusch & Teuteberg 2014]

Walterbusch, M., Teuteberg, F.: Datenverluste und Störfälle im Cloud Computing: Eine quantitative Analyse von Service Level Agreements, Störereignissen und Reaktionen der Nutzer. Proceedings: Multikonferenz Wirtschaftsinformatik, Paderborn, 2014.

Workshop 2015

07./08. Mai 2015
am Wissenschaftlichen Zentrum für Informationstechnik-Gestaltung (ITeG) der Universität Kassel

Suchmaschinen-getriebene Umweltportale – Nutzung einer ElasticSearch-Suchmaschine

Thorsten Schlachter, Clemens Düpmeier, Christian Schmitt
Karlsruher Institut für Technologie, Institut für Angewandte Informatik
thorsten.schlachter@kit.edu, clemens.duepmeier@kit.edu, christian.schmitt@kit.edu

Wolfgang Schillinger
Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg
wolfgang.schillinger@lubw.bwl.de

Abstract

In order to merge data from different information systems in web portals, querying of this data has to be simple and with good performance. If no direct, high-performance query services are available, data access can be provided (and often accelerated) using external search indexes, which is well-proven for unstructured data by means of classical full text search engines. This article describes how structured data can be provided through search engines, too, and how this data then can be re-used by other applications, e.g., mobile apps or business applications, incidentally reducing their complexity and the number of required interfaces. Users of environmental portals and applications can benefit from an integrated view on unstructured as well as on structured data.

Einleitung

Um Daten aus verschiedenen Informationssystemen/-quellen als integrierte Ansichten gemeinsam in Portalen darstellen zu können, müssen diese Daten möglichst gezielt und performant abfragbar sein. Je nach Anforderung an die Stärke dieser Integration

müssen dabei die Daten selbst miteinander verknüpft sein, oder die Kopplung bezüglich bestimmter Eigenschaften (z.B. Ort) muss ermittelt werden.

Viele potenzielle Datenquellen erfüllen weder die Anforderungen an (einfache) Abfrageschnittstellen noch die notwendige Performanz und/oder Verfügbarkeit. Eine mögliche Lösung hierfür ist eine redundante Datenhaltung, die optimal auf die Abfragen durch die Portalanwendungen zugeschnitten ist.

Ein gängiges Beispiel für dieses Vorgehen stellen Volltextsuchmaschinen dar. Diese laden und analysieren asynchron die verfügbaren Daten und bereiten sie in für Suchanfragen optimierte Datenstrukturen auf (Indexbildung). Volltextsuchmaschinen bieten diese Funktionalität im Wesentlichen für un- oder schwach strukturierte Daten (Dokumente) an. Verschiedene NoSQL-Datenbankkonzepte und darauf aufsetzende Indexierer legen nahe, dieses Prinzip auch für strukturierte Daten umzusetzen. Da semistrukturierte Daten in eine strukturierte Form überführt werden können [Abiteboul, 1999], ist im Folgenden nur noch von unstrukturierten bzw. strukturierten Daten die Rede.

Dieser Beitrag beschreibt wie auch strukturierte Daten über Suchmaschinen zur Verfügung gestellt und von Portalen sowie anderen Anwendungen, zum Beispiel mobilen Apps oder Fachanwendungen, genutzt werden können.

Dies kann für viele Anwendungsfälle zu einer Reduzierung sowohl der Komplexität von Anwendungen als auch der Anzahl der erforderlichen (technischen) Schnittstellen führen. Außerdem lassen sich die Daten während der Aufbereitung für die Suchmaschine automatisiert anreichern, z.B. um abgestimmte Deskriptoren, wodurch die spätere Abfragbarkeit und Verknüpfbarkeit im Portal verbessert werden kann. Konkret geht es um die Nutzung einer ElasticSearch-Suchmaschine für die Landesumweltportale verschiedener Länder und die Bereitstellung von strukturierten Daten (Sach- und Geodaten) als Ergänzung zu einer „klassischen“ Volltextsuchmaschine.

1 Umgang mit größer werdenden Datenmengen

Wenn Daten dem Nutzer bedarfsorientiert in verschiedenen Granularitäten, vom Einzelobjekt bis zur aggregierten Ansicht, zur Verfügung gestellt werden sollen, bedeutet dies insgesamt eine Zunahme von Datenobjekten. Viele herkömmliche Ansätze zur Abfrage solcher Daten sind dem nicht mehr gewachsen, und der Entwurf adäquater Nutzerschnittstellen gewinnt daher an Bedeutung.

Während für den Zugriff auf unstrukturierte textuelle Daten üblicherweise Volltextsuchmaschinen verwendet werden, ist die Abfrage von strukturierten Daten häufig durch komplexe Suchformulare, Selektoren oder Kartenansichten realisiert, in der Regel ergänzt um eine dedizierte Anwendungslogik, die den Zugriff auf die zugrunde liegenden Datenbanken regelt.

Moderne Anwendungen sollten dem Nutzer, insbesondere Nicht-Fachleuten, eine einfach zu bedienende Schnittstelle zur Verfügung stellen, welche die Komplexität der Datenrecherche verbirgt, so wie es die meisten Internet-Suchmaschinen mit ihrem einzelnen Suchschlitz tun. Die Präsentation der Suchergebnisse sollte angepasst an deren Typ erfolgen, z.B. in Ergebnislisten, Diagrammen oder Kartenansichten.

2 Nutzererwartungen, Anforderungen und Ziele

Die Nutzung von Mobilgeräten hat das Verhalten der Nutzer und deren Erwartung an die Benutzerschnittstellen massiv beeinflusst. Daten werden auf Mobilgeräten häufig kontextbezogen abgerufen, passende Informationen werden abhängig von Standort, Zeit und Situation angezeigt. Für viele Nutzer sind lange Navigationspfade mit vielen Interaktionen daher nicht mehr akzeptabel. Die Navigation erfolgt häufig explorativ, basierend auf bereits dargestellten Informationen. Jede Interaktion kann eine Änderung der Abfrage auslösen, d.h. die dargestellten Informationen müssen ggf. aktualisiert werden. Dies stellt erhebliche Anforderungen sowohl an die Flexibilität der Nutzerschnittstelle als auch an ihre Performanz, ebenso wie an die der zugrundeliegenden Dienste. Für den Nutzer sind spürbare Latenzen nach Interaktionen und bei Daten-Updates nicht akzeptabel.

Leider erfüllen viele Systeme und Dienste diese Anforderungen (noch) nicht. Um eine flexible und hochperformante Abfrage von strukturierten Daten zu ermöglichen, ist es daher sinnvoll, diese Daten hierfür speziell aufzubereiten – ein Ansatz, der für unstrukturierte Daten durch Volltextsuchmaschinen seit langem erfolgreich praktiziert wird. Die Prinzipien von Suchindexen lassen sich auch auf strukturierte Daten anwenden.

In Suchmaschinen-getriebenen Webportalen wird daher nicht auf die Originaldaten und -abfrageschnittstellen, sondern auf hochoptimierte Suchindexe zugegriffen, d.h. alle primären Informationen werden über die Suche erreicht. Sowohl Zugriffsstrukturen, z.B. Menüs, als auch Inhalte werden als Resultat einer oder mehrerer Suchanfragen dargestellt. Suchmaschinen-getriebene Websites und Portale können daher als hochdynamisch charakterisiert werden.

2.1 Klebstoff für verteilte Informationen

Die Nutzung von Suchmaschinen und -indices kann auch zur Lösung weiterer Probleme dienen, die sich aus der klassischen service-orientierten Architektur ergeben: Die Landschaft von Umweltinformationssystemen ist sehr heterogen, dies gilt insbesondere auch für die genutzten Schnittstellen und Dienste. Eine deutliche Reduzierung der notwendigen Schnittstellen innerhalb der Portalsoftware wäre wünschenswert – wenn diese auch zur Synchronisierung von Daten mit der Suchmaschine an anderer Stelle benötigt werden. Allerdings profitieren dann auch andere Anwendungen, z.B. mobile Apps, von der Verfügbarkeit einer einzelnen – oder zumindest von wenigen – Abfrageschnittstellen.

Ein Suchindex kann auch für solche Daten genutzt werden, die nicht per Service zur Verfügung stehen, oder die sich hinter komplexen Abfragen verbergen. Doch selbst wenn Daten bereits als Services zur Verfügung gestellt werden, haben diese häufig eine zu lange Antwortzeit, um sinnvoll aus Portalen oder mobilen Anwendungen genutzt werden zu können.

Während des Imports in einen bzw. der Synchronisierung mit einem Suchindex können Daten gefiltert oder um weitere Attribute ergänzt werden, z.B. können Ortsangaben durch Nutzung von Gazetteer-Services um explizite Geokoordinaten ergänzt werden,

mit deren Hilfe dann z.B. eine Umkreissuche realisiert werden kann. Bei der Anreicherung von Daten kann durch Verwendung von übergreifenden (kontrollierten) Vokabularen und Schlüsseln die semantische Verknüpfbarkeit von Daten unterschiedlicher Herkunft sichergestellt werden, ähnlich wie dies bei Linked Data [Berners-Lee, 2006] der Fall ist. Durch die Anwendung von Verfahren des maschinellen Lernens ist auch eine komplexere Anreicherung der Daten möglich.

2.2 Ziele des Suchmaschinen-getriebenen Ansatzes

Zusammenfassend werden durch die Suchmaschinen-getriebene Architektur die folgenden Ziele verfolgt:

- Verwendung von Daten aus der Originalquelle und Konzentration in einer leistungsfähigen Suchmaschine
- Umfassende Sicht auf alle verfügbaren, relevanten Informationen über eine einfache Benutzerschnittstelle
- Bereitstellung von Hilfsmitteln für eine explorative Recherche
- Sowohl die Abfrage von Objektklassen als auch von einzelnen Objekten
- Alle Daten enthalten Verweise (Links) auf ihr Quellsystem, da nur die wichtigsten Informationen im Portal oder der mobilen Anwendung gezeigt werden
- Reduzierung der Anzahl und Komplexität von Datenschnittstellen (für Portale und andere Anwendungen)
- Verwendung des Suchindexes durch eine Vielzahl von Anwendungen
- Keine oder wenig zusätzliche Last für Originalsysteme durch zusätzliche Suchanfragen
- Die Abfrage einer großen Anzahl von Objekten muss so performant wie eine „einfache" Volltextsuche sein
- Der aktuelle Zustand der Umwelt (Messungen) muss dargestellt werden

- Kontextinformationen (mobile Geräte, individuelle Einstellungen) müssen in der Suche Verwendung finden.

3 Grundlegende Architektur

Die in Abbildung 1 dargestellte Archtitektur für Suchmaschinen-getriebene Umweltportale basiert auf einem JEE (Java Enterprise Edition) Portalserver *Liferay Portal* [Liferay, 2015], der sowohl als Web-Frontend als auch als Basis für die zugrundeliegenden Services dient. Die Portalsoftware bietet eine große Auswahl von Standard-Komponenten wie eine Nutzer- und Rechteverwaltung, Designvorlagen für responsive Layouts [Zillgens, 2012], Content-Management, etc. Für unsere Zwecke ist jedoch insbesondere die modulare Systemarchitektur und die Erweiterbarkeit dieser Basissoftware von Interesse. Neue Anzeigekomponenten können durch Konzepte wie Portlets und Web Widgets leicht hinzugefügt werden. Die Anwendungslogik dieser Erweiterungen nutzt zwei Suchmaschinen für den primären Datenzugriff.

3.1 Nutzung spezialisierter Suchmaschinen

Verschiedene Suchmaschinen haben ihre jeweils spezifischen Stärken und Einsatzgebiete. Deshalb kommen in den beschriebenen Umweltportalen zwei spezialisierte Suchmaschinen zum Einsatz (Abb. 1). Unstrukturierte Daten werden durch eine traditionelle Volltextsuchmaschine per Web-Crawler indiziert, eine Google Search Appliance [Google, 2015].

Strukturierte Daten hingegen werden in Form von JSON-Objekten in einem separaten Suchindex einer ElasticSearch-Suchmaschine [Elastic, 2015] gespeichert und indexiert. Letztere kann komplett über eine REST-Schnittstelle beschickt und administriert werden. Die für die einmalige oder fortlaufende Synchronisation notwendigen Prozesse laufen außerhalb der Portal-Software und sollten Bestandteil von übergreifenden Workflows sein, die auch das Qualitätsmanagement für die Daten einschließen.

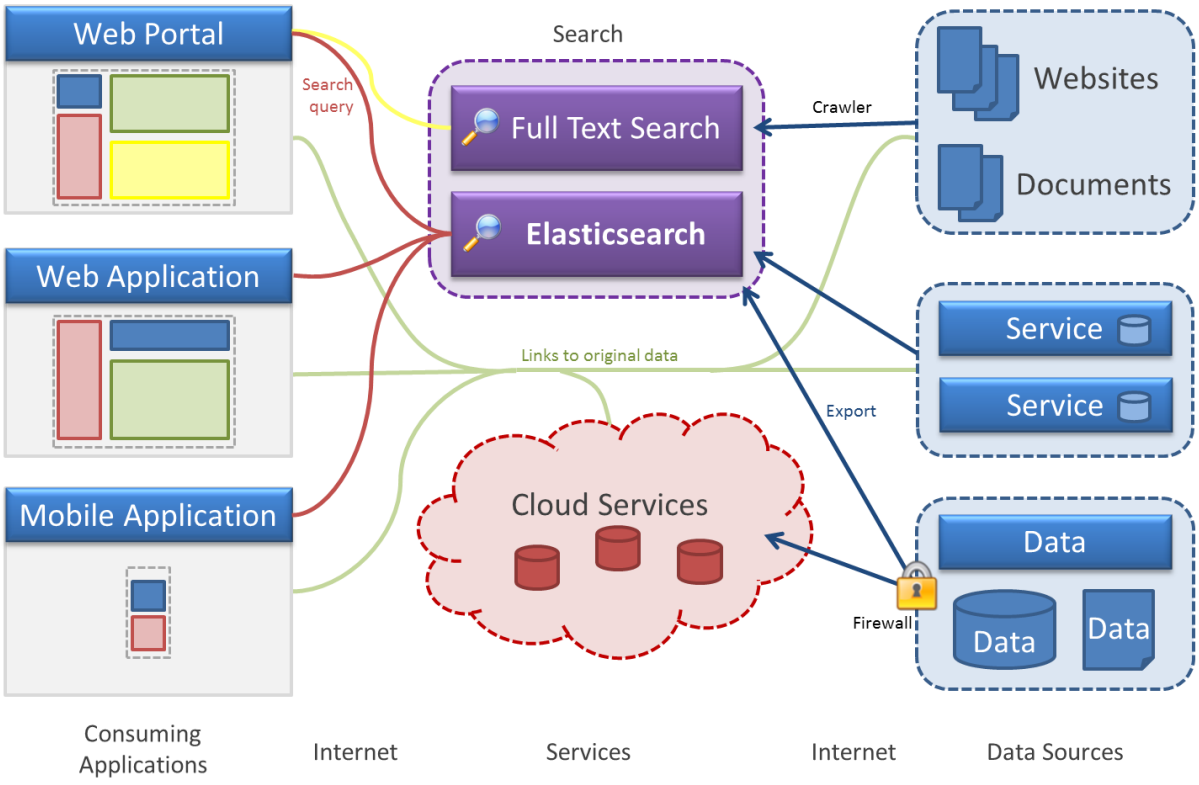


Abbildung 1: Hauptkomponenten einer Suchmaschinen-getriebenen Architektur für Webportale und mobile Anwendungen inkl. zweier Suchmaschinen und Cloud-Diensten zur Ergänzung von Original-Systemen.

Wenn spezielle Daten nicht als Services zur Verfügung stehen, so können diese zusätzlich durch Nutzung von (generischen) Cloud-Diensten für spezielle Anwendungsfälle verfügbar gemacht werden, z.B. Beispiel können Geodaten aus Dateien über die Google Maps Engine bereitgestellt werden und anschließend durch Anwendungen in Form von Kartenlayern genutzt werden.

Externe Suchmaschinen können weitere Daten liefern, z.B. statistische Daten [Statistik-BW, 2015] oder Beschreibungen von Verwaltungsverfahren [Service-BW, 2015]. Dabei ist die Verwendung standardisierter Schnittstellen bei deren Integration hilfreich, z.B. von OpenSearch [Opensearch, 2015].

Auch Portal-eigene Informationen (Content, Assets) können über die strukturierte Suchmaschine in vielen Fällen performanter abgefragt werden als über das Portal selbst. Über entsprechende Konnektoren ist eine Synchronisation von Inhalten aus Portal- oder CMS-Systemen mit der Suchmaschine häufig „out-of-the-box“ möglich. Durch intelligente Vorverarbeitung und Synchronisation mit einer Suchmaschine kann

die Portalsoftware spürbar entlastet werden. Liferay 7 wird mit einer eingebauten ElasticSearch-Suche kommen.

3.2 Such-Funktionalität

Neben redaktionell gepflegten Inhalten und Navigationshilfen bilden die Suchmaschinen den Kern der Portalanwendung. Vom Nutzer gestellte Suchanfragen werden vorverarbeitet und semantisch angereichert, z.B. durch die Verwendung von Gazetteer-Services zur Erkennung ihres geographischen, zeitlichen oder thematischen Bezugs, und anschließend an die Suchmaschinen weitergeleitet.

Durch diese Anreicherung können Suchanfragen erheblich präziser ausgeführt werden und heben sich daher von einer herkömmlichen, auf dem Vergleich von Zeichenketten basierenden Volltextsuche ab. So können Anfragen mit Geobezug beliebige Objekte in der Nähe nur aufgrund ihrer Lage identifizieren. Durch die Verwendung angereicherter Suchanfragen können die Suchmaschinen die Suchergebnissen auch nach zusätzlichen, konfigurierbaren Relevanzkritieren gewichten.

Die Präsentation der Suchergebisse erfolgt abhängig von Typ der Suchergebnisse durch verschiedene Präsentations-Komponenten, z.B. klassische Trefferlisten für unstrukturierte Informationen, tabellarischen Ansichten für strukturierte Informationen, Diagramme für Messreihen oder Kartenansichten für Daten mit Geobezug. Die einzelnen Komponenten sind dabei als Web Widgets realisiert, gekapselte Javascript-Komponenten, die jeweils einen Visualisierungstyp implementieren. Sobald eine breite Unterstützung durch die gängigen Webbrowser gewährleistet ist können diese durch standardisierte Web Components [W3C, 2014] ersetzt werden.

Web Widgets und Web Components sind unabhängig von ihrer umgebendenden Anwendung, d.h. sie können in weiteren Anwendungen wiedergenutzt werden, z.B. in hybriden Mobilanwendungen [Wikipedia, 2015] oder anderen Webanwendungen und -portalen.

Für den Datenzugriff nutzen die Widgets in der Regel REST-Dienste. Zur nahtlosen Integration in die umgebende Portalanwendung wird jedes Web Widget in einem Hüll-

Portlet (Wrapper-Portlet) verpackt. Die einzelnen Widgets werden dabei so konfigurierbar und generisch wie möglich gestaltet, z.B. kann die jeweilige Ausgabe durch Templates [Mustache, 2015] beschrieben werden. Die Wiederverwendbarkeit von Daten und Templates für mehrere Einsatzbereiche kann durch die Verwendung von (ggf. redundatenten) allgemeinen Attributen wie „Titel“, „Beschreibung“ und „Link“ weiter gefördert werden.

3.3 Ergebnis-Mashup

Bei der Anzeige von Suchergebnissen können spezialisierte Komponenten Menüs, Auswahlen zur Verfeinerung der Suche, Trefferlisten oder einzelne Objekte anzeigen. Bei jeder Komponente kann es sich dabei wiederum eine mehr oder weniger komplexe Anwendung handeln, z.B. eine Ansicht zur Darstellung von Objekten auf einer Karte. Es werden in der Regel mehrere Portlets (inkl. Web Widgets) auf einer Trefferseite angezeigt. Dadurch entsteht eine übergreifende, integrierte Ansicht aller Suchergebnisse (Abb. 3).

Um dem Nutzer dabei eine möglichst interaktive Oberfläche „aus einem Guss“ zur Verfügung zu stellen, müssen die einzelnen Komponenten miteinander kommunizieren können. Ein Ereignisbus (Event-Bus) implementiert die lose Kopplung von Komponenten (Abb. 2). Jede Komponente kann dabei sowohl auf Ereignisse hören als auch eigene Ereignisse auf den Bus schicken. Jede Komponente ist dabei selbst dafür verantwortlich, auf welche Ereignisse sie reagiert und auf welche nicht. Um jedoch eine kohärente Interaktion der einzelnen Komponenten zu erreichen, muss die Anwendung eine Reihe von Event-Typen (z.B. Auswahl eines Ortes, Selektion eines bestimmten Objekts) definieren, auf die (möglichst) alle Komponenten reagieren.

Das Event-Konzept kann dabei um weitere Mechanismen erweitert werden, z.B. um das Zusammentragen von Personalisierungsinformationen (wie bevorzugten Orten oder Themen) oder Sensordaten (aktueller Standort) auf Mobilgeräten.

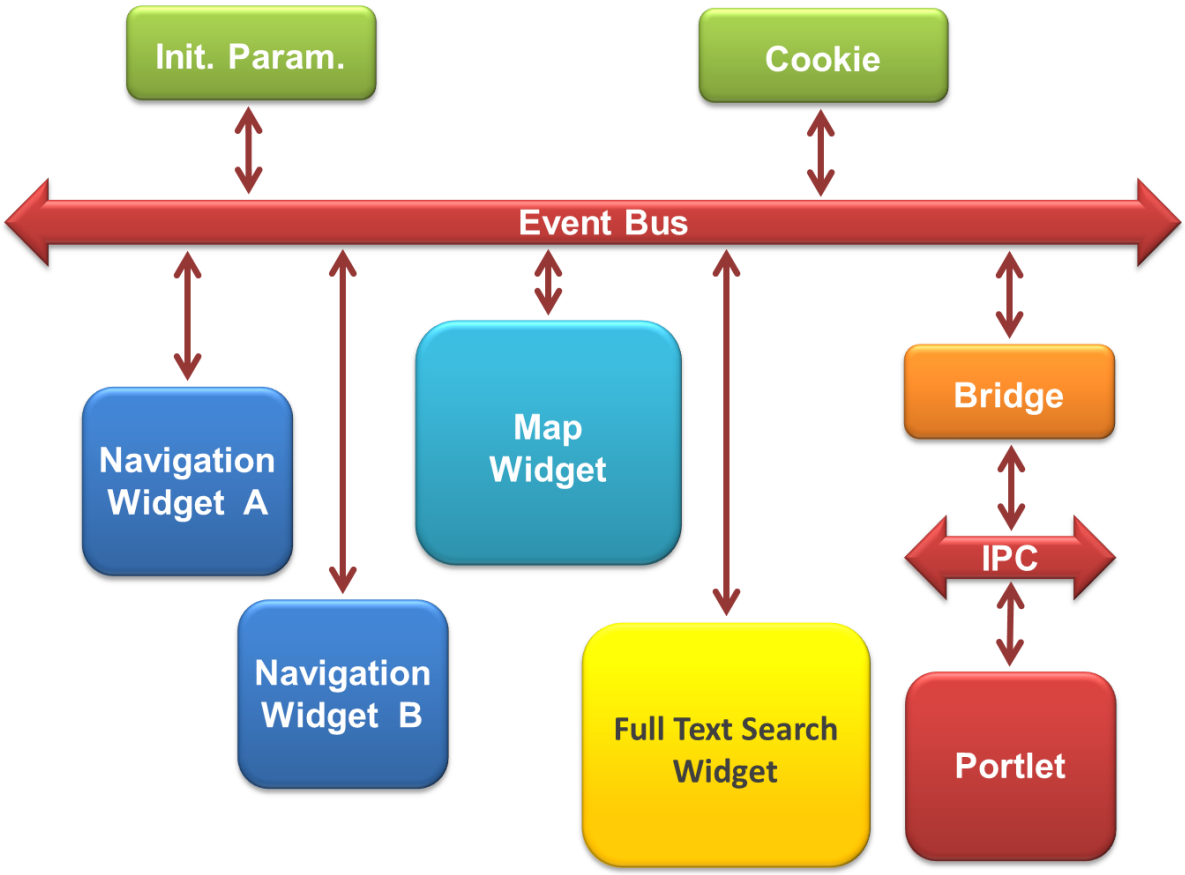


Abbildung 2: Der Ereignisbus verbindet verschiedene Komponenten auf einer Ergebnisseite. Weitere Komponenten können über Schnittstellen (Bridge) angebunden werden, z.B. die Inter-Portlet-Kommunikation von Liferay

4 Case-Study „Landesumweltportale“

Der beschriebene Suchmaschinen-basierte Ansatz wurde bei der Reimplementierung der Landesumweltportale (LUPO) [Schlachter, 2014] genutzt. Zusätzlich zu den schon in der Vorversion vorhandenen schwach bzw. unstrukturierten Daten (rund 2.000.000 Dokumente aus etwa 2.000 Quellen) wurden nun auch viele strukturierte Datensätze in einem ElasticSearch-Index integriert, z.B. diverse Schutzgebietstypen, Energieanlagen wie Windräder, Biogasanlagen und Solaranlagen. Wo möglich wurden auch die zugehörigen Geodaten übernommen. Dieser Index stellt nun die Basis für einen inhaltliche und funktionale Erweiterung des Portals dar, deren augenscheinlichste Neuerung die Kartenansicht mit ortsbasierter Suche ist.

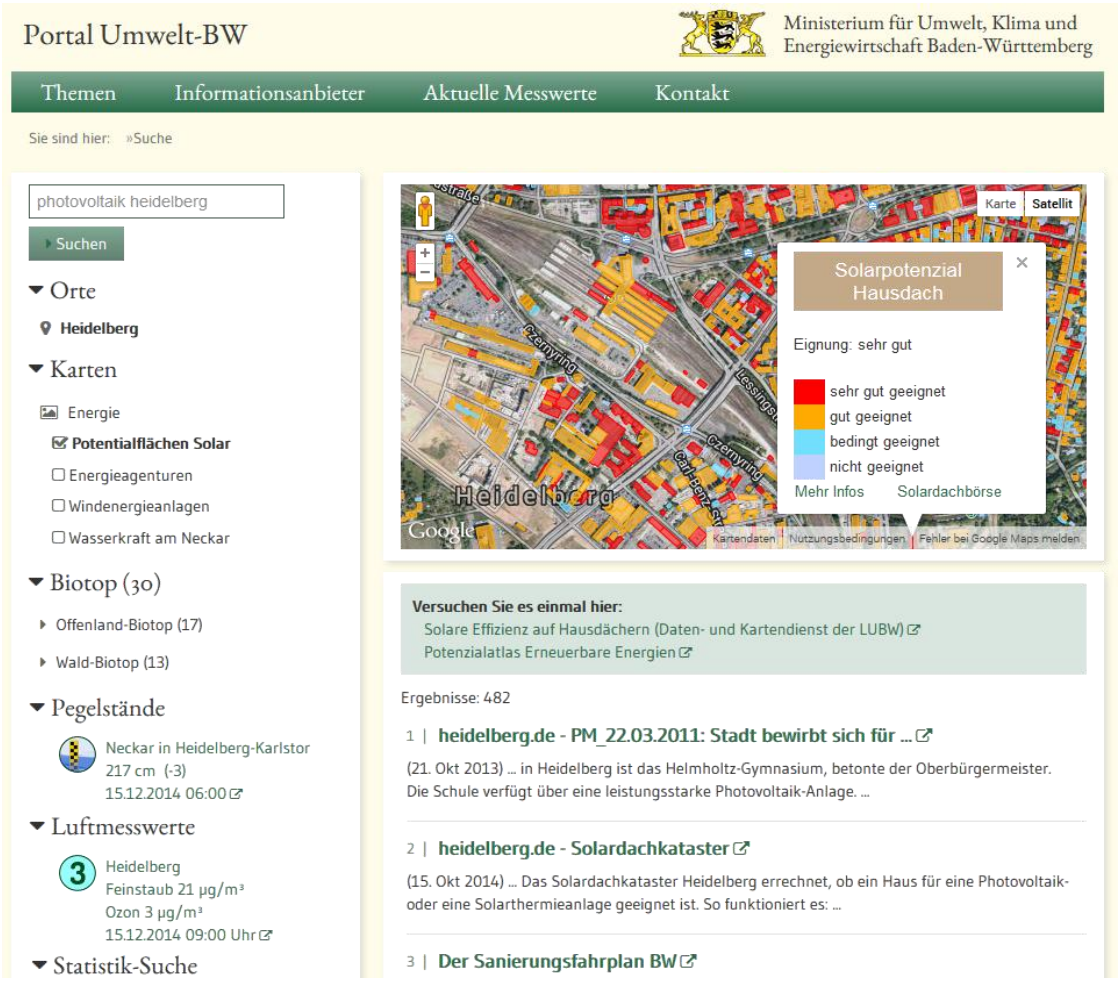


Abbildung 3: Ergebnisseite im Landesumweltportal Baden-Württemberg. Klassen von Objekten, einzelne Objekte und Messwerte werden in der linken Spalte dargestellt, passende Geoinformationen innerhalb der Kartenansicht. Unten rechts befindet sich die Trefferliste der Volltextsuche.

Jede Suchanfrage, egal ob diese manuell eingegeben oder durch eine Interaktion ausgelöst wird, generiert nun eine zunächst leere Ergebnisseite. Nach der Vorverarbeitung und semantischen Erweiterung der Suchanfrage werden beide Suchmaschinen parallel und asynchron abgefragt. Während die Treffer der Volltextsuche in einer klassischen Trefferliste dargestellt werden, kann die Bereitstellung der strukturierten Suchergebnisse eine ganze Reihe von Darstellungen anstoßen, z.B. können Objekte einer Klasse in einer Listenansicht und gleichzeitig auf der Karte dargestellt werden, während Messwerte in einer spezialisierten Komponente dargestellt werden (Abb. 3).

Die Weiternavigation kann dann z.B. durch Klicken auf ein bestimmtes Objekt erfolgen, der dann das Zentrieren der Karte auf die entsprechende Position und die Auswahl des passenden Kartenlayers auslösen kann.

Die Ergebnislisten können auch quantitative Informationen enthalten, z.B. wie viele Schutzgebiete welchen Typs passend zur Suchanfrage gefunden wurden.

Viele Komponten der Landesumweltportale (Daten, Dienste, Suchmaschine, UI-Komponenten) werden auch in der mobilen App „Meine Umwelt“ [Schlachter, 2013] genutzt. Dabei handelt es sich um eine hybride App mit einem Kern auf Basis von HTML5, die mit einer nativen Hülle für iOS, Android und Windows phone verfügbar ist.

5 Zusammenfassung und Ausblick

Das beschriebene Vorgehen kann so zusammengefasst werden: Moderne Webanwendungen und mobile Anwendungen laden Daten dynamisch. Hierfür wird eine große Anzahl von unterschiedlichen Schnittstellen und Datenquellen verwendet. Um die Anwendungen zu vereinfachen und um die Last auf die Originalsysteme zu reduzieren, können Daten in einem separaten Suchindex verwaltet werden:

- Überlasse die schwere Arbeit der Suchmaschine (Vorverarbeitung, Indexierung, Abfragen)
- Reichere existierende Daten (semantisch) an
- Fasse individuelle Objekte zu Klassen zusammen
- Nutze generische Formate und Mechanismen, z.B. Darstellungstemplates, damit eine ganze Reihe von Anwendungen „out-of-the-box“ abgedeckt werden kann.

Was unterscheidet diesen Ansatz von vorigen?

- Minimierung der Schnittstellenvielfalt in den konsumierenden Anwendungen
- Leichte Integration von neuen Datensätzen
- Schnelles (!) Auffinden von Objektklassen und einzelnen Objekten
- Feingranulare Ortssuche auf Objekt-Level (im Unterschied zu ganzen Layern)
- Kombination von strukturierter und unstrukturierter Suche
- Reduktion von redundant verwalteten Metadaten

- Wiederverwendbarkeit von Web Widgets
- Mehrfachnutzung von Suchindexen und Diensten
- Mehr Interaktivität für den Nutzer
- Entlastung der Portal-Software

Das Konzept der Suchmaschinen-getriebenen Umweltportale steht und fällt mit der Relevanz seiner Such-Indices. Automatisierte Updates der Daten und die notwendige Datenkonsistenz dürfen nicht unterschätzt werden. Zunächst bedeutet auch der Betrieb einer Suchmaschine zusätzlichen Aufwand und wirft Fragen bezüglich Konsistenz, Redundanzvermeidung und Betriebkkosten auf. Jedoch wiegen die Vorteile des Suchmaschinen-basierten Ansatzes diese Punkte leicht auf. Das Nutzerfeedback ist sehr positiv und die Erwartungen an die neuen Umweltportale wurden in vielen Fällen übertroffen.

Die Verwaltung von unstrukturierten und strukturierten Daten durch Suchmaschinen stellt einen ersten Schritt in Richtung einer einheitlichen Zugriffsplattform (Unified Information Access Platform, UIA) dar [Probstein, 2010], auch wenn zentrale Aspekte einer UIA wie Datenprozessierung und Datenanalyse bestenfalls im Ansatz enthalten sind. Jedoch bieten die Basistechnologien ein großes Potenzial für Erweiterungen.

Die meisten geplanten Erweiterungen betreffen weniger die Architektur als die umsetzenden Anwendungen, in denen insbesondere im Bezug auf Nutzerfreundlichkeit und Ergonomie noch Verbesserungen möglich sind. Durch die lose Koppung der Komponenten ist die Integration weiterer Navigationshilfen und weiterer Funktionalität sehr einfach.

Die bessere Verknüpfung der Daten innerhalb des Suchindex ist sicherlich der vielversprechendste Ansatz für die Verbesserung der Qualiät von Suchergebnissen. Die Nutzung gemeinsamer Konzepte, Vokabulare und Schlüssel macht vormals isolierte Dateninseln leichter verknüpfbar und ermöglicht ganz Anwendungen.

6 Literaturverzeichnis

[Abiteboul, 1999]

Abiteboul, S.; Buneman, P.; and Suciu, D.: „Data on the Web: From Relations to Semistructured Data and XML", Morgan Kaufmann, ISBN 978-1558606227, 1999

[Berners-Lee, 2006]

Berners-Lee, T.: „Linked Data", http://www.w3.org/DesignIssues/LinkedData.html, 2006 (besucht am 20.04.2015)

[Elastic, 2015]

Elasticsearch: https://www.elastic.co/de/ (besucht am 20.04.2015)

[Google, 2015]

Google Search Appliance: http://www.google.de/enterprise/search/ (besucht am 20.04.2015)

[Liferay, 2015]

Liferay Portal: http://www.liferay.com (besucht am 20.04.2015)

[Mustache, 2015]

Mustache: http://mustache.github.io (besucht am 20.04.2015)

[Opensearch, 2015]

OpenSearch: http://www.opensearch.org/Specifications/OpenSearch/1.1 (besucht am 20.04.2015)

[Probstein, 2010]

Probstein, S.: „Five Advantages of Unified Information Access (UIA)"; CIO; 2010; www.cio.com/article/2416284/ (besucht am 20.04.2015)

[Service-BW, 2015]

Service-BW: http://service-bw.de/zfinder-bw-web/processes.do (besucht am 20.04.2015)

[Schlachter, 2013]

Schlachter, T.; Düpmeier, C.; Weidemann, R.; Schillinger, W.; Bayer, N.: „My environment - a dashboard for environmental information on mobile devices", International Symposium on Environmental Software Systems (ISESS 2013), 2013

[Schlachter, 2014]

Schlachter, T. et al.: „LUPO - Weiterentwicklung der Landesumweltportale", in: Weissenbach, K. et al. (Eds.): Umweltinformationssystem Baden-Württemberg F+E-Vorhaben MAF-UIS Moderne anwendungsorientierte Forschung und Entwicklung für Umweltinformationssysteme, Phase II 2013/14, KIT Scientific Reports 7665, ISBN 978-3731502180, 2014, pp. 65-74

[Schlachter, 2015]

Schlachter, T., et.al.: „Towards a Search Driven System Architecture for Environmental Information Portals", International Symposium on Environmental Software Systems (ISESS 2015 in Melbourne/Australien), IFIP Advances in Information and Communication Technology Vol. 448, Springer International Publishing; S. 351-360, 2015, http://link.springer.com/chapter/10.1007/978-3-319-15994-2_35

[Statistik-BW, 2015]

Statistisches Landesamt Baden-Württemberg: http://www.statistik-bw.de (besucht am 20.04.2015)

[W3C, 2014]

Web Components: http://www.w3.org/TR/components-intro/ (besucht am 20.04.2015)

[Wikipedia, 2015]

Multi-channel app development: http://en.wikipedia.org/wiki/Multi-channel_app_development (besucht am 20.04.2015)

[Zillgens, 2012]

Zillgens, C.: „Responsive Webdesign: Reaktionsfähige Websites gestalten und umsetzen“, Carl Hanser Verlag, ISBN 978-3446430150, 2012

UmweltWiS: Von Umweltinformationssystemen zu „Umweltwissenssystemen“?

Mareike Dornhöfer, m.dornhoefer@uni-siegen.de
Madjid Fathi, fathi@informatik.uni-siegen.de
Lehrstuhl für Wissensbasierte Systeme & Wissensmanagement, Universität Siegen

Abstract

Environmental Information Systems (EIS) are a kind of information system, addressing either the public or business sector. Different environmental data is gathered and visualized for the user. Now there is a question whether the application of semantic technologies and the interconnection and inference of data may transform EIS to environmental knowledge systems? The given article addresses this issue and proposes a concept for an “environmental knowledge system” called UmweltWiS.

Einleitung

Umweltinformationssysteme (UIS) sind eine Form von Informationssystemen, welche den öffentlichen oder betrieblichen Sektor adressieren. Hierzu werden Daten aus verschiedenen Quellen zusammengeführt und visuell für den Anwender aufbereitet. Durch die Anwendung und Einbringung von semantischen Technologien und eine Vernetzung der Daten, stellt sich jedoch die Frage inwiefern sich UIS hin zu Umwelt-wissenssystemen weiterentwickeln? Der vorliegende Beitrag diskutiert daher den Einsatz semantischer Technologien, die Möglichkeiten zur Vernetzung und Inferenz von Daten hin zu Wissen und stellt ein Konzept eines „Umweltwissenssystems“ (UmweltWiS) vor.

1 Umweltinformationssystem vs. Wissensbasiertes System

Klassische Informationssysteme sind betriebsindividuell und können zur *„Entscheidungsfindung, Koordination, Steuerung und Kontrolle“* sowie zur Problemanalyse von *„komplizierten Sachverhalten“* und im Rahmen der Produktentwicklung eingesetzt werden [Laudon et al, 2010, S. 17]. Umweltinformationssysteme (UIS) können sowohl betrieblich, als auch öffentlich sein. In der einfachsten Form lassen sich UIS durch das Zusammenbringen von Daten aus verschiedenen Umweltdatenbanken etablieren [Zifkovtis, Brunner, 2012, S. 233]. Die angestrebten Umweltinformationen sollen für verschiedene Zielgruppen, wie Öffentlichkeit, Kommissionen, Entscheidungsträger nutzbar sein und entsprechend öffentlich zugänglich gemacht werden [EU Kommission, 2013, S. 2], [Schlachter et al., 2014, S. 77]. Das Planungspapier der Europäischen Kommission aus 2013 [EU Kommission, 2013, S. 2] referenziert in diesem Zusammenhang sieben Kernanforderungen aus 2008 [EU Kommission, 2008] an die bereitzustellenden Umweltinformationen: (1) Quellnahe Verwaltung, (2) vielseitige Einsetzbarkeit, (3) Verfügbarkeit für (4) unterschiedliche Endbenutzer, (5) Ermöglichung einer Vergleichbarkeit, (6) öffentlicher Zugang auf Basis (7) *„kostenfreie(r), quelloffene(r) Softwarestandards"*. Die bereitgestellten Informationen werden als eine Entscheidungsgrundlage gesehen, welche maßgeblich von deren Qualität, Relevanz, dem Umfang, der Granularität und Vergleichbarkeit abhängt. Gerade letztere wird jedoch auf Grund von unterschiedlicher Repräsentation zuweilen bemängelt [EU Kommission, 2013, S. 4]. Basierend auf der von Fischer-Stabel bereits 2005 [Fischer-Stabel, 2005, S. 7] für Umweltinformationssysteme publizierten Architektur, lässt sich ein UIS aus einem Content Management System, Experteninterface, einem öffentlichen Zugang, einer Meta-Daten-Komponente, Fach-Informationssystemen, Basis-Informationssystemen, weiteren externen Systemen und einem Berichtssystem aufbauen. Die einzelnen Bausteine agieren dabei über eine Kommunikationskomponente miteinander.

Betrachtet man auf der anderen Seite Wissensbasierte Systeme (WBS), so werden diese vor allem zur Entscheidungsuntersützung und im speziellen der Problemlösung eingesetzt, setzen aber intelligente Methoden voraus, die aus bestehenden Wissensfragmenten mit Hilfe von Inferenztechniken Wissen folgern oder neues Wissen ableiten

[Bodendorf, 2006, S. 147]. In der Literatur wird häufig von der Wissenspyramide [Bodendorf, 2006, S. 2] oder der Wissenstreppe [North, 2011, S. 36] gesprochen, welche auf den Prinzipien Zeichen, Daten, Information und Wisssen aufbauen, wobei Wissen nur entstehen kann, wenn der Anwender oder das System in der Lage sind, aus vorliegenden Daten interpretierbare Informationen zu bilden, die wiederum durch Kontext und Vernetzung zu Handlungsempfehlungen werden. Die Darstellung des Wissens und dessen Repräsentation bzw. Verarbeitung sollten in einem WBS idealerweise getrennt voneinander gehalten werden [Beierle et al., 2008, S. 8]. Schematisch sind WBS daher in der Regel aus einer Wissensbasis, einer Wissensverarbeitungskomponente (zur Inferenz, Problemlösung), Schnittstelle für Experten zum Einpflegen neuen Wissens und einer Benutzerschnittstelle aufgebaut [Bodendorf, 2006, S. 154-156], [Beierle et al., 2008, S. 18]. Die Wissensrepräsentation der Artefakte in der Wissenbasis und die eingesetzte Wissensverarbeitung hängen dabei eng zusammen. Es werden in diesem Zusammenhang z.B. logische, relgelbasierte, fallbasierte [Bodendorf, 2006], [Beierle et al., 2008] oder semantische Technologien [Dengel, 2012] eingesetzt. Dies kann bis zu Ansätzen aus dem Themenbereich der künstlichen Intelligenz [Bodendorf, 2006, S. 177ff.] führen. Innerhalb dieses Beitrags soll daher beleuchtet werden, inwiefern eine Kombination aus semantischen und inferenzbasierten Ansätzen eine Aufwertung eines UIS hin zu einem “Umweltwissenssystem“ ermöglicht.

2 Semantische Wissensrepräsentation von Umweltinformationen

Die semantische Repräsentation von Wissen bedeutet immer Anreicherung der vorhandenen Daten um Meta-Daten (z.B. zu einem Messwert Standort, Erfassungsdatum, Bedingungen, …), in expliziter Form oder durch intrinsische Meta-Daten, welche z.B. als Schlüsselbegriffe bereits in dem Text vorhanden sind [Dengel, 2014, S. 13-15]. Durch eine entsprechende Modellierung z.B. als Baumstruktur (XML[16]) oder Graphstruktur (RDF[17], RDFS[18] oder OWL[19]), werden die Datenfragmente formal miteinander in

[16] W3C, XML, http://www.w3.org/standards/xml/core, Abruf 07.03.2015

[17] W3C, RDF 1.1 Semantics, http://www.w3.org/TR/rdf11-mt, Abruf 07.03.2015

Kontext gesetzt, wobei RDF- bzw. Ontologieansätze insbesondere zum Wissensaustausch Anwendung finden [Dengel, 2012, S. 132], [Hitzler et al., 2008, S. 36-37]. RDF, kurz für Ressource Description Framework, bildet in seiner einfachsten Form Verbindungen (Graphstrukturen) zwischen zwei Ressourcen und ist so in der Lage Zusammenhänge zwischen (Web-) Objekten zu beschreiben. Die Verbindung wird durch eindeutige Bezeichner, sogenannte URIs realisliert. Es lassen sich in Kombination mit RDF Schema einfache Ontologien bilden, welche ausgehend von ihrer Definition ein in sich abgeschlossenes Themengebiet oder eine Domain konzeptuell aufbereiten. Nach Dengel bilden Ontologien „*Wörterbücher einer umfassenden, vielsprachigen und erdumspannenden Informationswelt und sind damit wesentliche Grundlage für ein automatisches Verständnis durch Semantische Technologien*“ [Dengel, 2012, S. 70]. Dieser Gedanke wird etwa zur Umsetzung von Linked Open Data[20] aufgegriffen, welches den Anspruch hat, eine maschinenlesbare Aufbereitung und Vernetzung von Daten im Internet zu ermöglichen [Dengel, 2012, S. 183]. Um nun innerhalb dieser Strukturen zu navigieren und Abfragen durchzuführen, wird die SPARQL[21] Abfragesprache eingesetzt, welche sich von ihrer Syntax her an die Datenbankabfragesprache SQL anlehnt, aber auf RDF-Strukturen ausgeführt wird.

Neben den angsprochenen Graphstrukturen werden u.a. Taxonomien oder Thesauren (z.B. GEMET[22] oder UMTHES[23]) im Bereich der Umweltinformatik eingesetzt. Sie unterstützen bei der eigentlichen Definition und Ordnung (z.B. hierarchisch) von Begriffen und deren Synonymen. Bei beiden genannten Thesauren erfolgt die Repräsentation des Angebots mittlerweile ebenso per RDF Syntax. Mit Hinblick auf die Open (Government) Data Ansätze [Abecker et al., 2013] und die zuvor adressierte und durch die EU geforderte Veröffentlichung von Geo- oder Umweltinformationen ergeben sich durch eine

[18] W3C, RDF Schema, http://www.w3.org/TR/2014/REC-rdf-schema-20140225/, Abruf 07.03.2015

[19] W3C, OWL, http://www.w3.org/TR/owl2-overview/, Abruf 07.03.2015

[20] W3C, Linked Open Data: http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData, Abruf 06.03.2015

[21] W3C, SPARQL: http://www.w3.org/TR/sparql11-query/, Abruf 07.03.2015

[22] European Environment Information and Observation Network, GEMET, GEneral Multilingual Environmental Thesaurus, http://www.eionet.europa.eu/gemet, Abruf 06.03.2015

[23] Umweltbundesamt, UMTHES, http://data.uba.de/umt/de.html, Abruf 06.03.2015

semantische Anreicherung der Daten verschiedene Vorteile, wie etwa eine Interoperabiliät durch den Einsatz von vordefinierten Meta-Strukturen, Datenvernetzung, Möglichkeiten zur standardisierten Publikation und dem Austausch von umweltbezogenem Wissen sowie Abfrage- und Inferenzmöglichkeiten z.B. mit Hilfe von Regeln. Als Beispiel könnten Messwerte für einen Standort entsprechend um Meta-Informationen angereichert und miteinander vernetzt werden, um diese dann zum Austausch bereitzustellen. Eine Vernetzung und Austausch von UI zwischen nationaler und europäischer Ebene wird somit standardisiert und verbessert möglich und unterstützt die Erfassung von grenzübergreifenden Problemen wie etwa Hochwasser oder Stürme. Auch auf nationaler Ebene ergibt sich der Vorteil, dass durch die einmalige Erfassung und Aufbereitung von bestimmten Umweltinformationen, diese in unterschiedlichen UIS (z.B. auf Länderebene) weiterverwendet werden können. Es wird daher Redundanz in der Beschaffung und Aufbereitung der Informationen vermieden und Arbeitsaufwände reduziert.

3 Architekturkonzept „UmweltWissensSystem“ (UmweltWiS)

Ein integratives Konzept auf der Basis von Vorarbeiten zu UIS, WBS und semantischen Technologien wird im Folgenden als „Umweltwissenssystem“ (UmweltWiS) dargeboten. Abbildung 1 zeigt schematisch dessen Architektur, angelehnt an die in der Theorie beschriebenen Modelle für WBS. Dabei werden die Aspekte Linked Open Data Import und Export sowie semantische Repräsentation und Vernetzung der Datenfragmente in das Zentrum der Betrachtung gestellt.

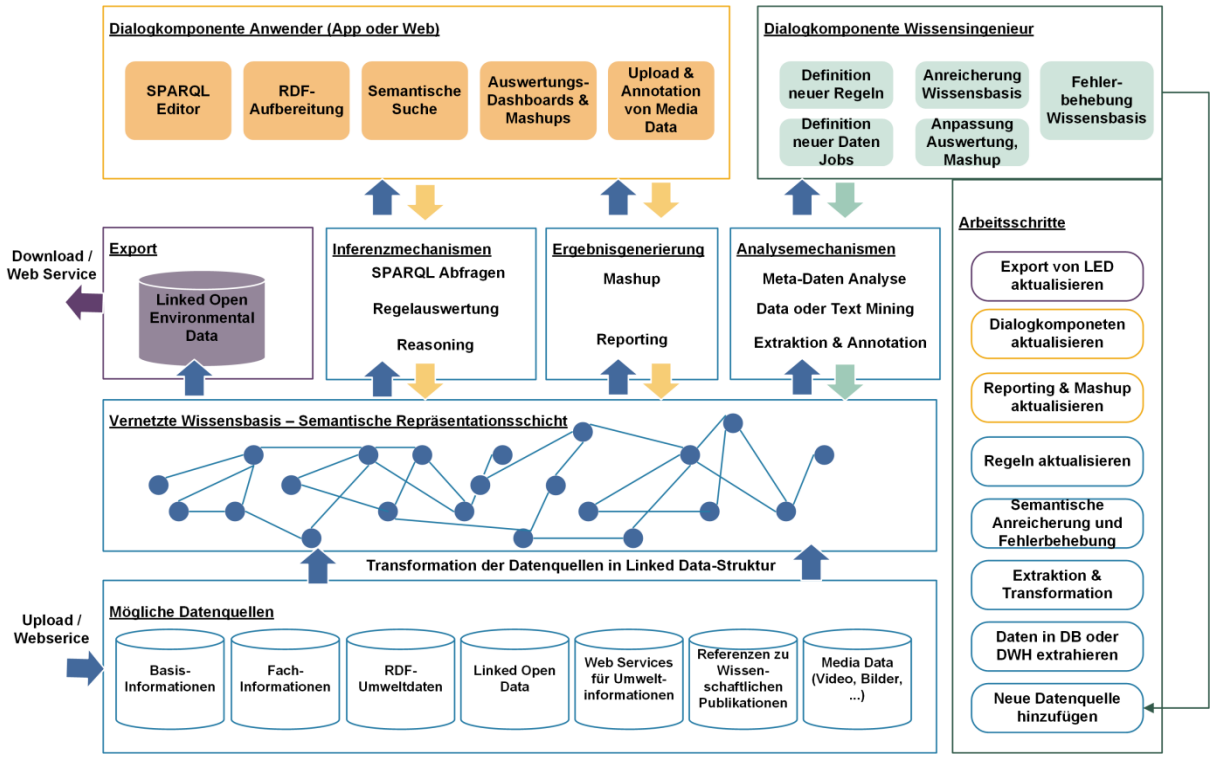


Abbildung 1: Architekturschaubild Umweltwissenssystem

Die Architektur lehnt sich dabei leicht an die von Faerber et al. vorgestellte Architektur eines ontologiebasierten KnoweldgeNavigator [Faerber, 2009, S. 23] und den Lebenszyklus von Linked Data an. Im Sinne dieses Lebenszykluses werden die Phasen: Extraktion, Speicherung und Abfrage, Manuelle Revision, Verlinkung, Klassifikation/Anreicherung, Qualitätsanalyse, Evolution/Fehlerbehebung und Suche bzw. Exploration der verknüpften Daten durchlaufen [Auer et al., 2011, S. 2]. Es werden verschiedene *Datenquellen* angebunden, welche semantisch vorstrukturiert sein können (z.B. Linked Open Data Sammlungen, als RDF oder XML repräsentierte Umweltinformationen) (1), die wie auch bei UIS aus bereits existierenden Basis- oder Fachinformationssystemen bereitgestellt und zugeführt werden (2), die über Web Services von anderen (öffentlichen) Angeboten[24] abgefragt werden (3), aber auch thematisch zugehörige wissenschaftliche Publikationen im Umweltbereich (4), Medien wie Bilder, Videos, Diagramme (5) oder Feedback der Anwender durch Social Media Features der Dialogkomponente (6). Es ist hierbei zu beachten, dass die bereitgestellten Daten nicht in gleicher Weise strukturiert sind und somit ein erster Schritt die Bildung einer

[24] z.B. Pegelonline, Gewässerkundliches Informationssystem, https://www.pegelonline.wsv.de, Abruf 06.03.2015

einheitlichen Datenstruktur ist. Zudem ist anzumerken, dass es sich nur um Beispielquellen handelt und das Konzept so ausgelegt ist, dass es nicht nur zur Etablierung eines öffentlichen UmweltWiS eingesetzt werden kann, sondern auch zur Umsetzung eines betrieblichen UmweltWis. Im Falle eines betrieblichen Systems wären Messungen aus Betriebsvorgängen, Stoffstrom-analysen, Prozessen oder Industrie 4.0 Komponenten von Interesse. Hier wäre es dann auch möglich Daten aus bestehenden Systemen via Web-Service direkt anzubinden und in die Betrachtung zu integrieren.

Über die *semantische Wissensrepräsentation* mittels RDF und RDFS ist es möglich die unterschiedlichen Datenquellen zu vereinheitlichen und miteinander zu vernetzen. Auf diese Weise wird der Forderung nach dem Einsatz von standardisierten Formaten für Linked Data von vornherein Rechnung getragen [Berners-Lee, 2006], [Auer et al., 2011, S. 5]. Insgesamt ist bei der Entwicklung frühzeitig die semantische Struktur der Wissensbasis zu erarbeiten, da diese später entscheidend für die Struktur der internen Datenbank und die Vernetzung der Quellen sein wird. Diese muss dabei gleichzeitig flexibel und strukturiert sein, so dass im späteren Verlauf durch die Definition von neuen Data Jobs einfach zusätzliche Datenquellen extrahiert und integriert werden können, bzw. eine passende Erweiterung der Struktur möglich ist. Betrachtet man die Verfahren, welche bei LOD zur Generierung von semantischen Strukturen aus unstrukturierten oder semi-strukturierten Daten verwendet werden, so existieren folgende Möglichkeiten: (1) Konvertierung eines kompletten Datendumps in RDF-Strukturen, auch wenn es hierbei zu falschen Zuordnungen kommen kann, (2) Online-Konvertierung bei Anfrage eines speziellen Datensatzes in RDF oder (3) Spiegelung von ganzen Datensammlungen in ein Linked-Data fähiges System, wie etwa einem Semantic Media Wiki und damit Durchführung einer impliziten Transformation [Dengel, 2012, S. 188-190]. Mittels Extraktionstechniken und Mapping ist so die Bildung einer RDF-Struktur möglich [Auer et al., 2011]. Der Extraktionsvorgang kann dabei abhängig von der Datenquelle durch eine Erkennung von Mustern, Entitäten, Fakten oder durch die Spiegelung auf die vorgegebene Ontologie-Struktur stattfinden [Dengel, 2012, S. 213-222]. Im Hinblick auf die Analyse bieten wissensbasierte Ansätze wie Text oder Data Mining die Möglichkeit zusätzliches Wissen und Zusammenhänge aus vorhandenen

Publikationen zu extrahieren und in die Struktur einzufügen. Eine manuelle Nacharbeit, Korrektur oder semantische Anreicherung der zugeordneten Daten ist dabei ggf. durch einen Wissensingenieur notwendig. Des Weiteren wird dem Anwender im Zuge von Social Media Aktivitäten die Möglichkeit gegeben, bestimmte Umweltinformationen in die Datenbasis zu laden, mit der Struktur zu verknüpfen oder semantisch zu annotieren. Praktisch könnte dies z.B. durch die Erstellung eines Fotos von einem bestimmten aktuellen Sachverhalt (z.B. Beobachtung Hochwasser) oder der Kommentierung eines bereits vorhandenen Beitrags erfolgen. Der mobile Sektor ist dabei bereits für UIS ein Zukunftsthema, welches u.a. durch den Landesumweltportal-Zusammenschluss LUPO adressiert wird [Schlachter et al., 2014].

Die *Inferenz- und Analysemechanismen,* die auf die umgesetzte Struktur angewendet werden, müssen entsprechend konzipiert werden. RDF(S)-Strukturen werden in der Regel über SPARQL Abfragen durchsucht. Zusätzlich werden Reasoning-Mechanismen eingesetzt, die weitergehende Auswertungen und Empfehlung von Handlungsempfehlungen ermöglichen. Die Einbindung von weiteren Medien oder etwa RSS-Feeds reichert die Wissensbasis zusätzlich an. Die sich ergebende Auswertung wiederum soll die Generierung von Reports und ein Mashup, also ein Zusammenführen und Aufbereiten von Daten ermöglichen.

Die *Dialogkomponenten* teilen sich in die zwei Sichten für Anwender und Wissensingenieure. Der *Anwender* soll die Dialogkomponente möglichst intuitiv bedienen können. Es werden daher interaktive Elemente, unterstützte Abfragemöglichkeiten und passende RDF-Graph Darstellungen gewählt. Im Sinne des zuvor genannten „Mashups“ von Wissen, können die einzelnen Ergebnisse in Form von durch Wissen angereicherte Karten, Diagramme oder Netzen dargestellt und je nach Anfrage ausgewertet werden. Der versiertere Anwender kann durch einen passenden Editor[25] z.B. direkt mit SPARQL Abfragen arbeiten, während der „normale“ Anwender die Freitextsuche und/oder Eingabe von Regeln zur Abfrage nutzt, und dabei durch die Oberfläche unterstützt wird. Regeln basieren in diesem Zusammenhang auf logischen Zusammenhängen. In Zeiten

[25] z.B. in Anlehnung an SPARQL Client unter http://sparqlclient.eionet.europa.eu, Abruf 07.03.2015

von Social Media ist der Anwender zudem darauf bedacht seine Meinung einzubringen. Daher werden in der Architektur Möglichkeiten zum Upload von z.B. Bildern, eigenen Messwerten (z.B. durch die Bereitstellung von strukturierten monatlichen Messwerten aus privaten Messstationen) oder der Kommentierung von vorhandenen Umweltereignissen oder Auswertungen hinzugefügt. Die *Expertensicht* ermöglicht die Anlage neuer Data Jobs, die Definition der Extraktionsmethoden, die Bearbeitung, Anreicherung oder Korrektur der vorhandenen Wissensbasis und die Anpassung der Inferenzmöglichkeiten.

Abschließend wird im Sinne der Anwendung ein Export von Linked (Open) Environment Data (LED[26]) im RDF-Format angeboten, wobei nur die Daten ausgegeben werden, welche gemäß den genutzten Quellen auch frei weitergegeben werden dürfen. Dies kann entweder in Form von Datenkollektionen zum Download oder auch in Form von Webservices erfolgen. Welche Variante gewählt wird, hängt davon ab, wie zeitkritisch die LED sind.

Die vorgestellte Architektur ist aktuell in der Konzeptionsphase, so dass an dieser Stelle bislang nur Umsetzungspotentiale dargestellt werden können. Eine Vorarbeit zum Einsatz von semantikbasierten Technologien in UIS wurde durch [Soori, 2015] erarbeitet und als Beispiel Use Case realisiert. Die aktuell in der Konzeption bedachten Anwendungsfälle erstrecken sich auf den öffentlichen und den betrieblichen Sektor. Im öffentlichen Bereich wäre es denkbar eine Zusammenführung von umweltrelevantem Wissen zu bestimmten Orten oder Regionen umzusetzen. Es könnten dann z.B. „Normalwerte“ für Pegelstände, Ozonwerte, Temperaturen etc. hinterlegt werden und anhand von Abweichungen z.B. Hochwasser, Sturmwarnung, etc. Handlungsanweisungen ausgegeben werden. RSS Informationen, Bilder, Videos, Hintergrundinfomationen oder Kommentare könnten zugeordnet und in die Auswertung mit einebzogen werden. Würde man von einer längeren Laufzeit ausgehen, so ließen sich Zeitreihen, Jahresvergleiche etc. auswerten und als LED ausgeben. Im betrieblichen Umfeld wäre

[26] Informationen zu LED u.a. Umweltbundesamt, http://www.umweltbundesamt.de/themen/nachhaltigkeit-strategien-internationales/information-als-instrument/linked-environment-data, Abruf 08.03.2015

es denkbar dieses zur Unterstützung der Erstellung einer Ökobilanz einzusetzen und anhand von produkt- oder abteilungsbezogenem Abfall- oder Chemikalienaufkommen auszuwerten. Ein Einsatz im Kontext von Industrie 4.0 Ansätzen, also eine Anreicherung der Wissensbasis durch ermittelte Sensorwerte ist ebenso denkbar.

4 Fazit & Ausblick

In dem vorgestellten Beitrag wurde der Vergleich zwischen Umweltinformationssystemen und einem Ansatz zu einem Umweltwissenssystemen „UmweltWiS“ auf Basis von Methoden aus dem Bereich der Wissensbasierten Systeme und dem Wissensmanagement diskutiert und ein entsprechendes Konzept vorgestellt. Während Teile der Architektur so auch in UIS Anwendung finden, sind die Integration von unterschiedlichen (Umwelt-)Datenquellen in einer gemeinsamen RDF-Struktur und die Auswertung mittels Abfrage-, Analyse und Inferenzmethoden als zusätzlicher Nutzen der Anwendung zu sehen. Die nächsten Schritte werden in der Technologieauswahl und der Entwicklung eines Prototypen gesehen.

5 Literaturverzeichnis

[Abecker et al., 2013]

Abecker, Andreas; Heidmann, Carsten; Hofmann, Claus; Kazakos, Wassilios: Veröffentlichung von Umweltdaten als Open Data, In: Umweltbundesamt (Hrsg.), Umweltinformationssysteme - Wege zu Open Data - Mobile Dienste und Apps, 20. Workshop des Arbeitskreises "Umweltinformationssysteme", S. 67-83, 2014.

[Auer et al., 2011]

Auer, Sören; Lehmann, Jens; Ngonga Ngomo, Axel-Cyrille: Introduction to Linked Data and Its Lifecycle on the Web, In: A. Polleres et al. (Ed.), Reasoning Web 2011, LNCS 6848, Berlin: Springer, S. 1-75, 2011.

[Beierle et al., 2008]

Beierle, Christoph; Kern-Isberner, Gabriele: Methoden wissensbasierter Systeme – Grundlagen, Algorithmen, Anwendungen. Wiesbaden: Vieweg + Teubner, GWV Fachverlage GmbH, 2008.

[Berners-Lee, 2006]

Berners-Lee, Tim: Principles of Linked Data, http://www.w3.org/DesignIssues/LinkedData.html, 2006, Abruf 06.03.2015

[Bodendorf, 2006]

Bodendorf, Freimund: Daten- und Wissensmanagement. 2. Aufl., Berlin/Heidelberg: Springer, 2006.

[Dengel, 2012]

Dengel, Andreas [Hrsg.]: Semantische Technologien - Grundlagen - Konzepte – Anwendungen. Heidelberg: Spektrum Akademischer Verlag, 2012.

[EU Kommission, 2008]

Kommission der Europäischen Gemeinschaften: Mitteilung der Kommission an den Rat, das europäische Parlament, den europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen – Hin zu einem gemeinsamen Umweltinformationssystem (SEIS), http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52008DC0046&from=EN, Online 01.02.2008, Abruf, 08.03.2015.

[EU Kommission, 2013]

Europäische Kommission: Arbeitsunterlage der Kommissionsdienststellen – Gemeinsame Umweltinformationssystem der EU – (SEIS) – Perspektiven für die Umsetzung, http://ec.europa.eu/environment/seis/pdf/seis_implementation_de.pdf, Online 25.01.2013, Abruf 08.03.2015.

[Faerber et al., 2009]

Faerber, Matthias; Archne, Oliver; Jochaud, Florent; Jablonski, Stefan: "KnowledgeNavigator: Ontologiebasierte Datenintegration in der Umweltforschung, In: Umweltbundesamt (Hrsg.), Umweltinformationssysteme - Suchmaschinen und Wissensmanagement - Methoden und Instrumente, S. 21-30, 01/2009.

[Fischer-Stabel, 2005]

Fischer-Stabel, Peter [Hrsg.]. 2005. Umweltinformationssysteme - Grundlegende Konzepte und Anwendungen. 1. Aufl., Berlin: Wichmann, VDE Verlag GmbH, 2005.

[Hitzler et al., 2008]

Hitzler, Pascal, Krötsch, Markus; Rudolph, Sebastian; Sure, York: Semantic Web. Grundlagen. Berlin: Springer, 2008.

[Laudon, 2010]

Laudon, Kenneth C.; Laudon Jane P.; Schoder, Detlef: *Wirtschaftsinformatik* - Eine Einführung, 2. Aufl., München: Pearson, 2010.

[North, 2011]

North, Klaus: Wissensorientierte Unternehmensführung – Wertschöpfung durch Wissen, 5. Aufl., Wiesbaden: Gabler Verlag, Springer Fachmedien, 2011.

[Schlachter et al., 2014]

Schlachter, T.; Düpmeier, C.; Weidemann, R; Schillinger, W.; Nonnenmann, B; Rossi, R; Hibbeler, C.; Koch, L.; Dombeck, T.: LUPO mobil - Umweltdaten mobil: Konzepte und technologische Einblicke in die "Meine Umwelt"-App. In: Weissenbach, Kurt; Schillinger, Wolfgang; Weidemann, Rainer (Hrsg.): Moderne anwendungsorientierte Forschung und Entwicklung für Umweltinformationssysteme, F+E-Vorhaben MAF-UIS; UIS BW, Umweltinformationssystem Baden-Württemberg, Phase II, 2012/2014, S. 75-90, 2014

[Soori, 2015]

Soori, Hamid Ahmad: Konzeption und Umsetzung eines öffentlichen Umweltinformationssystems unter Verwendung semantikbasierter Technologien, Diplomarbeit, Lehrstuhl für Wissensbasierte Systeme und Wissensmanagement, Universität Siegen, 2015.

[Zsifkovits et al., 2012]

Zsifkovits, Helmut; Brunner, Uwe: Konzeption und Planung von Umweltinformationssystemen. In: Tschandl, Martin; Posch, Alfred: *Integriertes Umweltcontrolling - Von der Stoffstromanalyse zum Bewertungs- und Informationssystem*. Wiesbaden: Gabler Verlag, Springer Fachmedien, 2012, S. 232-253

Energieatlas Schleswig-Holstein

Friedhelm Hosenfeld
Institut für Digitale Systemanalyse & Landschaftsdiagnose (DigSyLand),
hosenfeld@digsyland.de

Malte Albrecht
Ministerium für Energiewende, Landwirtschaft, Umwelt und ländliche Räume
des Landes Schleswig-Holstein (MELUR-SH),
Malte.Albrecht@melur.landsh.de

Abstract

The energy atlas Schleswig-Holstein represents an instrument supporting the state authorities in performing their tasks in order to promote the energy transition. The web application energy atlas makes governmental data about wind turbines and biogas plants available combined with comprehensive evaluation functions including digital maps presenting the energy plants in the context of other thematic layers like land use. The energy atlas is one of the first applications integrated in a special area of a new data center, observing new challenges and options.

Kurzzusammenfassung

Vorgestellt wird der Energieatlas Schleswig-Holstein, in der ersten Ausbaustufe ein behördeninternes Instrument zur Identifizierung und Bewertung des Ausbaus der Erneuerbaren Energien in Schleswig-Holstein unter Verwendung vorliegender Daten.

In einer einfach zu bedienenden Web-Anwendung auf der Basis der Auswertungsplattform Disy CadenzaWeb werden Informationen zu Onshore-Windkraftanlagen und Biogasanlagen für Recherchen und geografische Auswertungen für die Landesbehörden Schleswig-Holsteins zugänglich gemacht.

Der Energieatlas ist eine der ersten Anwendungen des Umweltressorts, die im neuen Rechenzentrum RZ² des zentralen Landes-IT-Dienstleisters Dataport den Produktivbetrieb aufgenommen haben.

1 Überblick

1.1 Motivation und Grundlagen

Zur geeigneten Unterstützung der Energiewende in Schleswig-Holstein wurde ein Instrument benötigt, um Informationen, die bereits in behördlichen Informationssystemen über Energieanlagen vorliegen, den verantwortlichen Landesbehörden in geeigneter Form zu erschließen und zu präsentieren.

Mit dem Energieatlas Schleswig-Holstein sollen unter anderem folgende Informationen verfügbar gemacht werden:

- Weitere Potenziale zum Ausbau der Erneuerbaren Energien
- Konkurrenzen zwischen Erneuerbaren Energien und anderen flächenrelevanten Nutzungsformen (vgl. [Koldrack 2014; Lückehe et al. 2014])
- Weiterer Netzausbaubedarf
- Entwicklung der verschiedenen Energieformen in Schleswig-Holstein

In der Konzeption des Energieatlas wurden folgende Aspekte berücksichtigt:

- Nutzung vorhandener Informationen der Landesbehörden
- Aufbau eines flexibel erweiterbaren Werkzeugs
- Bereitstellung zunächst behördenintern aber ressortübergreifend (später auch für die Kommunen)

Der Energieatlas wurde geplant als ein Baustein zum Monitoring der Umsetzung der Energiewende in Schleswig-Holstein.

Der technisch-operative Betrieb sollte möglichst weitgehend beim IT-Dienstleister des Landes Dataport zentralisiert werden.

Die Federführung zur Realisierung des Energieatlas liegt beim Ministerium für Energiewende, Landwirtschaft, Umwelt und ländliche Räume (MELUR). Damit wurde der Energieatlas auch als Pilotanwendung zur Erprobung des Vorgehensmodells gemäß Data Warehouse-Rahmenkonzept MELUR konzipiert (vgl. [Hübner et al. 2013]).

1.2 Zeitliche Entwicklung

Bereits 2011 wurde in dem „Integrierten Energie- und Klimakonzept für Schleswig-Holstein“ (IEK) von der damaligen schleswig-holsteinischen Landesregierung der Anstoß für den Aufbau eines Energieatlas Schleswig-Holstein gegeben, der dann Anfang 2012 zur Konzeption eines entsprechenden Informationssystems führte.

Mit der verstärkten energiepolitischen Schwerpunktsetzung, die sich nach dem Regierungswechsel 2012 auch in der Erweiterung der Bezeichnung des Ministeriums für *Energiewende*, Landwirtschaft, Umwelt und ländliche Räume (MELUR) äußerte, wurde unter dessen Federführung gemeinsam mit der Staatskanzlei die Entwicklung eines Prototypen vorangetrieben. Begleitet von einer ressortübergreifenden Projektgruppe wurde die Firma DigSyLand mit der Entwicklung einer zunächst behördeninternen Web-Anwendung beauftragt, die perspektivisch bereits die Weiterentwicklung zu einer öffentlichen Informationsplattform und die dezentrale Erfassung von weiteren relevanten Daten berücksichtigen sollte.

Im Sommer 2012 wurde die erste prototypische Anwendung im Intranet des MELUR-Ressorts bereitgestellt und in den folgenen Monaten auf der Basis der Anforderungen der Projektgruppe weiterentwickelt.

Mit der Inbetriebnahme des neuen Dataport-Rechenzentrums RZ² wurde 2013 mit den Planungen begonnen, den Energieatlas im Rahmen des Projekts ZeBIS (Zentraler Betrieb der Informationssysteme im Geschäftsbereich des MELUR) in einer zu konzipierenden *MELUR-Servicearea* innerhalb des neuen Rechenzetrums zu realisieren.

Konzeption und Aufbau der MELUR-Servicearea wurden 2014 begonnen, so dass der Energieatlas als eine der ersten Anwendungen des MELUR in der MELUR-Servicearea den Produktivbetrieb aufnahm.

2 Informationsangebot des Energieatlas

2.1 Recherche nach Windkraft- und Biogasanlagen

Die Recherche nach Windkraft- und Biogasanlagen nach verschiedenen Kriterien bietet einen möglichen Einstieg in den Energieatlas. Die Kriterien umfassen verwaltungsrelevante Angaben wie Anlagenbezeichnungen, Betreiber, Genehmigungs- und Inbetriebnahmedaten, Adressangaben aber auch anlagentypspezifische Daten: Bei Windkraftanlagen (WKA) können Leistungs- und Größenparameter abgefragt werden, bei den Biogasanlagen (BGA) z.B. die Feuerungswärmeleistung und die genehmigten Einsatzstoffe.

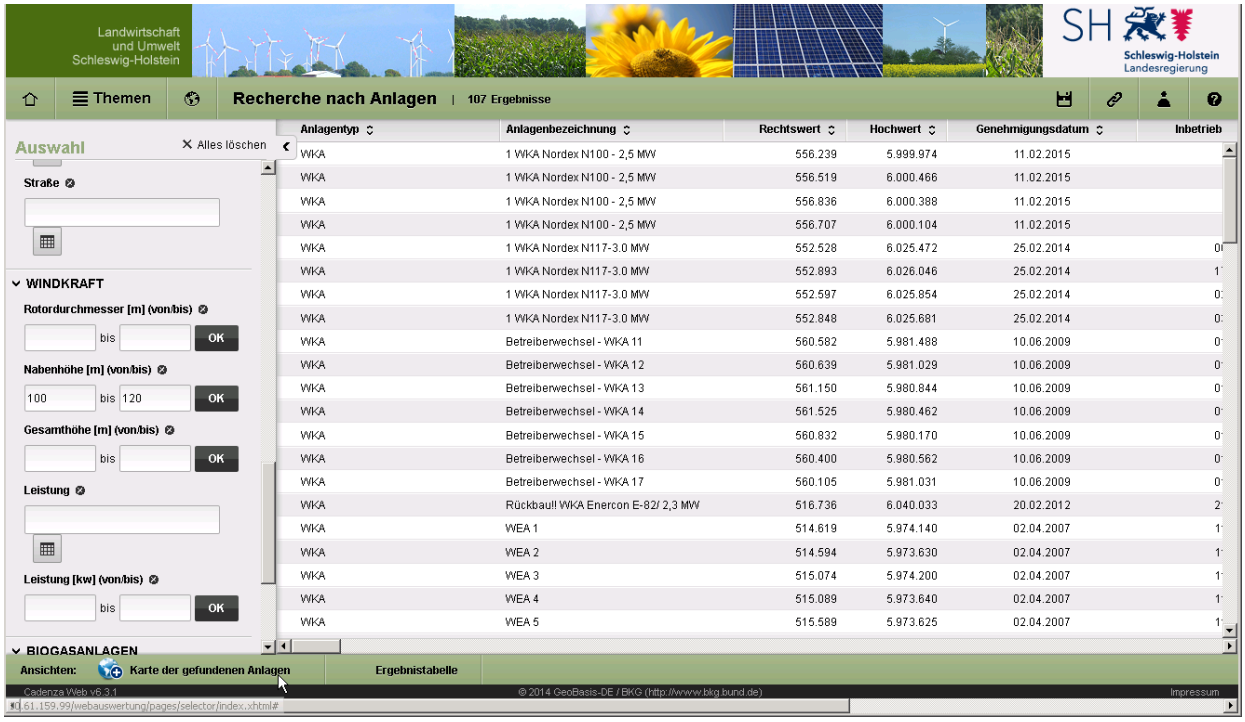


Abbildung 1: Recherche nach Windkraftanlagen mit unterschiedlichen Kriterien

Als Standardfunktionalität von CadenzaWeb werden bereits bei der Zusammenstellung der Recherchekriterien die gefundenen Ergebnisse tabellarisch ausgegeben (s. Abb 1).

Die gefundenen Tabellendaten können als Excel-Datei zur weiteren Bearbeitung und Analyse abgespeichert werden. Zudem können für durchgeführte Recherchen sogenannte Permalinks vergeben werden: Web-Adressen (URLs), die später wieder aufgerufen und auch an Dritte weitergegeben werden können.

Abbildung 2: Visualisierung der durch Recherche ermittelten Anlagen auf der Karte in CadenzaWeb

Besonders wichtig ist die Darstellung der gefundenen Energieanlagen auf der Karte, um sie dort bei Bedarf mit anderen Themen zu kombinieren und andere flächenrelevante Nutzungsformen zu visualisieren (siehe Abb. 2).

Informationen über die dargestellten Anlagen können per Mausklick bequem abgefragt werden.

2.2 Kartenthemen

Verschiedene fachliche und Verwaltungs-Kartenthemen stehen neben den Standorten der Energieanlagen zur Verfügung, um diese mit Rechercheergebnissen zu kombinieren oder als alternativen Einstieg über die Karten zu nutzen.

Dazu zählen Verwaltungseinheiten wie Gemarkungen und Flurstücke, die per WMS vom Landesamt für Vermessung und Geoinformation Schleswig-Holstein (LVermGeo-SH) eingebunden werden, ebenso wie ausgewiesene Windeignungsgebiete, Natura-2000-Schutzgebiete und nicht-beihilfefähige Punkt-Geometrien der Landwirtschaft (s. Abb. 3).

Abbildung 3: Präsentation von energierelevanten Kartenthemen mit Recherchefunktionen

2.3 Auswertungskarten

Ergänzt werden die Hintergrund-Kartenthemen durch sogenannte Auswertungskarten, die durch Aggregation von anlagenbezogenen Parametern auf den räumlichen Einheiten *Gemeinde* und *Kreis* gebildet werden. Die Visualisierung umfasst beispielsweise die Anzahl der Windkraftanlagen pro Gemeinde und wird durch unterschiedliche Größenklassen farblich symbolisiert (s. Abb. 4).

Die Zusammenstellung der räumlich aggregierten Daten findet im Rahmen des ETL-Prozesses im Rahmen der nächtlichen Aktualisierung des Datenbestandes statt, so dass die präsentierten Informationen immer konsistent sind (s. Abb. 5).

3 Technische Grundlagen

3.1 Konzeptionelle Vorgaben

Für den ersten Schritt zum Aufbau eines Energieatlas sollte soweit wie möglich auf bereits bei den Landesbehörden vorhandenen Daten und Softwaresystemen aufgebaut werden.

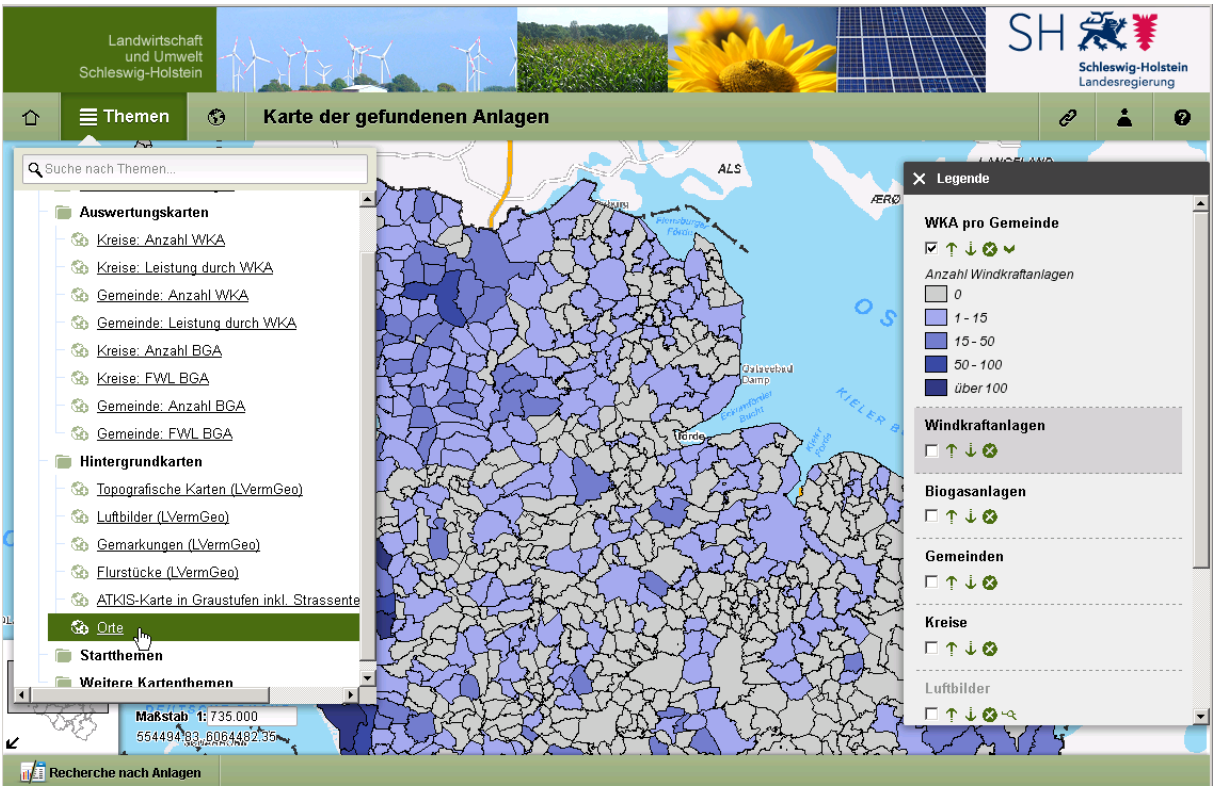


Abbildung 4: Auswertungskarten aggregieren Anlageninformationen auf Gemeinde- und Kreisebene

Die Konzeption sah vor, Daten der immissionsschutzrechtlich genehmigten Anlagen aus dem „**L**änder**I**nformations**S**ystem für **A**nlagen“ LIS-A zu nutzen. Als Genehmigungsbehörde verwaltet und pflegt das Landesamt für Landwirtschaft, Umwelt und ländliche Räume (LLUR) die benötigten Angaben in LIS-A. Zu den für den Energieatlas relevanten Daten zählen die Windenergieanlagen an Land (mit einer Höhe ab 50 m und einige weitere) und Biomasse-Anlagen.

3.2 Auswertungsplattform Disy Cadenza

Als Plattform zur Realisierung des Energieatlas als Web-Anwendung wurde Cadenza Web festgelegt. Disy Cadenza wird im Landwirtschafts- und Umweltressort in mehreren Fachabteilungen bereits seit vielen Jahren erfolgreich zur Auswertung von Umweltdaten eingesetzt, so dass sowohl auf die technischen Grundlagen als auch das erforderliche Know How aufgebaut werden konnte. Unter anderem unterstützt Cadenza die Erfüllung von Auswertungsaufgaben und Berichtspflichten für die EG Wasserrahmenrichtlinie (WRRL) [Hosenfeld et al 2012a] sowie das Management von Naturschutz-

maßnahmen [Hosenfeld et al 2012b] und die Umsetzung der Umgebungslärmrichtlinie in Schleswig-Holstein [Hosenfeld et al. 2014].

Cadenza bietet eine Plattform, die einerseits als Java-basierte Desktop-Anwendung *Cadenza Professional* verfügbar ist und die Fachleute in die Lage versetzt, kombinierte Auswertungen von Sach- und Geodaten mit dem Bedienungskomfort und der Funktionsvielfalt einer typischen Desktop-Anwendung durchzuführen. Bei den schleswig-holsteinischen Landesbehörden kommt die Java-Webstart-Technik zum Einsatz, um den Installationsaufwand zu minimieren.

Andererseits kann die zweite Variante *CadenzaWeb* genutzt werden, um Cadenza-basierte Web-Anwendungen aufzubauen. CadenzaWeb enthält ebenfalls umfangreiche Auswertungs- und Präsentationsfunktionen, ist aber im Vergleich zur Desktop-Version eingeschränkt, so dass CadenzaWeb insbesondere für die Adressierung von Nicht-Fachleuten gut geeignet ist.

Beide Cadenza-Varianten greifen auf die gleichen Datenquellen und Verwaltungsstrukturen, das Cadenza-Repository, zu. Das Verwaltungswerkzeug Repository Manager dient zur einheitlichen Administration der Cadenza-Plattform.

Für die Umsetzung des Energieatlas wurde CadenzaWeb gewählt, da Mitarbeitende aller Landesbehörden die Anwendung so ohne zusätzlichen Installationsaufwand mit ihrem Standard-Web-Browser nutzen können (s. Abb. 5 und 6).

3.3 Datenmanagement

Die Datenhaltung findet im Oracle-RDBMS statt, das als Standard innerhalb der Landesverwaltung eingesetzt wird.

Cadenza bietet den integrierten Zugang zu Sachdaten und Geodaten, die in Oracle in den Standard-Oracle-Formaten für räumliche Daten gehalten werden. Zusätzlich können Geodaten in Shape-Dateien und durch OGC-Dienste wie WMS und WFS bereitgestellte Daten eingebunden werden (s. Abb. 6).

3.4 ETL-Prozesse zur Datenzusammenstellung

In enger Abstimmung mit dem Fachdezernat und dem IT-Dezernat im LLUR wurden zur Übernahme der relevanten Daten aus LIS-A ETL-Prozesse (ETL: Extraktion, Transformation, Laden) umgesetzt, die die Daten in einer für den Energieatlas verwendbaren Form in einer Staging-Datenbank innerhalb des LLUR zusammenstellen (s. Abb. 5). Die ETL-Prozesse umfassen auch die Umsetzung der in LIS-A als Sachdaten gehaltenen Koordinatenangaben in Oracle-Geometrien und die räumliche Aggregation bestimmter Anlageninformationen als Basis für die Auswertungskarten (vgl. Kap. *2.3*). Mit der nächtlichen Ausführung der ETL-Prozesse stehen im Energieatlas immer die aktuellen Daten zur Präsentation und Recherche zur Verfügung.

Mit Hilfe einer Cadenza-Test-Instanz innerhalb des LLUR ist der Zugriff auf die Staging-Datenbank und damit die fachliche Prüfung der präsentierten Informationen innerhalb des LLUR und anderer Behörden des Ressorts möglich (s. Abb. 5).

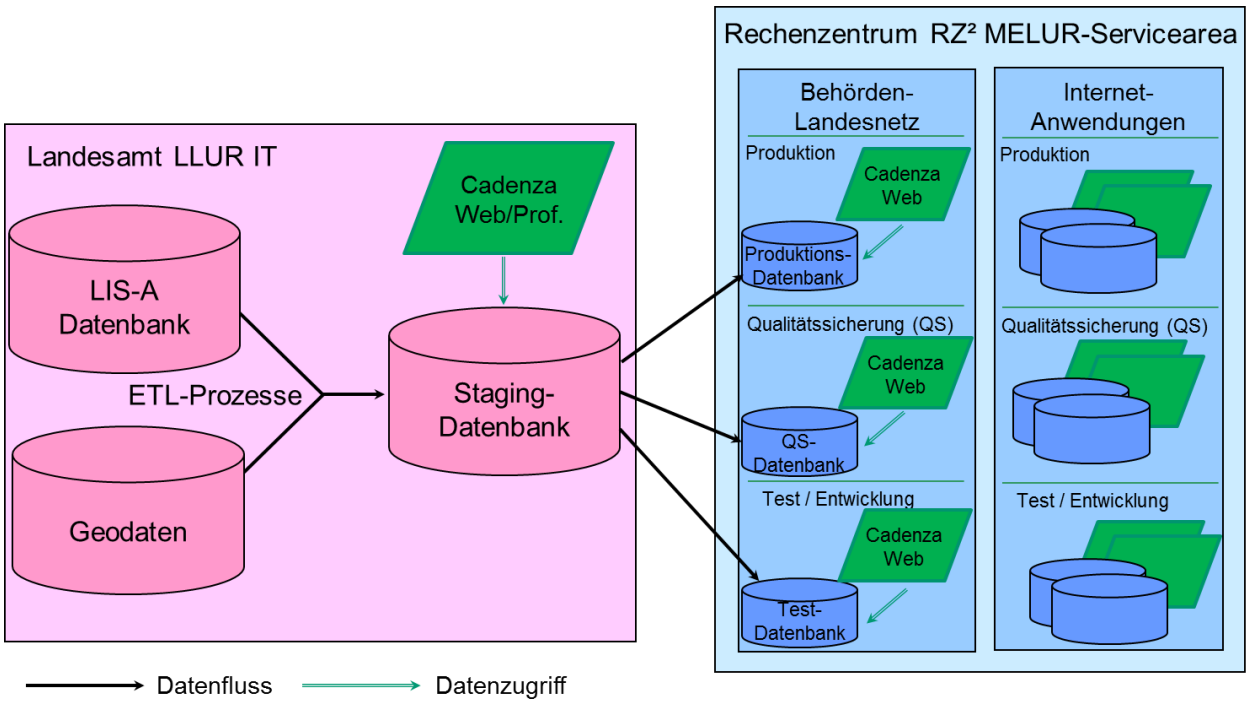


Abbildung 5: Für den Energieatlas umgesetzter Datenfluss in das Rechenzentrum RZ²

3.5 Ressort-übergreifende Bereitstellung im Rechenzentrum RZ²

Bereits die Konzeption des Energieatlas sah die Bereitstellung der Anwendung bei Dataport, dem zentralen Dienstleister des Landes vor, da nicht nur das MELUR-Ressort,

sondern auch weitere Landesbehörden als Nutzer adressiert werden. Dazu zählt insbesondere der in der Staatskanzlei angesiedelte Bereich der Landesentwicklung und Rauminformation, der ebenfalls in der Projektgruppe des Energieatlas repräsentiert ist.

Der Energieatlas bildete eine der beiden ersten Anwendungen, die im Rahmen des Projekts ZeBIS auf einer zentrale Cadenza-Plattform innerhalb der MELUR-Servicearea des neuen Dataport-Rechenzentrums RZ² installiert wurde (vgl. [Hübner et al. 2013]).

Eine Herausforderung bestand darin, die hohen technischen und sicherheitsrelevanten Vorgaben des Rechenzentrums mit den Bedürfnissen des Umweltressorts, das flexibel auf aktuelle Anforderungen reagieren können und heterogene Datenstrukturen und –flüsse berücksichtigen muss, zur Deckung zu bringen.

In enger Abstimmung mit den beteiligten Firmen Disy und DigSyLand, Dataport und dem LLUR konnten geeignete Zugriffsmechanismen und Replikationsprozesse entwickelt werden, so dass die erforderlichen Daten jetzt nächtlich in das RZ²-Rechenzentrum repliziert werden und dort tagesaktuell für Recherchen zur Verfügung stehen.

Innerhalb der MELUR-Servicearea werden Internet-Anwendungen und nur Landesnetz-intern zugängliche Anwendungen in unterschiedlichen Zonen verwaltet, die durch Sicherheitsmechanismen wie Firewalls voneinander getrennt sind. Jede dieser Zonen ist in drei Umgebungen aufgeteilt (s. Abb. 5):

- Test- und Entwicklungsumgebung
- Qualitätssicherungsumgebung (QS)
- Produktionsumgebung

Die Test- und Entwicklungsumgebung ist auch für externe Entwickler mittels VPN (Virtual Private Network) zugänglich, so dass Anwendungen dort auf einer Integrationsplattform mit produktionsnahen Bedingungen getestet werden können, was insbesondere beim Einsatz unterschiedlicher Komponenten verschiedener Software-Entwickler von entscheidender Bedeutung ist.

Die beiden anderen Umgebungen werden vollständig von Dataport betrieben. Nur Anwendungen und Updates, die den QS-Prozess in der QS-Umgebung erfolgreich durchlaufen haben, werden in die Produktionsumgebung übernommen.

Zeitgleich mit dem von Disy entwickelten Informationssystem für die Hochwasserkarten Schleswig-Holstein, das für die Öffentlichkeit freigegeben ist, ging der Energieatlas als behördeninterne Anwendung Ende 2014 in den Produktivbetrieb.

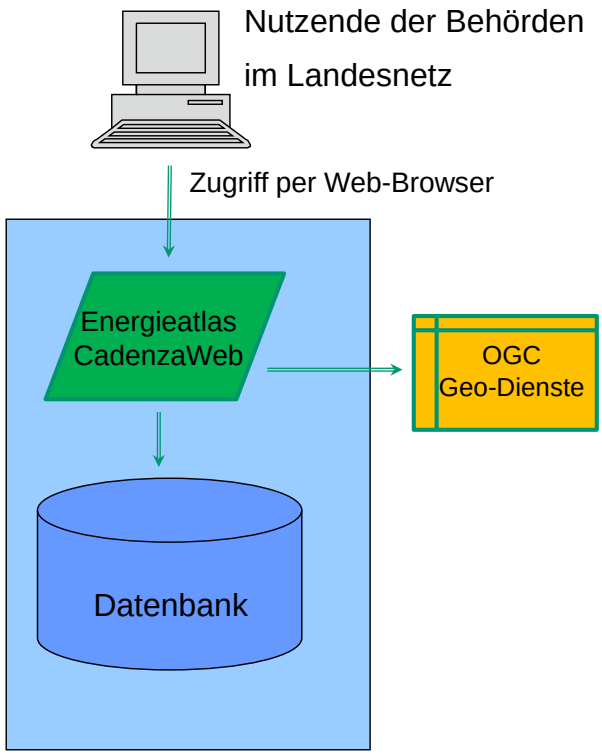


Abbildung 6: Die Landesbehörden können mittels Web-Browser auf den mit CadenzaWeb realisierten Energieatlas zugreifen

Diese Anwendungen bilden damit die Keimzelle der MELUR-Servicearea als erster Schritt zum Aufbau eines Data Warehouse für das MELUR-Ressort [Hübner et al. 2013]. Auf dieser Basis findet die konzeptionelle Weiterentwicklung der Lösungsarchitektur der MELUR-Servicearea statt.

4 Zusammenfassung und Ausblick

Mit dem Energieatlas wurde ein flexibel erweiterbares Instrument geschaffen, um den Landesbehörden Schleswig-Holsteins ressortübergreifend zu ermöglichen, durch den Genehmigungsprozess aufgenommene Daten aus dem Fachinformationssystem LIS-A zu Windkraftanlagen an Land und Biogasanlagen zu recherchieren, auszuwerten und zu visualisieren (s. Abb. 6).

Die Realisierung des Energieatlas bildet damit eine Pilotanwendung zur Erprobung des Vorgehensmodells gemäß Data Warehouse-Rahmenkonzept des MELUR, das als Beispiel für die Integration weiterer Verfahren in die MELUR-Servicearea dienen kann,

das den übergeordneten Zielen des Zentralen IT-Managements der Landesverwaltung SH (ZIT) dient.

Im Unterschied beispielsweise zum Potenzialatlas Erneuerbare Energien Baden-Württemberg [Müller 2013] legt der Energieatlas Schleswig-Holstein den Schwerpunkt zunächst auf die Auswertung von Daten über vorhandene Energieanlagen, während für Angaben zu weiteren Energiepotenzialen derzeit die Datenlage noch nicht ausreichend vorhanden ist.

Als weitere Ausbauschritte ist die Bereitstellung einer Version für die Öffentlichkeit und damit die Integration in in das Energiewendeportal des Landes (http://www.schleswig-holstein.de/DE/Schwerpunkte/Energiewende/energiewende_node.html) geplant.

Zudem sollen Daten, die von den Kommunen (Kreise und kreisfreie Städte) im Rahmen ihrer Genehmigungsprozesse, beispielsweise für kleinere Windkraftanlagen erhoben werden, mittelfristig in den Energieatlas einfließen. Ebenso wird die Integration von Netzdaten für die Zukunft konzipiert.

5 Literaturverzeichnis

[Hübner et al. 2013]

Hübner, H.; Kumer, D.; Albrecht, M. & Kazakos, W. (2013): Konzeption und Aufbau eines Data Warehouses für die Umweltverwaltung Schleswig-Holstein. Vortrag beim 20. Workshop des Arbeitskreises „Umweltinformationssysteme“ der Fachgruppe „Informatik im Umweltschutz“, veranstaltet an der Hochschule für Technik und Wirtschaft in Berlin in Zusammenarbeit mit dem Umweltbundesamt am 23. und 24. Mai 2013. http://www.ak-uis.de/ws2013/Vortraege.zip (do12-Huebner-DwhMELUR.pdf)

[Hosenfeld et al. 2012a]

Hosenfeld, F.; Lietz, J. & Trepel, M. (2012): Entwicklung einer Bewertungsdatenbank zur Umsetzung der Wasserrahmenrichtlinie in Schleswig-Holstein. In: Umweltbundsamt: Umweltinformationssysteme - Informationsgewinnung und Datenaufbereitung für

maritime Informationssysteme, 22. Workshop des Arbeitskreises „Umweltinformationssysteme“ der Fachgruppe „Informatik im Umweltschutz“, Elsfleth, 2011, UBA Texte 03/12, pp. 69 - 78.

[Hosenfeld et al. 2012b]

Hosenfeld, F.; Kleinbub, J.; Krüger, K.; Kumer, D.; Langner, K. & Rinker, A. (2012): Development of a Database Supporting the Management of Nature Conservation Measures in Schleswig-Holstein. In: H.-K. Arndt, G. Knetsch, W. Pillmann (Eds.): Light up the Ideas of Environmental Informatics, Dessau 2012, pp. 623 – 629.

[Hosenfeld et al. 2014]

Hosenfeld, F.; Gliessmann, L. & Rinker, A. (2014): Mapping Service Environmental Noise Schleswig-Holstein. In: Gomez, J.M.; Sonnenschein, M.; Vogel, U.; Winter, A.; Rapp, B. & Giesen, N. (Eds.): Information and Communication Technology for Energy Efficiency, EnviroInfo 2014 28th International Conference on Informatics for Environmental Protection, Oldenburg, pp. 493-500.

[Koldrack 2014]

Koldrack, N. (2014): Calculation of current land use for renewable energy. In: Gomez, J.M.; Sonnenschein, M.; Vogel, U.; Winter, A.; Rapp, B. & Giesen, N. (eds.): Information and Communication Technology for Energy Efficiency, EnviroInfo 2014 28th International Conference on Informatics for Environmental Protection, Oldenburg, pp 93 – 100.

[Lückehe et al. 2014]

Lückehe, D.; Kramer, O. & Weisensee M. (2014): An evolutionary approach to geo-planning of renewable energies. In: Gomez, J.M.; Sonnenschein, M.; Vogel, U.; Winter, A.; Rapp, B. & Giesen, N. (eds.): Information and Communication Technology for Energy Efficiency, EnviroInfo 2014 28th International Conference on Informatics for Environmental Protection, Oldenburg, pp. 501 – 508.

[Müller 2013]

Müller. M. (2013): Potenzialatlas Erneuerbare Energien Baden-Württemberg. In: Page, B.; Fleischer, A.G.; Göbel, J. & Wohlgemuth, V. (eds): Environmental Informatics and Renewable Energy. Proceedings of the 27th EnviroInfo Conference Hamburg 2013, pp. 55 – 61.

River Basin Information System - RBIS

Beispiel Vu Gia Thu Bon RBIS (Vietnam)

Franziska Zander, Dr. Sven Kralisch, Prof. Dr. Alexander Brenning
Friedrich-Schiller-Universität Jena, Institut für Geographie,Lehrstuhl für Geoinformatik
Franziska.Zander@uni-jena.de; Sven.Kralisch@uni-jena.de

Abstract

The Vu Gia Thu Bon RBIS (River Basin Information System) has been set up and further developed within the research project LUCCi (Land use and climate change interactions in central Vietnam; http://www.lucci-vietnam.info). It serves as project database to manage environmental basic and result data. The project itself was funded from 2010 to 2015 within the BMBF program “sustainable land management”.

The Vu Gia Thu Bon RBIS is based on the software platform RBIS (River Basin Information System), which has been developed at the Chair of Geographic Information Science at the Friedrich-Schiller-University of Jena in Germany for more than 13 years. During that time RBIS was applied in several research projects worldwide.

RBIS is a modular structured, web-based data and sharing platform with full read and write access. It is built on open source software and standards to support reuse and extensibility of the software. It aims to provide functions to describe, manage, analysis, visualize, link and present different type of environmental data (e.g. time series data, geodata, documents …) and the corresponding metadata. Due to the fine gained user and permission management and multilingual support (e.g. Vietnamese) RBIS can also serve as information system for local stakeholder and external scientists.

The software platform RBIS using the example Vu Gia Thu Bon RBIS will be briefly presented.

1 Einleitung

Das *Vu Gia Thu Bon RBIS* (River Basin Information System) wurde im Rahmen des Forschungsprojektes LUCCi (*Land use and climate change interactions in central Vietnam*; http://www.lucci-vietnam.info) eingerichtet und weiterentwickelt. Innerhalb des Projektes diente das *Vu Gia Thu Bon RBIS* als Projektdatenbank für Basis- und Ergebnisdaten. Das Projekt LUCCi wurde von 2010 bis 2015 innerhalb des BMBF-Förderschwerpunktes „Nachhaltiges Landmanagement“ gefördert. Dabei waren deutsche, internationale sowie auch vietnamesische Partner (Universitäten und Institute) beteiligt.

Das *Vu Gia Thu Bon RBIS* basiert auf der Softwareplattform RBIS (*River Basin Information System*), welche seit mehr als 13 Jahren am Lehrstuhl für Geoinformatik an der Friedrich-Schiller-Universität entwickelt wird und seitdem in zahlreichen Forschungsprojekten eingesetzt wurde.

RBIS ist eine modular strukturierte und webbasierte Datenverwaltungs- und Austauschplattform, die für Nutzer mit entsprechenden Rechten nicht nur lesenden, sondern auch schreibenden Zugriff bietet. Ziel von RBIS ist es nicht nur Daten verschiedenen Typs (z. Bsp. Zeitreihendaten, Geodaten, Dokumente…) zu verwalten, sondern auch Funktionen bereitzustellen um diese zu analysieren, visualisieren, verknüpfen und zu präsentieren. Dadurch kann RBIS nicht nur als Projektdatenbank und Informationssystem für Wissenschaftler im Projekt dienen, sondern auch für lokale Stakeholder und andere projektexterne Wissenschaftler. Zum Schutz vor unautorisierten Zugriffen und von gespeicherten Daten, die zu unveröffentlichten Qualifikationsarbeiten gehören oder anderen bestimmten Nutzungsbedingungen unterliegen, gibt es ein Nutzer- und Rechtmanagement.

2 Systemaufbau

RBIS baut auf Open Source Software auf, um eine leichte Erweiterbarkeit, aber auch eine Weitergabe des gesamten Systems mit Daten als VirtualBox Installation zu ermöglichen. Auf der Serverseite wird ein Standard-Linux-Web-Stack mit Apache-

Webserver, PHP, PostgreSQL (http://www.postgresql.org) zusammen mit der PostGIS-Erweiterung (http://www.postgis.org) für Geodaten-Unterstützung eingesetzt. Zudem kommen der MapServer (http://mapserver.org) und OpenLayers (http://openlayers.org) sowie JavaScript-Bibliotheken wie jQuery (http://jquery.com), die Erweiterung jQuery UI (http://jqueryui.com) und weitere zum Einsatz.

Das technische Kernstück der Anwendung beruht auf XML-Dateien, welche u.a. folgende Informationen für die jeweilige Art von Datensätzen beinhalten:

- Visualisierung und Gruppierung einzelner Elemente
- Information für das Anlegen und Ändern
- Festlegung der Filterfunktion in der Anwendung
- Sortierung in Auswahllisten, Übersetzung
- Verknüpfung von Datensätzen
- räumliche Lage und Verknüpfung mit der Karte
- Default Werte
- interne Abbildung der Daten in der Datenbank

Die Übersichtsansicht ist in Abbildung 1 dargestellt.

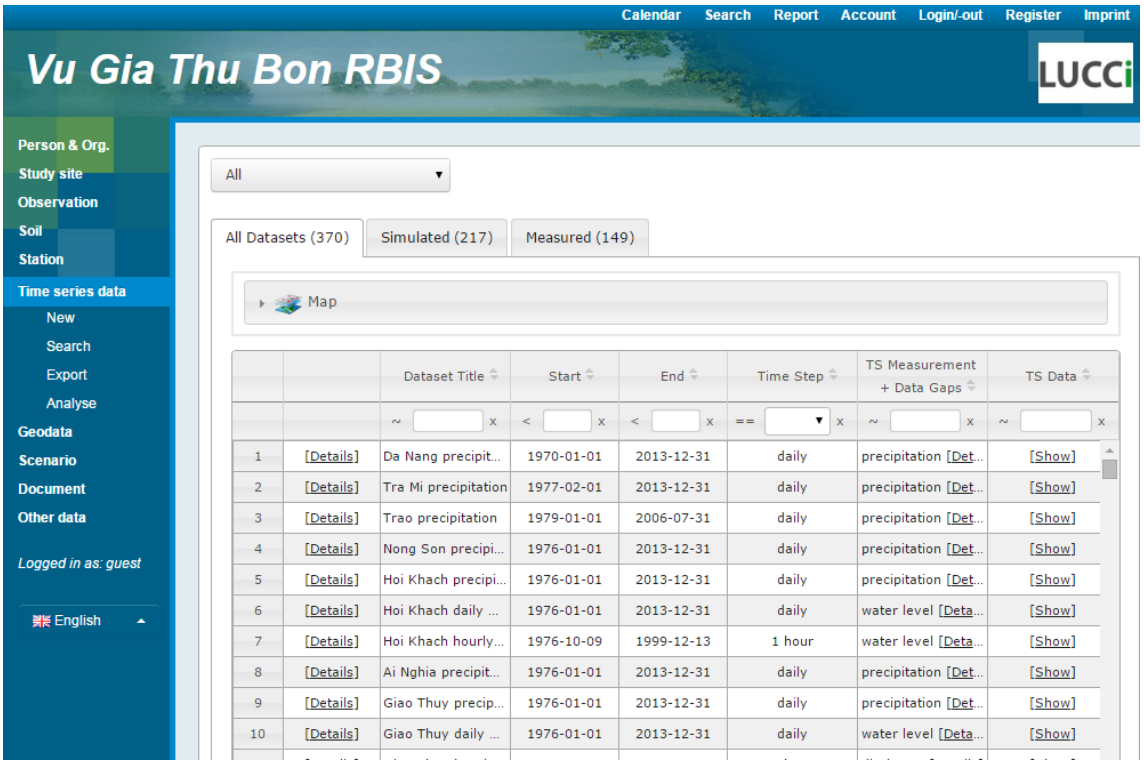


Abbildung 1: Screenshot Zeitreihenübersicht im Vu Gia Thu Bon RBIS
(Quelle: http://leutra.geogr.uni-jena.de/vgtbRBIS)

2.1 Module

RBIS ist modular aufgebaut, um den verschiedenen Anforderungen aus Projekten gerecht zu werden. Die im Fall vom Vu Gia Thu Bon aktivierten Module sind in Abbildung 1 in der linken Spalte als Menüeintrag zu sehen. Die Hauptmodule von RBIS sind in Abbildung 2 dargestellt, wobei der Fokus auf der Beschreibung der Daten mit Metadaten liegt (auch wenn die Daten selbst nicht im RBIS hinterlegt sind). Ausgehend von der Metadatenbeschreibung für Geodaten auf der Basis von ISO 19115 werden Elemente davon auch für die Beschreibung der anderen Datentypen wie zum Beispiel Zeitreihendaten verwendet. Neben der reinen Speicherung und Beschreibung von Datensätzen stellt RBIS für bestimmte Datentypen auch weitere Funktionen zur Verfügung. Ein Beispiel dafür sind die Zeitreihendaten, welche als Tabellen in der Datenbank verwaltet werden und somit eine Vielzahl zusätzlicher Möglichkeiten für eine Analyse bieten.

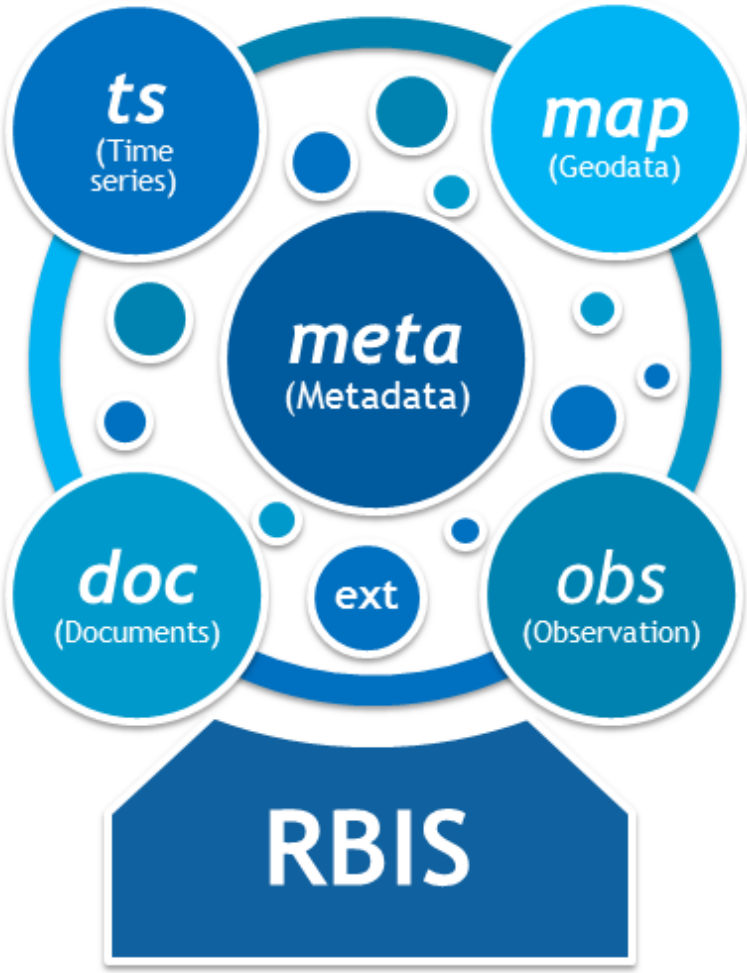


Abbildung 2: RBIS Module.

Dazu zählt neben der Visualisierung auch die Lückenerkennung sowieso die Möglichkeit diese mit einem regelbasierten Interpolations-Werkzeug zu füllen. Dabei werden die Informationen über verwendete Methode und Zeitreihen vollständig für jede einzelne Lücke erfasst.

An jeden RBIS Datensatz kann eine beliebige Anzahl von Dokumenten angehängt werden. Dokumente (Veröffentlichungen, Stationsbeschreibungen, …) oder Datensammlungen („Other data“) können jedoch auch einzeln mit Metadaten beschrieben und anschließend mit anderen Datensätzen verknüpft werden.

Ein Beispiel weiterer RBIS-Module ist ein Bodenmodul (Bodenprofile und Bodenhorizonte), welches auch im Vu Gia Thu Bon RBIS zum Einsatz kommt.

2.2 Suche

Um gespeicherte Information und Datensätze zu finden, gibt es globale und auf Datentypebene verschiedene Text und räumliche Sortier-, Such- und Filtermöglichkeiten. Die räumliche Suche kann auf alle Datensätze angewendet werden, die selbst oder über eine Verknüpfung entweder Koordinaten, Begrenzungskoordinaten oder eine Polygongeometrie besitzen. Zu den räumlichen Suchfunktionen gehört neben einer einfachen Suche mit einem Begrenzungsrechteck auch die Suche oder permanente Filterung über die abgespeicherten Untersuchungsgebiete (zum Beispiel Einzugsgebiete) mit oder ohne zusätzlichen Umkreissuche (10, 20 …100km)(Abbildung 3).

Suchergebnisse können als Permanentlink abgespeichert oder die entsprechenden Metadaten der Ergebnisdatensätze können als CSV-Datei exportiert werden.

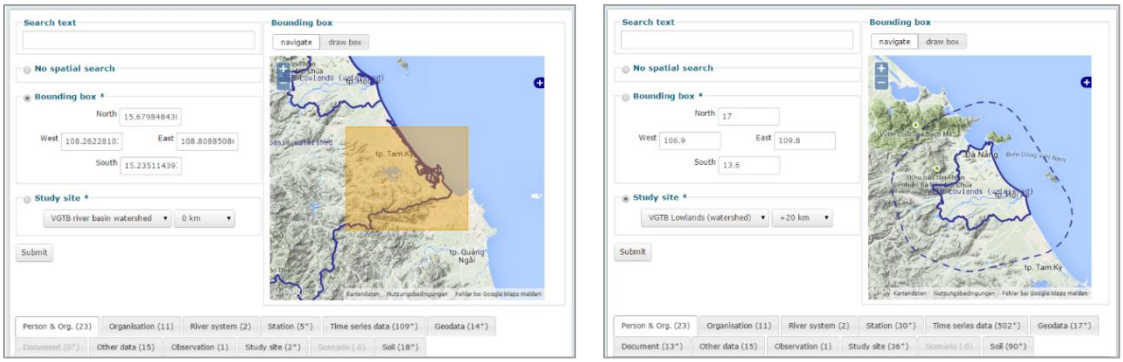


Abbildung 3: Globale Suche in RBIS.

2.3 Nutzer- und Rechteverwaltung

RBIS verfügt über eine Nutzer- und Rechteverwaltung, welche ursprünglich auf der Annahme beruhte, dass zumindest projektintern alle Daten für alle zugänglich sein sollten (Rechteverwaltung auf der Ebene von Datensatztypen). In der Praxis hat sich jedoch gezeigt, dass in einigen Fällen wie zum Beispiel bei unveröffentlichten Forschungsergebnissen auch eine Rechteverwaltung auf Datensatzebene durch den jeweiligen Eigentümer benötigt wird. Der Eigentümer kann bestimmen, welche Personen oder Nutzergruppen sich die Metadaten anschauen, verändern und angehängte Daten heruntergeladen dürfen.

In der Regel sollten Metadaten für alle lesbar sein (auch ohne Benutzerkonto). Dies hat den großen Vorteil, dass diese auch über Suchmaschinen gefunden werden können.

Anfragen für das Recht, bestimmte Datensätze herunterladen zu dürfen, können über die Anwendung gestellt werden. Der Eigentümer erhält eine Mail und der Nutzer wird über die positive oder negative Entscheidung mit einer Begründung oder Hinweis zum Umgang mit den Daten informiert.

2.4 Schnittstellen

RBIS besitzt nicht nur Schnittstellen, um Daten einfach und automatisiert importieren zu können, sondern bietet auch Möglichkeiten für externe Anwendungen auf die gespeicherten Informationen zuzugreifen bzw. zu durchsuchen. Hierzu zählt eine CSW Schnittstelle, welche im Rahmen der Kooperation mit dem begleitenden Projekt GLUES aus dem BMBF-Förderschwerpunkt „Nachhaltiges Landmanagement“ eingerichtet wurde. Alle im Vu Gia Thu Bon RBIS gespeicherten Metadaten zu Zeitreihen und Geodaten, die für einen Gastzugang sichtbar sind, werden über eine CSW Schnittstelle zur Verfügung gestellt. Die Umsetzung erfolgte mit pycsw (http://pycsw.org). Derzeitig gibt es 2 CSW Clients, welche auf diesen zugreifen.

Zum einen eine auf CSW-Abfragen basierende globale Suche über RBIS Installationen (http://leutra.geogr.uni-jena.de/RBISsearch) und zum anderen die im Rahmen von GLUES entwickelten Anwendungen wie GeoMetaFacet (http://geoportal.glues.geo.tu-dresden.de) (Abbildung 4).

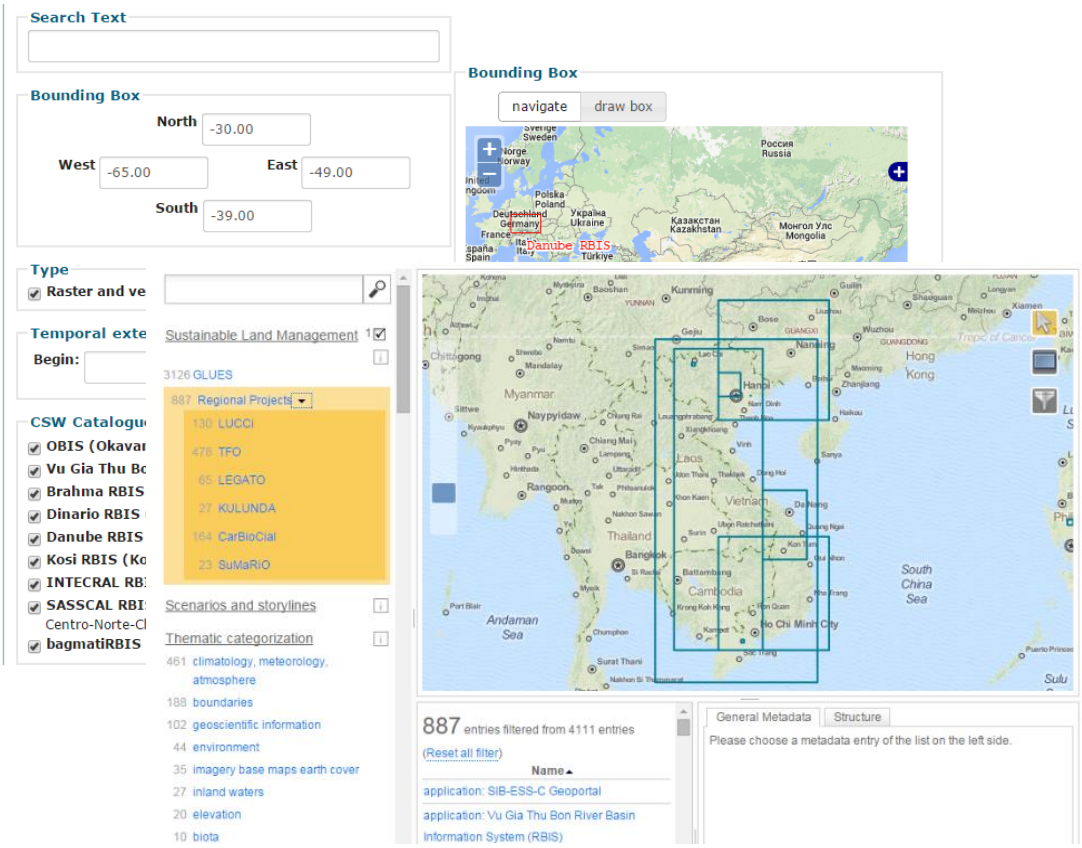


Abbildung 4: CSW-Client RBIS; GeoMetaFacet.

Neben den Schnittstellen, welche der im Kartenmodul von RBIS zum Einsatz kommende MapServer mit sich bringt (zum Beispiel WMS), gibt es auch Schnittstellen zu ebenfalls am Lehrstuhl entwickelten Softwareprodukten, mit deren Hilfe sich der komplette Workflow für integriertes Landmanagement abbilden lässt. RBIS dient hierbei als Datenplattform und ist damit zentraler Bestandteil der ILMS-Software (Integriertes Land Management System) [Kralisch et al. 2012].

3 Anwendungen von RBIS

Die verwendete RBIS-Software wird nicht nur in LUCCi eingesetzt, sondern auch in mehr als 20 weiteren Forschungsprojekten weltweit. Inhaltlich sind diese meist zu Arbeiten assoziiert, die sich mit IWRM (*Integrated Water Resources Management*) oder „*climate change impact assessment*“ beschäftigen. Dazu zählen Installationen für einzelne Promotionsvorhaben wie zum Beispiel das Oti RBIS (Togo, Benin) [Badjana et al., 2015], Jakarta RBIS (Indonesien) [Miga et al., 2013] und Kosi RBIS (Nepal) [Nepal,

2012], bilaterale Kooperationsprojekte (Seyhan RBIS (Türkei), Centro-Norte-Chile RBIS) oder multidisziplinäre Forschungsprojekte wie TFO (http://www.future-okavango.org) mit OBIS (Okavango Basin Information System) [Kralisch et al., 2013] oder SASSCAL (http://www.sasscal.org) mit dem SASSCAL-IS.

Darüber hinaus wird RBIS am ITT an der FH Köln im Rahmen des CNRD-Netzwerkes (http://www.tt.fh-koeln.de/cnrd/) als Plattform für Daten aus derzeit 8 verschiedenen Einzugsgebieten verwendet (http://rbis.itt.fh-koeln.de/).

4 Ausblick

Das Vu Gia Thu Bon RBIS soll am Projektende an die vietnamesischen Partner übergeben und wenn möglich auch in Vietnam installiert werden. Die RBIS Software selbst soll im Rahmen anderer Projekte weiterentwickelt werden.

Aktuelle größere Neu- und Weiterentwicklungen umfassen die Erweiterung der für den CSW bereitgestellten Datensätze von Zeitreihen- und Geodaten auf Bodendatensätze und die Entwicklung von Modulen zur Verwaltung von Simulationsmodellen. Damit verbunden ist auch die vollständige Beschreibung der Eingangs- und Ausgabedaten auf der Grundlage der „lineage“ -Komponente in dem ISO Standard 19115-1/2 [ISO, 2003 & 2009]. Darüber hinaus ist eine Anpassung der gesamten Tabellenstruktur und Codelisten an die überarbeitete ISO 19115-1 Version von 2014 geplant.

5 Literaturverzeichnis

[Badjana et al., 2015]

H. M. Badjana, F. Zander, S. Kralisch, J. Helmschrot, W.-A. Flügel: An information system for integrated land and water resources management in the Kara River basin (Togo and Benin), in International Journal of Database Management Systems (IJDMS) Vol.7, No.1, February 2015.

[ISO, 2003 & 2009]

ISO, INTERNATIONAL STANDARDS ORGANIZATION: International Standard ISO 19115 Geographic information – Metadata., Reference Number ISO19115:2003(E).

ISO, INTERNATIONAL STANDARDS ORGANIZATION: International Standard ISO 19115-2 Geographic information - Metadata - Part 2: Extensions for imagery and gridded data., Reference Number ISO19115:2009(E).

[Kralisch et al. 2012]

Kralisch, S., Böhm, B., Böhm, C., Busch, C., Fink, M., Fischer, C., Schwartze, C., Selsam, P., Zander, F. & Flügel, W.-A.: ILMS - a Software Platform for Integrated Water Resources Management. Proceedings of the International Congress on Environmental Modelling and Software, Sixth Biennial Meeting (R. Seppelt, A. A. Voinov, S. Lange & D. Bankamp, eds.). Leipzig, Germany, 2012.

[Kralisch et al., 2013]

S. Kralisch, F. Zander, and W.-A. Flügel: OBIS - a Data and Information Management System for the Okavango Basin, J. Oldeland, C. Erb, M. Finckh and N. Jürgens, Eds. Biodiversity and Ecology 5, 2013, pp. 213-220.

[Miga et al., 2013]

M. M. Julian, M. Fink, C. Fischer, P. Krause and W.-A. Flügel: Implementation of J2000 hydrological model in the western part of Java island, Indonesia, in The Journal of MacroTrends in Applied Science, 2013.

Die Implementierung von Umweltbewertungsverfahren in einer dienstebasierten Umgebung

Jörn Kohlus, Elisabeth Kompter, Landesbetrieb für Küstenschutz, Nationalpark und Meeresschutz in Schleswig-Holstein, joern.kohlus@lkn.landsh.de, elisabeth.kompter@lkn.landsh.de

Abstract

In the following article the results of the RichWPS project, which was promoted by the Federal Ministry of Education and Research (BMBF), are presented from the perspective of its application to perform mandatory tasks of the environmental reporting. The Marine Strategy Framework Directive (MSFD) explicitly claims the use of INSPIRE (INfrastructure for SPatial InfoRmation in Europe) for the provision and processing of data. INSPIRE is based on a network services architecture. A contribution of RichWPS in terms of the implementation of the MSFD is the transfer of assessment methods consistently in a service architecture and the development of tools for a simple and comfortable orchestration of Web Processing Services (WPS) to handle such complex processes.

1 Einleitung

Der Beitrag stellt Ergebnisse des BMBF-Projektes RichWPS aus Sicht der Anwendung für Aufgaben des Umweltberichtswesens vor. Die Meeresstrategierahmenrichtlinie (MSRL) der EU benennt explizit INSPIRE als Orientierung gebende Richtlinie zur Datenbereitstellung.

Für die MSRL sind nicht nur Daten bereit zu stellen, sondern auch Bewertungen vorzunehmen und Maßnahmen einzuleiten. INSPIRE setzt auf eine dienstebasierte Architektur. Bewertungsverfahren konsequent in eine Dienstearchitektur zu übertragen

und Werkszeuge für eine einfache Orchestrierung solcher komplexer Verfahren zu entwickeln, ist ein Beitrag von RichWPS für die Umsetzung der MSRL.

2 Aufgaben und Motivation

Bei der Zusammenarbeit in Forschung und Entwicklung ergibt sich meist eine kooperative Situation mit den beteiligten Hochschulen und Unternehmen. Dabei gerät oft ein wenig aus dem Blickwinkel, dass Behörden vor allem Dienstleister zur Umsetzung rechtlich oder politisch gesetzter Aufgaben sind. Mit der kontinuierlichen, eher zunehmenden Zahl an rechtlich bindenden Vorgaben, der zunehmenden Zentralisierung verkoppelt mit dem Zuwachs interner Vorgaben, dem resultierenden steigenden Arbeitsaufwand bei gleichzeitiger Mittelreduktion gehen Handlungsspielräume für zweckmäßige Innovationen zurück. Eine Veränderung, deren Auswirkungen auf den Arbeitskreis UIS erkannt und diskutiert wurden.

Für verbleibende Initiativen ist der Nachweis ihrer Verbindung und Erforderlichkeit für die Umsetzung der gesetzten Anforderungen nötig. Nach dem hohen Innovationsdruck durch die Umsetzungsanforderungen der Wasserrahmenrichtlinie ist es nun die Meeresstrategierahmenrichtlinie (2008/56/EG), die einen Teil der Bundesländer und einige Bundesministerien mit neuen, komplexen Aufgaben konfrontiert. Es geht einerseits darum, dass sich Länder- und Bundeseinrichtungen auf die Spezifizierung geeigneter qualitative Deskriptoren zur Festlegung eines guten Umweltzustandes einigen. Andererseits um die Festlegung von Kriterien und Verfahren zur Bestimmung und Bemessung des Meeresumweltzustandes; eine langwierige Aufgabe, die Expertise erfordert aber per sé nicht wissenschaftlich ist. Daran anschließend sind Maßnahmen festzulegen, mit denen ein guter Meereszustand erreicht werden soll. Und nicht zuletzt legt die Richtlinie fest, dass die Bearbeitung und Bereitstellung der Daten mit den Verfahren nach INSPIRE (INfrastructure for Spatial InfoRmation in Europe, 2007/2/EG) vorzunehmen ist. Bei der Umsetzung der MSRL ergibt sich damit auch die Notwendigkeit, sich den teilweise ungelösten Aufgaben zu INSPIRE mit dem richtungsweisenden Basiskonzept einer dienstebasierten Geodateninfrastruktur zu stellen.

Die von INSPIRE, den Geodatenzugangs- wie Umweltinformationsgesetzen beschriebenen Ziele sind anspruchsvoll und umfassen die Erwartung, dass Geobasis- und Geofachdaten weitestgehend interoperabel sind, alle verfügbaren Daten mit Metadaten beschrieben und recherchierbar sind, ein kostenfreier Zugang zu Umweltdaten für Öffentlichkeit und Wirtschaft gesichert wird, die Daten mit standardisierten Diensten für verschiedenste Nutzungen und miteinander nutzbar bereitgestellt werden und jederzeit verfügbar sind. Transparenz und Offenlegung der Daten sind Schlagworte.

Ein solcher Erwartungshorizont vermittelt vor allem die gesellschaftliche Rezeption einer modernen Datenhaltung. Die Bereitschaft, die hierfür erforderlichen Mittel anzubieten, ist deutlich geringer ausgebildet und zudem gibt es konzeptionelle wie technische Grenzen: Zusätzliches Personal zum Betrieb Tag und Nacht hochverfügbarer Rechner oder zur Eingabe von Metadaten soll in den Verwaltungen möglichst nicht eingestellt werden, Daten sind nicht sofort elektronisch verfügbar und benötigen eine Bearbeitung mit Expertise, nicht alle Daten sind mit allen Daten geeignet zu verbinden und nicht alle Standards so präzise, dass sie einen reibungsfreien Datenaustausch sicherstellen.

Ausgehend von zwei Masterarbeiten zur Umsetzung eines Bewertungsverfahrens für Makrophyten im Rahmen der Mitwirkung am Aufbau der Marinen Dateninfrastruktur Deutschland (MDI-DE) unternahm die Nationalparkverwaltung zusammen mit den anderen Projektpartnern den Versuch, die oben dargelegten Ziele konsequent am Beispiel der Makrophytenbewertung umzusetzen.

3 Umweltindikator Makrophyten und das Bewertungsverfahren

Die Seegrasarten *Zostera noltii* und *Zostera marina* reagieren empfindlich auf die Sedimentstabilität und Eutrophierung. Die Seegräser der Nordsee wurden durch einen Pilzbefall bis auf Restbestände im Nordfriesischen und Dänischen Wattenmeer nahezu ausgerottet. So ist das Seegras selbst aufgrund seiner Bestandsgefährdung zu einem eigenständigen Qualitätsparameter geworden. Laut der OSPAR-Empfehlung 2012/4

(Übereinkommen zum Schutz der Meeresumwelt des Nordostatlantiks, OSPAR) sollen die OSPAR-Mitgliedstaaten Programme ausarbeiten, um zu beurteilen, ob bestehende Maßnahmen zum Schutz der Seegraswiesen im Wattenmeergebiet wirksam und effektiv sind sowie den ökologischen Zustand der Seegraswiesen systematisch untersuchen (OSPAR Commission 2012).

Im Gegensatz zu den Seegräsern reagieren Grünalgen mit starkem Zuwachs auf erhöhte Nährstoffeinträge. Dabei können sich dichte Matten bilden, deren Absterben im Herbst zu Sauerstoffmangel insbesondere im Benthos führen kann. Makrophytische Algen gelten als Anzeiger für eine erhöhte Eutrophierung.

Seit über fünfzehn Jahren werden die Makrophyten der Wattflächen im Rahmen des Trilateralen Monitoring- und Bewertungsprogramm (engl. „Trilateral Monitoring and Assessment Program“, TMAP) (CWSS 2008) mit ergänzenden Parametern beobachtet. Aufgrund der Größe und der Dynamik der Vorkommen werden die Bestände dreimal im Jahr per Befliegung kartiert (u.a. Reise et al. 2010a, 2010b, 2011, 2012, 2013). Sicher erkannt werden Flächen mit einer Bedeckung von mehr als 20% im eulitoralen Watt, eine weitere Deckungsklasse von etwa über 60% kann identifiziert werden.

Abb. 1: Wasserkörper der Wasserrahmenrichtlinie im Schleswig-Holsteinischen Wattenmeer.

Das für die Implementierung genutzte Bewertungsverfahren für die Wasserrahmenrichtlinie greift auf ein von der Expertengruppe „Makrophyten und Zoobenthos für die Wasserrahmenrichtlinie“ der Arbeitsgemeinschaft Bund/Länder-Messprogramm (ARGE BLMP) im Jahre 2010 vorgeschlagenes Verfahren zurück, bei dem die Daten der Befliegung verwendet werden. Wichtiger Vorteil der Befliegung ist die Vollständigkeit der Erfassung, da die Entwicklung von Teilgebieten unterschiedlich verlaufen kann, wobei manchmal Zustandsveränderungen bis zu 10% im Jahr auftreten können. Zudem lässt sich über drei Befliegungen – die jahreszeitliche Variation liegt noch über den jährlichen Veränderungen – das Bestandsmaximum, das Grundlage eines konsistenten Vergleichswertes ist, erkennen und erfassen.

Bei der ökologischen Zustandsbewertung der Makrophyten gehen nach dem methodischen Konzept von Reise (vgl. Dolch et al. 2009) mehrere Faktoren ein:

- Ausdehnung sowie Bewuchsdichte der Seegraswiesen;
- Artendiversität der Wiesen (Vorhandensein beider heimischer Zostera-Arten);

- Ausdehnung und Dichte der Grünalgenmatten.

Zur Bewertung werden *Ecological Quality Ratios* (EQR, Birk & Böhmer 2007) entsprechend einer Bewertungsmatrix der jeweils für die kumulierten Bewertungsgebiete Nordfriesland und Dithmarschen (vgl. Tab. 1 und Tab. 2) bestimmt. Die Zusammenfassung erfolgt aufgrund der Mobilität der Grünalgenmatten zwischen den Wasserkörpern und der typischen Grenzlage der Seegrasvorkommen um die Wasserscheiden, so werden im Einklang mit der Expertengruppe „Makrophyten und Zoobenthos für die Wasserrahmenrichtlinie“ Gesamtwerte nicht für die einzelnen Wasserkörper (vgl. Abbildung 1), sondern für die gesamten ökologischen Teilsysteme Nordfriesisches und Dithmarscher Wattenmeer bestimmt (ARGE BLMP 2010).

Die Klassengrenzen beziehen sich auf einen Referenzzustand und können regional unterschiedlich gesetzt werden. Als Kriterien gehen der prozentuelle Anteil der Flächendeckung eulitoraler Seegraswiesen unterschieden in Bereiche mit Bedeckung von mehr als 20% bzw. 60% sowie die Artenabundanz innerhalb des jeweiligen Untersuchungsgebietes ein. Die Flächenanteile von Grünalgenmatten im Bezug zur Wattfläche der Makroalgen werden unterschieden nach zwei Dichtestufen und antiproportional in die Berechnung eingesetzt (Dolch et al. 2009). Die Gesamtbewertung des ökologischen Zustandes wird als arithmetisches Mittel aller EQR der Teilparameter über den Betrachtungszyklus von sechs Jahren berechnet.

	Bewertungsmatrix Nordfriesland Makrophytobenthos-Index							
	Qualitäts-kategorien	0	1	2	3	4	Gewichtung %	Norm-EQR gemäß Gewichtung für 6-Jahre-Interval
		Schlecht	Unbe-friedigend	Mäßig	Gut	Sehr gut		
	Norm-EQR	0 – 0,19	0,2 – 0,39	0,4 – 0,59	0,6 – 0,79	0,8 – 1,0		
Modul Seegras [6]	Eulitorale Fläche (%) [1]	< 2	2 - 4,9	5 - 9,9	10 - 19,9	20 - 100	50	Mittelwerte aller Parameter-EQRs über 6 Jahre
	Anteil ≥ 60 % Bedeckung (%) [2]	< 6	6 - 11,9	12 - 24,9	25 - 49,9	50 - 100	10	
	Präsenz beider Arten (%) [3]	< 20	20 - 39,9	40 - 59,9	60 - 79,9	80 - 100	10	
Modul Grünalgen [7]	Eulitorale Fläche (%) [4]	100 - 15	14,9 - 7	6,9 - 3	2,9 - 1	< 1	20	
	Anteil ≥ 60 % Bedeckung (%) [5]	100 - 50	49,9 - 25	24,9 - 12	11,9 - 6	< 6	10	

Tabelle 1: Bewertungsmatrix Nordfriesland Makrophytobenthos-Index (in Dolch et al. 2009)

Bewertungsmatrix Dithmarschen Makrophytobenthos-Index								
	Qualitäts-kategorien	0	1	2	3	4	Gewichtung %	Norm-EQR gemäß Gewichtung für 6-Jahre-Interval
		Schlecht	Unbe-friedigend	Mäßig	Gut	Sehr gut		
	Norm-EQR	0 – 0,19	0,2 – 0,39	0,4 – 0,59	0,6 – 0,79	0,8 – 1,0		
Modul Seegras [6]	Eulitorale Fläche (%) [1]	< 0,3	0,3 - 0,69	0,7 - 1,49	1,5 - 2,9	3 - 100	50	Mittelwerte aller Parameter-EQRs über 6 Jahre
	Anteil ≥ 60 % Bedeckung (%) [2]	< 6	6 - 11,9	12 - 24,9	25 - 49,9	50 - 100	10	
	Präsenz beider Arten (%) [3]	< 20	20 - 39,9	40 - 59,9	60 - 79,9	80 - 100	10	
Modul Grünalgen [7]	Eulitorale Fläche (%) [4]	100 - 15	14,9 - 7	6,9 - 3	2,9 - 1	< 1	20	
	Anteil ≥ 60 % Bedeckung (%) [5]	100 - 50	49,9 - 25	24,9 - 12	11,9 - 6	< 6	10	

Tabelle 2: Bewertungsmatrix Dithmarschen Makrophytobenthos-Index (in Dolch et al. 2009)

Um den Anforderungen der WRRL zu entsprechen, werden die Bewertungen der Teilsysteme den beteiligten Wasserkörpern zugeordnet.

4 Implementierung des Verfahrens

Eine erste Implementierung des Verfahrens erfolgte durch Rieger (Rieger 2011). Die eingehenden Geo- und Fachdaten wurden per Datenbank bereitgestellt, wobei die Geooperationen und statistischen Berechnungen mittels einer PostgreSQL Datenbank und dem räumlichen Aufsatz PostGIS prozessiert wurden. Die Kommunikation mit dem Anwender erfolgt über eine PHP-Schnittstelle auf dem Apache-WEB-Server (vgl. Abbildung 2). Um die Bewertungskarten, -texte und -diagramme dynamisch zu erzeugen und das Ergebnis unabhängig von der implementierten Umgebung zugänglich zu machen, wurden die Ergebnisdaten als Web-Dienste WMS und WFS zugänglich gemacht (Rieger et al. 2013, S. 180).

Ein Bewertungsvorgang wird durch die Angabe des Berichtsjahres in einer für die Bewertung implementierten interaktiven Weboberfläche angestoßen. Die Ergebnisse stehen in verschiedener Form und Funktionalität zur Verfügung: Zwecks Visualisierung der Berechnung in der Karte wird ein WMS-Dienst vom GeoServer angesprochen, die

Bewertungsgebiete mit den aktuell kalkulierten Werten können außerdem per WFS als georeferenzierte Datei im Shape-Format direkt aus dem Formular heruntergeladen werden. Mit der Funktion „Bericht erstellen“ kann der Anwender einen digitalen Bericht erzeugen und diesen im PDF-Format abspeichern („Bericht drucken“). Des Weiteren wird dem Anwender durch die Funktion „Zwischenergebnisse“ die Möglichkeit angeboten, sich die vollständige Berechnung in Text- und Tabellenform anzeigen zu lassen, wobei alle Bewertungsschritte transparent und didaktisch erläutert werden.

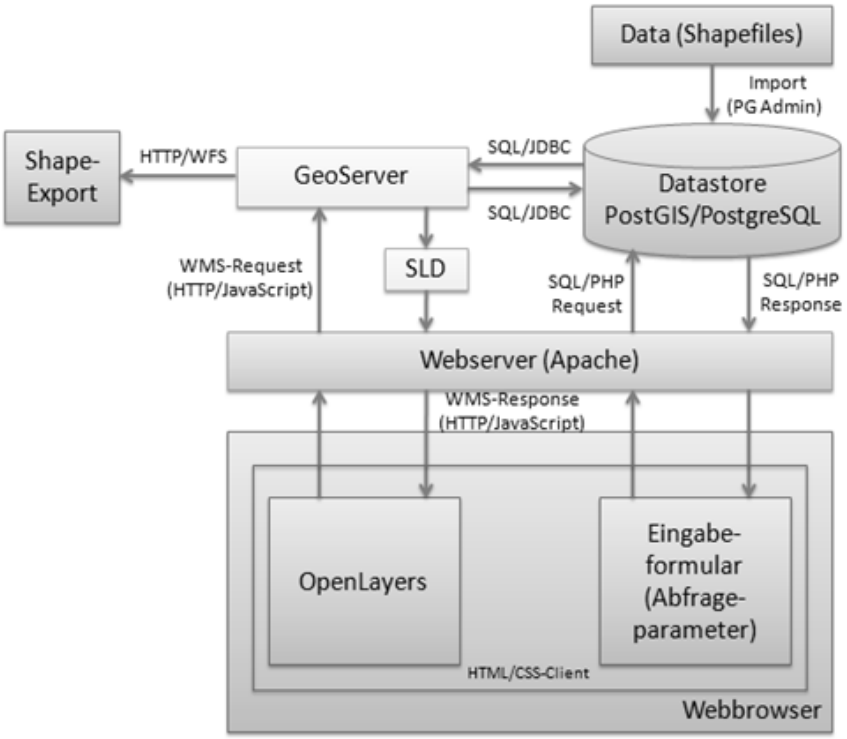


Abbildung 2: Datenbankbasierte Umsetzung mit PostGIS/PostgreSQL

Ein Bewertungsdiagramm zeigt die Entwicklung über die sechs einzelnen Jahre. Diese Grafik wird durch einen kurzen Bewertungstext, der dynamisch aus Textbausteinen und der Analyse der Trends generiert wird, ergänzt. Der generierte exemplarische Textbericht lässt sich vor dem Drucken von den beteiligten Experten editieren und ergänzen, um Sonderkonditionen beim Bewertungsvorgang festzulegen und zu erläutern.

5 Überführung in eine dienstebasierte Implementierung

Die Nachimplementierung von Rieger (Rieger 2012) stützte sich dann auf Quellen per Dienste und produzierte Ergebnisse, die auch in Form von Diensten bereitgestellt werden konnten. Das Verfahren erlaubt es damit, Daten per Web-Services auf einem über viele Server verteilten System – wie der MDI-DE – gemeinsam zu nutzen. Die Eingangsdaten werden bereits als Web-Services (Web Map Service (WMS) und Web Feature Service (WFS)) von den lokalen Infrastrukturknoten für die MDI-DE in Schleswig-Holstein (Helbing et al. 2012, Lübker et al. 2013) bezogen, die dort mittels der Software GeoServer aufbereitet wurden. Zur interaktiven Visualisierung der Geodaten wurde die JavaScript-Bibliothek OpenLayers als Map-Viewer eingesetzt, die u.a. die Einbindung von OGC (engl. „Open Geospatial Consortium“) konformen WMS und WFS ermöglichte. Im Gegensatz zu GIS-basierten Ansätzen werden die Ergebnisse für den Bewertungsservice vollständig automatisch und dynamisch erstellt und passen sich je nach Ausgangsdaten ohne Eingriff des Nutzers von selbst an.

Konform zu der von INSPIRE geforderten dienstebasierten Technologie sollte jetzt eine Implementierung mittels eines Web Processing Service (WPS) erfolgen. Die Umsetzung wurde mithilfe des WPS-Frontends Legato der Firma disy Informationssysteme GmbH und einer Testumgebung für WPS-Services erprobt. Es gelang, ein vereinfachtes Bewertungsverfahren mit interaktiver Auswahlmöglichkeit des Analysegebietes einzurichten (Rieger 2012, Wössner 2013). Der hierbei erstellte WPS umfasste allerdings die gesamte Verarbeitungskette mehr oder weniger monolithisch programmiert und kann nicht zweckdienlich für andere Aufgaben verwendet werden.

Im Rahmen des BMBF Projektes RichWPS wurden daher von den Partnern LKN, der Fa. disy, Lehrstuhl für Geoinformatik an der Hochschule Osnabrück und Bundesanstalt für Wasserbau die anschließenden Fragestellungen aufgegriffen, die sich auch aus der Perspektive von Teilhabern der Marinen Dateninfrastruktur Deutschland (MDI-DE) ergaben:

- Neben der Verpflichtung Umweltdaten konform zu INSPIRE bereitzustellen, besteht der Wunsch, ihre zweckmäßige Verwendung als Referenz per WPS anzubieten

- Ein Haupteinsatzbereich für WPS wird in der Harmonisierung von gleichartigen Daten verschiedener Infrastrukturknoten gesehen und
- ein weiterer in der Bereitstellung transparenter WPS für Bewertungsverfahren nach der MSRL und WRRL (Kohlus, Roosmann 2014)
- um durch das Personal bei der MDI-DE beteiligter Institutionen mit realistischem Aufwand bereit gestellt werden zu können, müssen programmierte Teile von WPS wiederverwendbar und ohne Programmierung anpassbar sein
- lange Diensteketten einfacher Prozessierungsschritte sind aufgrund von Zeitkriterien und Datenvolumen nicht realisierbar (u.a. time out), Teilprozesse müssen zu einem WPS zusammenstellbar sein.

Ausgehend von diesen Fragestellungen wurde als Zielstellung von RichWPS die Weiterentwicklung von konzeptionellen und technischen Grundlagen von Web Processing Services (Wosniok et al. 2014) identifiziert. Hierfür erfolgte eine Erprobung und Demonstration in exemplarischen, praxisnahen Prototypen für Modelle und Bewertungsverfahren, wie die Makrophytenbewertung. Im Rahmen von RichWPS wurde eine Software-Umgebung entwickelt, die es ermöglicht, komplexe Workflows benutzerfreundlich aus existierenden Diensten einer Geodateninfrastruktur (GDI) graphisch zu modellieren.

Im Zuge von RichWPS wurde der komplexe WPS-Prozess der Makrophytenbewertung in einzelne Prozessbausteine zerlegt. Dabei wurde versucht, wiederverwendbare Bausteine als Teilprozesse auszugliedern. Die Makrophytenbewertung wird anhand von fünf einzelnen Prozesse, die miteinander verknüpft sind, durchgeführt (siehe Abbildung 4). Der Prozess *MSRLD5Selection* selektiert unter Angabe des Bewertungsjahres die Makrophytenvorkommen in Seegras- und Grünalgen-Daten. Weiterhin wird eine Liste mit den für den Bewertungszeitraum relevanten Jahren ausgegeben, d.h. wird als Bewertungsjahr 2010 angegeben, dann entspricht der Bewertungszeitraum 2005 bis 2010. Die Liste der relevanten Jahre wird an den Prozess *SelectTopography* weitergegeben, um die zeitlich nächste Topographie zu ermitteln. In den Prozess *SelectReportingArea* werden die Wasserkörper der Gebiete Nordfriesisches und

Dithmarsches Wattenmeer selektiert, die wiederrum mit den Wattflächen der selektierten Topographien verschnitten werden (*Intersect*). Die bereits genannten Prozesse dienen der Vorbereitung für die Bewertung der Makrophytenvorkommen, die im Prozess *Characteristics* erfolgt. Eingangsdaten sind u.a. die Seegras- und Grünalgenvorkommen sowie die Verschneidungsprodukte von Wattflächen und Wasserkörpern. Die bewerteten Flächen für NF und DI werden in Form einer GML und XML an den Report-Prozess übergeben, der schließlich die Ergebnisse nutzt und einen PDF-Bericht generiert.

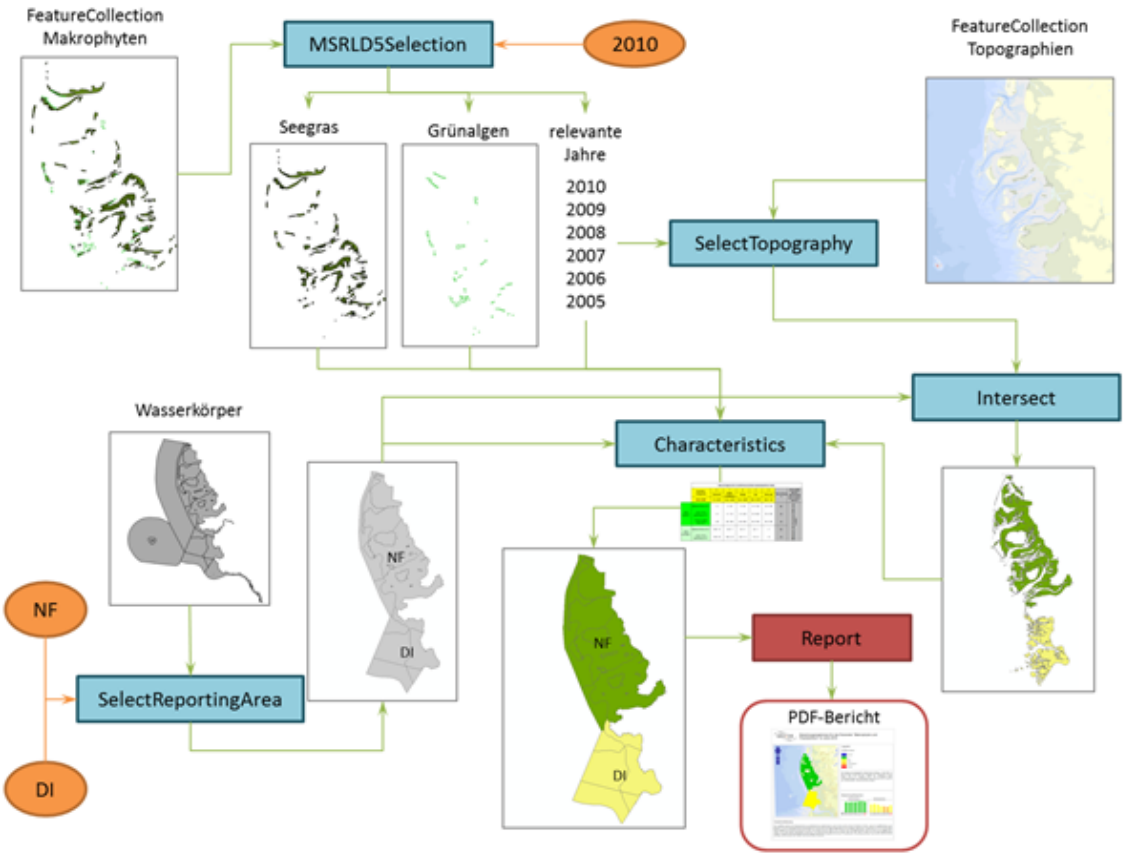


Abbildung 4: Die Makrophyten-Prozessbausteine (blau, rot).

Die einzelnen Prozessbausteine können zu einem komplexen Bewertungsprozess zusammengestellt und integriert werden. Hierfür wurde im Projekt der ModelBuilder entwickelt. Dieser stellt bei der Umsetzung von Bewertungsverfahren mittels Dienstetechnologie eine benutzerfreundliche Modellierungsoberfläche dar. Die Inbetriebnahme des ModelBuilders (MB) erfolgt über eine Java-Entwicklungsumgebung. Das

System erlaubt die Angabe eines http-Proxyservers und kann daher auch in der gesicherten Umgebung des LKN eingesetzt werden.

Die einzelnen Prozesse der Makrophytenbewertung sind auf einem Server der Hochschule Osnabrück bereitgestellt. Mittels des ModelBuilders können diese ausgewählt und einfach kombiniert werden (Bensmann et al. 2014, Alcacer-Labrador et al. 2014, Ziegenhagen et al. 2014). Das modellierte Makrophytenbewertung ist in Abbildung 5 dargestellt. Bevor das Modell ausgeführt werden kann, muss es zunächst auf dem Server als WPS bereitgestellt werden. Mit Hilfe eines Clients können anschließend die Eingangsdaten definiert und das Modell ausgeführt werden. Als Input-Daten stehen WFS der Seegras- und Grünalgenvorkommen, der Topographie und der Wasserkörper auf dem Schleswig-Holsteinischen Infrastrukturknoten der MDI-SH bereit. Das Ergebnis wird in Form eines PDF-Berichtes ausgegeben, der temporär auf dem Server gespeichert wird.

Der ModelBuilder unterstützt die Arbeit des Nutzers mit mehreren Funktionen: Mittels der Layout-Funktionen wird die Anpassung der graphische Darstellung von Modellen ermöglicht. Zwei weitere Funktionen (*Testing* und *Profiling*) helfen bei der Validierung des Modells und der Bestimmung der Laufzeit einzelner Prozesse. Mittels des *Testings* ist es möglich, auch Zwischenergebnisse in Form von GML-Dateien ausgeben zu lassen. Die Zwischenergebnisse helfen besonders bei der Validierung des Modells und des Endergebnisses. Weiterhin ist es somit möglich, vorhandene Fehlerquellen im Prozessablauf zu identifizieren. Besonders hilfreich ist das *Profiling*, wenn die verwendeten Prozesse auf unterschiedlichen Servern liegen. Damit wird es möglich den schnelleren Server zu identifizieren, um diesen schließlich für die Durchführung des Modells zu verwenden.

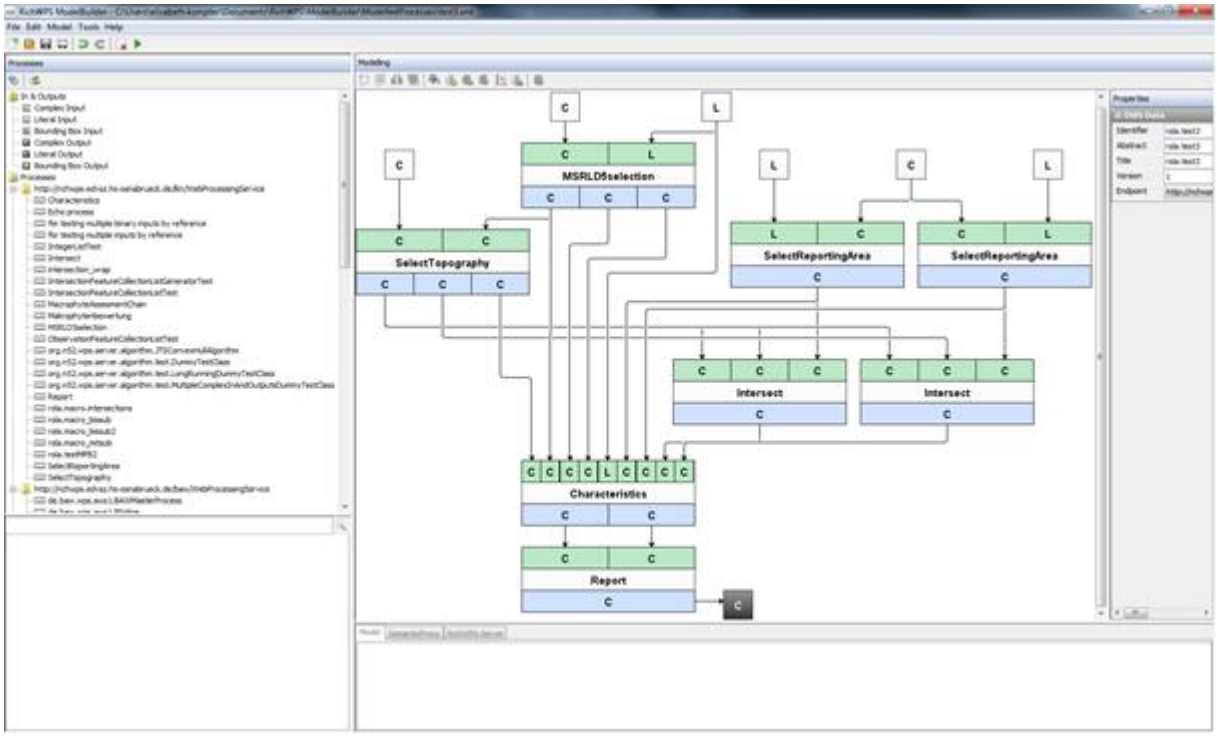


Abbildung 5: Benutzerschnittstelle des RichWPS ModelBuilders mit den Prozessen für die Makrophytenbewertung.

6 Ausblick

Mit dem Vorhaben RichWPS konnten mehrere Probleme bei der Anwendung von WPS in der Verwaltung reduziert werden. Komplexe Verarbeitungsketten lassen sich mit dem WPS-ModelBuilder zusammenstellen.

Die Evaluierung der bisher bekannten Bewertungsprozesse erbrachte, dass viele Verarbeitungsschritte durch wiederverwendbare Prozesse übernommen werden können, Teile der Prozessierung aber nur spezifisch für einen Bewertungsprozess benötigt werden. Der Aufwand, durch Programmierung komplexe WPS zu erzeugen, lässt sich mittels des ModelBuilders deutlich reduzieren, allerdings verbleibt der Bedarf zur Programmierung von Teilprozessen.

Der Einsatz der erzeugten komplexen WPS zur Harmonisierung von Daten verschiedener Diensteanbieter ist möglich, ebenso sind die erzeugten WPS geeignet, als Referenzverarbeitung neben den Datendiensten angeboten zu werden.

Am Ende des Projektes RichWPS verbleiben aus unserer Sicht zwei Herausforderungen: Die Implementierung eines erweiterten Systems auf verteilten Systemen mit alternativ nutzbaren Diensten auch unter Last sowie die Nutzung von WPS mit WFS-Diensten verschiedener Serveranbieter, denn noch zeigt sich, dass die Dienste bereits auf technischer Ebene nicht kompatibel miteinander sind.

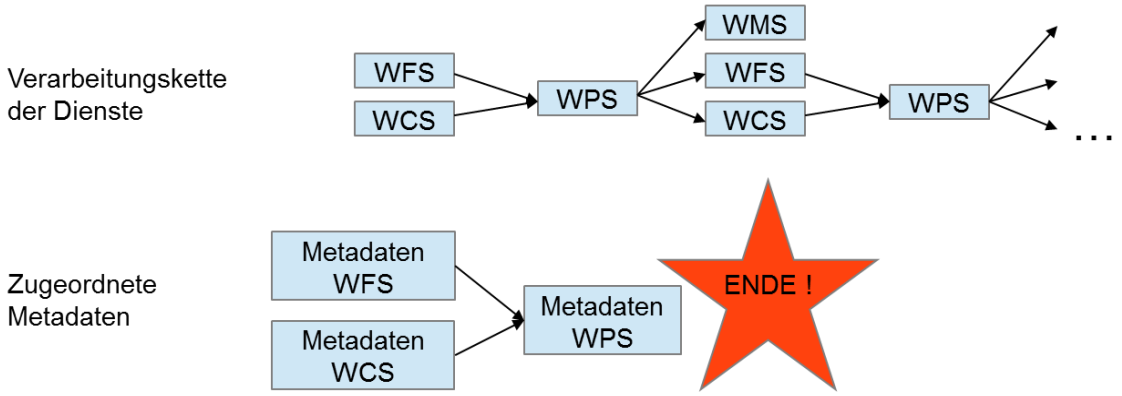


Abbildung 6: Fehlendes Konzept zur Metadatenbereitstellung für WPS-Datenprodukte

Zur Bereitstellung von Geodaten und Geodiensten gehört, dass hierzu entsprechende ISO-konforme Metadaten angeboten werden. Ein grundsätzlicher Mangel besteht unserer Ansicht nach darin, dass für Ergebnisdaten eines WPS keine Metadaten bereitstehen (Abbildung 6). Anders als bei der Nutzung von Diensten fehlt es hier an der Möglichkeit eine Verarbeitungskette zu bilden. Wir sehen darin auch einen entscheidenden Faktor bei der Akzeptanz von WPS für die Datenaufbereitung und Harmonisierung und sehen die Anforderung nach einer parallelen Prozessierung von Metadaten.

7 Literaturverzeichnis

[Alcacer-Labrador et al. 2014]

Alcacer-Labrador, D., Bensmann, F., Roosmann, R. (2014): Web Processing Service (WPS) Orchestration – A Practical Approach“, in FOSS4G-E Academic Track Bremen, Jul. 14.

[ARGE BLMP 2010]

ARGE BLMP - Arbeitsgemeinschaft Bund/Länder-Messprogramm für Nord- und Ostsee, Ministerium für Landwirtschaft, Umwelt und ländliche Räume des Landes Schleswig-Holstein (Hrsg.) (2010): Makrophyten. Monitoring-Kennblatt, Biologisches Monitoring – Flora – Makrophyten, Stand 08.06.2015. [http://www.blmp-online.de/Seiten/Monitoringhandbuch.htm].

[Bensmann et al. 2014]

Bensmann, F., Roosmann, R., Kohlus, J. (2014): „Aus Klein mach Groß: Komposition von OGC Web Services mit dem RichWPS ModelBuilder“, in 21. Workshop des Arbeitskreis Umweltinformationssysteme Hochschule Karlsruhe - Technik und Wirtschaft - (HsKA), Karlsruhe, May 14.

[Birk & Böhmer 2007]

Birk, S., Böhmer, J. (2007): Die Interkalibrierung nach EG-Wasserrahmenrichtlinie – Grundlagen und Verfahren. , S. 10-14, WaWi 9/2007.

[CWSS 2008]

CWSS – Common Wadden Sea Secretariat (Hrsg.)(2008): TMAP Handbook -TMAP guidelines for an integrated Wadden Sea monitoring. Version 0.9, May 2008. Aktueller Stand 08.06.2015 http://www.waddensea-secretariat.org/monitoring-tmap/about-tmap

[Dolch et al. 2009]

Dolch T., Buschbaum C., Reise K. (2009): Seegras-Monitoring im Schleswig-Holsteinischen Wattenmeer 2008. Ein Forschungsbericht für das Landesamt für Natur und Umwelt des Landes Schleswig-Holstein, Flintbek, 2009.

[Helbing et al. 2012]

Helbing, F., Binder, K., Duden, S., Lübker, T., Räder, M., Schacht, C., Zühr, D. (2012): Leitfaden zur Anbindung eine Infrastrukturknotens an die MDI-DE. [http://projekt.mdi-de.org/images/mdi-de/Publikationen/plugin-mdi-de_leitfaden_isk_2_0_publish.pdf].

[Kohlus, Roosmann 2014]

Kohlus, J., Roosmann, R. (2014): "Online-Bewertungsverfahren mit INSPIRE-Techniken? Umsetzungsansätze am Beispiel Seegras" in Arbeitskreis Umweltinformationssysteme Schleswig-Holstein, Jan. 14.

[Lübker et al. 2013]

Lübker, T., Helbing, F., Kohlus, J. (2013): Infrastrukturknoten – Technische Bausteine der MDI-DE. In: Traub, K.-P., Kohlus, J., Lüllwitz, T. (Hrsg.): Geoinformationen für die Küstenzone. S. 63-71, Band 4, Koblenz.

[OSPAR Commission 2012]

OSPAR Commission (2012): OSPAR Recommendation 2012/4 on furthering the protection and conservation of Zostera beds. Meeting of the OSPAR Commission, 25.-29. Juni 2012, Bonn. Annex 13, Ref. §5.38c.

[Reise et al. 2010a]

Reise, K., Buschbaum C., Dolch, T., Herre, E. (2010a): Vorkommen von Grünalgen und Seegras im Nationalpark Schleswig-Holsteinischen Wattenmeer 2009. Unveröfft. Forschungsbericht für den Landesbetrieb Küstenschutz, Meeresschutz und Nationalpark des Landes Schleswig-Holstein, List.

[Reise et al. 2010b]

Reise, K., Buschbaum, C., Herre, E. (2010b): Vorkommen von Grünalgen und Seegras im Nationalpark Schleswig-Holsteinisches Wattenmeer 2010. Unveröfft. Forschungsbericht für den Landesbetrieb Küstenschutz, Meeresschutz und Nationalpark des Landes Schleswig-Holstein, List.

[Reise et al. 2012]

Reise, K., Buschbaum, C., Dolch, T., Herre, E. (2012): Vorkommen von Grünalgen und Seegras im Nationalpark Schleswig-Holsteinisches Wattenmeer 2011. Unveröfft. Forschungsbericht für den Landesbetrieb Küstenschutz, Meeresschutz und Nationalpark des Landes Schleswig-Holstein, List.

[Reise et al. 2013]

Reise, K., Buschbaum, C., Dolch, T., Herre, E. (2013): Vorkommen von Grünalgen und Seegras im Nationalpark Schleswig-Holsteinisches Wattenmeer 2012. Unveröfft. Forschungsbericht für den Landesbetrieb Küstenschutz, Meeresschutz und Nationalpark des Landes Schleswig-Holstein, List.

[Rieger 2011]

Rieger, A. (2011): Erstellung eines Web-basierten, semiautomatischen Bewertungsverfahrens für die Wasserrahmenrichtlinie (WRRL) als Komponente der marinen Geodateninfrastruktur Deutschland (MDI-DE) am Beispiel Makrophyten. Masterthesis, HafenCity Universität Hamburg, Studiengang Geomatik.

[Rieger 2012]

Rieger, A. (2012): Einrichtung und Entwicklung eines Web Processing Services (WPS) für WRRL/MSRL-Bewertungsverfahren auf Basis des Legato-Dienstservers. Interner Bericht für das LKN, Tönning. Hamburg.

[Rieger et al. 2013]

Rieger, A., Kohlus, J.. Traub, K-P. (2013): Automatisiertes webbasiertes Verfahren zur ökologischen Bewertung von Makrophyten im Schleswig-Holsteinischen Wattenmer. In: Traub, K.-P., Kohlus, J., Lüllwitz, T. (Hrsg.): Geoinformationen für die Küstenzone. S. 171-184, Band 4, Koblenz.

[Wössner 2013]

Wössner, R. (2013): Untersuchungen zur praktischen Nutzbarkeit des OGC Web Processsing Service (WPS) Standards. Unveröffentlichte Masterthesis, Hochschule Karlsruhe Technik und Wirtschaft.

[Wosniok et al. 2014]

Wosniok, C., Bensmann, F., Wössner, R., Kohlus, J., Roosmann, R., Heidmann, C., Lehfeldt, R. (2014): "Enriching the Web Processing Service", in European Geosciences Union General Assembly 2014, Wien, April/May 14.

[Ziegenhagen et al. 2014]

Ziegenhagen, D., Kohlus J., Roosmann, R. (2014): "Orchestration of Geospatial Processes with RichWPS – A Practical Demonstration" in EnviroInfo 2014: Proceedings of the 28th International Conference on Informatics for Environmental Protection, pp. 525-532, Oldenburg, Sep. 14, BSI-Verlag, Carl von Ossietzky University Oldenburg, ISBN 978-3-8142-2317-9.

Quo vadis INSPIRE?

Dr. Heino Rudolf, M.O.S.S. Computer Grafik Systeme GmbH,
Buchenstr. 16b , 01097 Dresden, E-Mail: hrudolf@moss.de

Abstract

An entirely new approach to environmental data management – sustainable, scalable expandable and interoperable – including the data provision for INSPIRE

Environmental data are usually subject-specific structured and incompatible for multidisciplinary applications. INSPIRE itself has thematically tailored data models. We miss a Theme Crossed approach for environmental data management. This presentation shows how a harmonized data management can be constructed and how the environmental data can be made available according to the INSPIRE definitions of Annex III.

The contents and the amount of the data for the thematic tasks are substantially larger than the data specification of INSPIRE requires. That's why we added some attributes for INSPIRE in our central data base and we created one data pool for all reporting, INSPIRE included.

Johann Wolfgang von Goethe
Aus „Prometheus“

…
Hier sitz' ich, forme Menschen
Nach meinem Bilde,
Ein Geschlecht, das mir gleich sei,
Zu leiden, weinen,
Genießen und zu freuen sich
…

Zusammenfassung

Ein völlig neuer, innovativer Ansatz für das Umweltdatenmanagement – nachhaltig, skalier-/erweiterbar und interoperabel

Ich muss kein großer Prophet sein, um festzustellen, dass uns die Interoperabilität von Daten völlig neue Horizonte eröffnet und eröffnen wird. INSPIRE richtet sich an die Behörden und stellt Regelungen auf, um sie in diese Entwicklungen zu integrieren: Die Daten der Behörden zur Beschreibung unserer Umwelt sind interoperabel bereitzustellen. – Selbstverständlich können dadurch weitere Wertschöpfungspotenziale erschlossen werden; und für die Behörden ist es die große Chance, diesen Prozess mitzugestalten. Beim Annex I von INSPIRE funktioniert das im Allgemeinen auch, denn hier handelt es sich um die Geobasisdaten, für die die Vermessungsverwaltungen zuständig sind; und die Modelle passen auch auf ihre Arbeitsweise.

Der eigentliche Clou kommt natürlich mit den vielen Fachdaten aus Annex II und III. Die Idee ist konsequent und atemberaubend! – Nur scheint das so in der Praxis nicht zu funktionieren: wie viele Behörden versuchen herauszufinden, ob sie sich beteiligen müssen oder wie sie das alles so minimal wie möglich umsetzen können. – Warum?

In vielen Veröffentlichungen, Präsentationen und Beratungen stelle ich immer wieder den Gedanken der „Entmystifizierung der Modelle" in den Mittelpunkt: Zumeist sind nur ganz wenige Fachdaten zu übermitteln, was bei vielen Anwendern einfach abfallen könnte. Nur bei den komplizierten UML-Diagrammen und Basisstrukturen (ich nenne sie oft „Unterbauten") geht diese Pragmatik völlig verloren. (aus [1], Epilog)

Auf drei Ebenen habe ich versucht, das Umweltdatenmanagement grundlegend neu zu gestalten:

- einen neuen fachlichen Modellansatz, der systemanalytische Betrachtungen zum Ökosystem nutzt
- eine veränderte Art und Weise der Datenmodellierung, die zweistufig vorgeht: Systemanalysediagramm (als Abbild der Realität) und Anwendungsdiagramme (für die verschiedenen Anwendersichten auf die Realität)

- eine konsequent realitätsorientierte Objektdefinition, die die Objektklassen nicht nach geometrischen Aspekten bildet – sondern die Gegenstände, Handlungen und Prozesse der Realität modelliert und alle Elemente vierdimensional in Raum und Zeit betrachtet.

Ich bin davon überzeugt: Es ist ein Glück, dass es INSPIRE gibt! – Lasst es uns auch für alle Beteiligten und darüber hinaus erlebbar machen. Ich würde mich sehr freuen, wenn dazu meine Lösungsansätze beitragen können.“ (aus [1], Epilog)

1 Die Ausgangssituation

Umweltdaten liegen im Allgemeinen fachkonkret strukturiert vor und sind für themenübergreifende Anwendungen oftmals nicht kompatibel. Auch in den INSPIRE-Datenspezifikationen finden wir thematisch abgegrenzte Datenmodelle zu den einzelnen Anhängen. Sie nutzen zwar standardisierte Grundstrukturen, vor allem der ISO 191xx-Serie bzw. aus INSPIRE-GCM, ein fachübergreifendes Verständnis der Zusammenhänge, Vorgänge und Prozesse in unserer Umwelt finden wir nicht.

In der Praxis begegnen wir immer wieder bei Ist-Analysen einer gewaltigen Anzahl an Shape-Files und Fachanwendungen – mit zumeist eigenen, proprietären Datenstrukturen, keine Harmonisierungen der Daten, statt dessen Inkonsistenzen und Mehrfacherfassungen.

Diese Situation ist den meisten Verantwortlichen in den Umweltbehörden bewusst. Auch deshalb hegen viele große Erwartungen an die INSPIRE-Datenmodelle – mit der Hoffnung, jetzt Interoperabilität und Kompatibilität in den Daten herstellen zu können. Umso größer ist dann oft die Ernüchterung.

2 INSPIRE und Umweltbehörden

Mit einer kritischen Analyse möchte ich versuchen, objektive Gründe zu benennen, warum INSPIRE Annex III von den Behörden häufig nur schwerfällig umgesetzt wird:

- INSPIRE hat keinerlei Bedeutung für die eigentliche Arbeit der Fachbehörden:

- Nicht einmal die Berichterstattungen können mit INSPIRE erfüllt werden.
- Der Wunsch für ein themenübergreifendes Umweltdatenmanagement wird durch INSPIRE nicht befriedigt.
- Die Art und Weise der Modellierung steht im Widerspruch zur Arbeitsweise der Fachbehörden, die „Dynamik in den Daten“ verlangt.
- Die Modelle sind so kompliziert notiert, dass kaum einer die Zeit und Mühe investieren möchte, sie zu lesen – geschweige denn, sie zu verstehen.

INSPIRE ist etwas Zusätzliches – eine Pflicht, die keine Vorteile in der Arbeit bringt!

(Vergleiche mit [1], Epilog.)

3 Anforderungen an das Umweltdatenmanagement

Einerseits muss es möglich sein, mit diesen gewaltigen Datenmengen und –strukturen umgehen zu können. Die Daten haben einen großen Wert, und ihre Weiterverarbeitung über den eigentlichen Erfassungsrahmen hinaus schafft neue Werte und ist effektiv.

Andererseits müssen wir der „Dynamik in den Daten“ gerecht werden (vergleiche mit [1], Abschnitt 3):

- Die Vielfalt der Aufgabenstellungen (Vollzug, Überwachung, Planung …), verschiedene Bewertungen ein und derselben Datenerhebung und damit abweichende Sichten auf die Objekte, was zu grundsätzlich verschiedenen Herangehensweisen bei der Datenmodellierung und -verarbeitung führen kann.
- Sich ständig verändernde Anforderungen an das Datenmanagement und die Datenbereitstellung, neue Berechnungs- und Simulationsprogramme, Geodesign, neue Gesetze und Berichterstattungen …
- Die Notwendigkeit der Erhebung und Verwaltung von Umweltzuständen zu konkreten Zeitpunkten (insbesondere für Berichterstattungen zu EU-Richtlinien, die im Allgemeinen periodisch zu wiederholen sind)

- Erstellung und Verwaltung von verbindlichen Planungen und Bearbeitung von Planungsszenarien
- Ständig neue fachliche Inhalte, Betrachtungshorizonte, Bedürfnisse, spezielle (auch individuelle und politisch geprägte) Sichten.

Bonk: 04/2015

4 Die Idee

War in der Vergangenheit die Datenverarbeitung auf die Vereinfachung und Automatisierung der Prozesse fokussiert, so finden wir heute neue kreative Ansätze, im Rechner unsere Umwelt möglichst authentisch zu verwalten; d. h., sie einerseits realitätsnah abzubilden und andererseits die Daten um Bewertungen, Planungen, Simulationen u. ä. anzureichern. Dafür müssen wir riesige Datenmengen, komplexe Datenstrukturen mit Ursache-Wirk-Beziehungen managen können. Nutzen wir diese Datenbestände, so können wir heute intelligente Apps (z. B. Routenabfragen, Wetterinformationen, mobile Anwendungen…) entwickeln und anbieten.

Diesen Ansätzen folgend, sollte es uns doch gelingen, Gegenstände, Vorgänge und Prozesse unserer Realität/Umwelt mit einer einheitlichen Modellierungsmethodik nachzubilden. Die Datenverwaltung spiegelt ein Abbild der Realität wieder. Und dann erwecken wir dieses Abbild mit Leben: nutzen die Daten für verschiedene Ansichten

und Visualisierungen, für Simulationen und Planungen, für komplexe Anfragen und Berichte.

5 Der Ökosystemansatz

In der heutigen GIS-Welt ist die Geometrie das objektbildende Element. Von diesem Paradigma trenne ich mich. Ich lasse mich bei der Objektdefinition nicht von Karten oder anderen von Menschen geschaffenen Darstellungen der Realität leiten. Eine Objektklasse definiere ich als Abbild einer Menge realer Objekte. Und als reale Objekte verstehe ich nicht nur die Betrachtungsobjekte; werden z. B. Prozesse beschrieben (wie menschliche Handlungen, Aktivitäten oder Vorgänge in unserer Umwelt), dann behandle ich diese als eigenständige Objektklassen. Immer, wenn Daten und/oder Funktionen zu verarbeiten sind, dann wird eine Objektklasse kreiert. Und alle (!) Objektklassen existieren in Raum und Zeit.

Um die obige Idee (Abschnitt 4) umzusetzen, studierte ich die ökosystemaren Modellansätze und sortierte die Systemelemente so, dass damit eine Struktur für das Datenmanagement ableitbar wurde (vergleiche [1], Abschnitt 6 und Caput 4). Damit entstand ein neuer Ansatz für ein themenübergreifendes Umweltdatenmanagement.

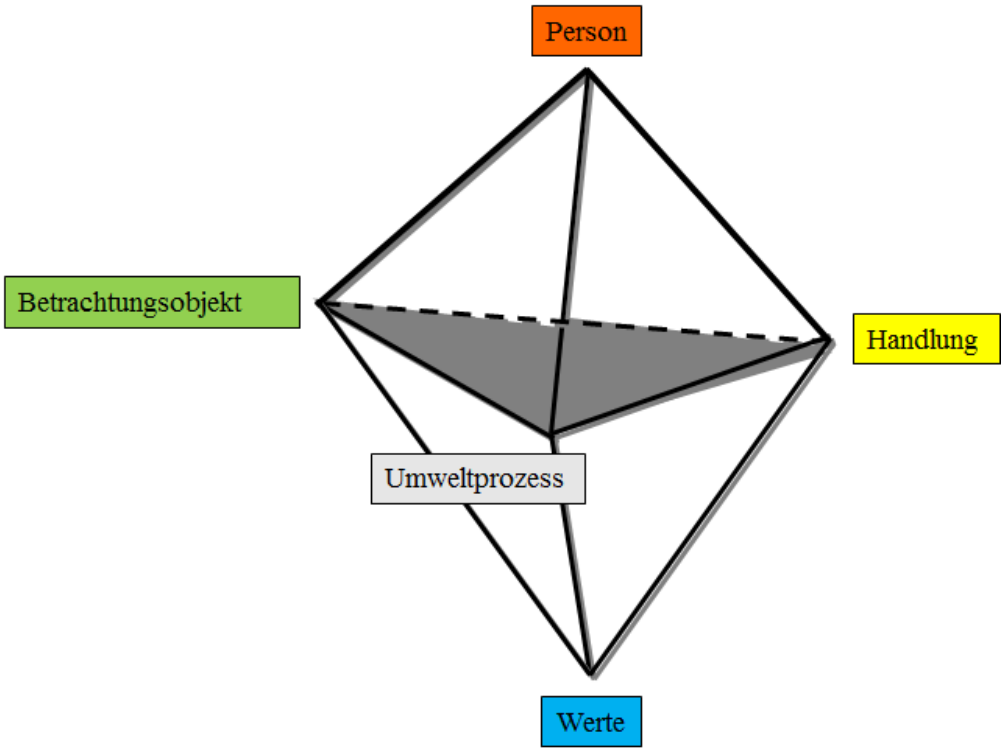


Abbildung 1: Der Doppeltetraeder

Zur Veranschaulichung des Modellansatzes wählte ich den Doppeltetraeder mit den Eckpunkten: „Betrachtungsobjekt“ (z. B. Anlagen, Gebäude, Gewässer, Messstellen, Schutzgebiete), „Umweltprozess“ (Speichern und Verarbeiten von Stoffen und Energien, Emissionen u. ä.), „Handlung“ (Beobachtungen, Messungen, Steuerungen und Regulierungen des Ökosystems), „Werte“ (physikalische, chemische, biologische oder andere Eigenschaft der Betrachtungsobjekte oder Umweltprozesse, die durch Handlungen erhoben werden können) und „Person“ (ein zentrales Personen- und Adressverzeichnis).

6 Der Modellierungsansatz

Bei allen Standard-Datenmodellen der ISO 191xx-Serie bzw. der INSPIRE Anhänge zeigt es sich, dass die praktizierte eindimensionale Modellierung nicht geeignet ist, themenübergreifende Strukturen aufzustellen. Einerseits wird versucht, die konkreten Anwendungsfälle und Phänomene im Modell abzubilden. Andererseits wird modularisiert, um Basisstrukturen aufzustellen und diese wiederverwendbar für beliebige andere Anwendungsfälle zu definieren und zu nutzen. Aus diesem Gebräu aus Modularisierungen, Basisklassen und ihren Ableitungen und der konkreten Nachbildung von Anwendungsfällen entsteht ein oft undurchsichtiger Modellmix. – Es macht selbst jedem „UML-Insider“ Mühe, die so entstandenen Modelle zu durchschauen, geschweige denn, sie für abgewandelte Aufgaben anzupassen, auszubauen bzw. zu verändern.

Wenn wir schon mit Multiplizitäten arbeiten, dann müssen wir anerkennen, dass wir verschiedene (*) Sichten auf unsere (einmalige) Realität haben. Diesem Gedanken folgend, modelliere ich zweistufig:

- ein „Systemanalysemodell“ als (abstraktes) theoretisches Abbild der Realität
- „Anwendungsmodelle“ zur Darstellung von Anwendungsfällen.

Das Systemanalysemodell wird so aufgesetzt, dass es problemlos möglich wird, das Modell fachlich zu erweitern. So können wir z. B. bestimmte Themen detailliert ausarbeiten, neue Themen ergänzen, beliebige Verknüpfungen jederzeit aufnehmen. Wir können es uns wie ein Netzwerk vorstellen, dass immer weiter und enger gespannt werden kann.

Sind im Systemanalysemodell noch alle, auch fachlich unsinnige Verknüpfungen denkbar, so wird in den Anwendungsdiagrammen festgeschrieben, was tatsächlich praktisch relevant ist. Das Anwendungsmodell ist frei von abstrakten, komplizierten und komplexen Abbildungen. Es ist geradlinig und enthält lediglich Objektklassen und Bewegungen zwischen ihnen. Und genau damit haben wir m. E. eines der Grundprobleme der schweren Verständlichkeit der ISO 191xx- und INSPIRE-Modelle überwunden. Für unsere Fachanwender heißt das: sie haben sogar die Chance, ein UML-Klassendiagramm zu verstehen. Und für uns Modellierer bedeutet das: wir sind gezwungen, auf die Nutzer einzugehen.

Und es entsteht ein weiterer Effekt: Eine Datenhaltung (beschrieben mit dem Systemanalysemodell) kann für viele Anwendungen genutzt werden!

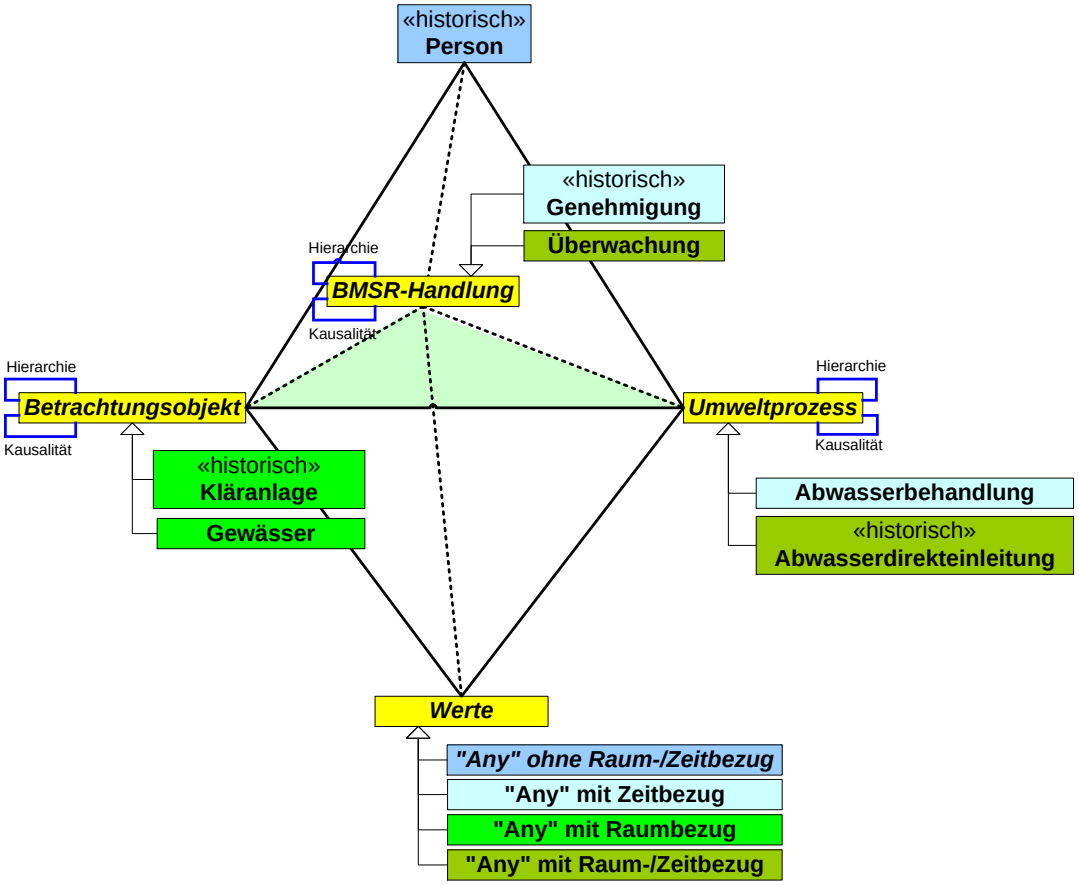


Abbildung 2: Auszug aus einem Systemanalysediagramm: Abwasserdirekteinleitung

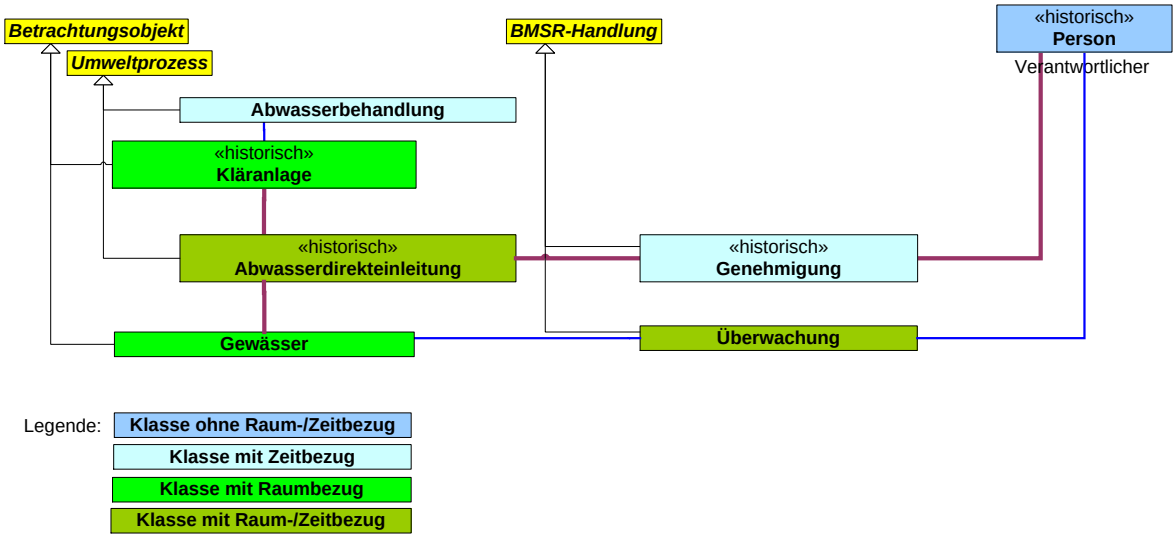


Abbildung 3: Beispiel für ein Anwendungsdiagramm: Abwasserdirekteinleitung

7 INSPIRE.Insight

Anwendungsbeispiele sind in [1], [2], [3] und [4] beschrieben.

Die Datenhaltung organisieren wir weder nach Berichtsschablonen noch nach INSPIRE. Die Daten werden themenübergreifend entsprechend des Strukturanalysediagramms verwaltet. Damit sind sie an keine Berichterstattungen gebunden; und auch nicht „INSPIRE-konform“ (was auch immer sich hinter diesem häufig verwendeten Begriff verbirgt) – aber sehr flexibel verwend-, auswert- und erweiterbar.

Auf dieser (ersten) Stufe, der Datenhaltung, finden wir sehr abstrakte Generalisierungen und Modularisierungen. Die eigentlichen Datenbereitstellungen definieren wir in den Anwendungsdiagrammen. Und diese sind dann sehr überschaubar. Es wird auch nicht nötig, auf dieser (zweiten) Stufe die komplexen INSPIRE-Strukturen der Datenspezifikationen nachzubilden.

Im Normalfall können wir die Anwendungsklassen direkt als Sichten (Filter- oder Verbundklassen) auf die Systemanalyseklassen aufsetzen. Und jetzt sind wir wieder „INSPIRE-konform“ – und zwar so, wie es angedacht ist: bei der Datenbereitstellung.

Oftmals ist das aber nicht mit den Performance-Anforderungen von INSPIRE vereinbar, z. B. dann, wenn aufwändige Datentransformationen für die Sichten notwendig werden. Dann lege ich die Daten sekundär in flachen Strukturen ab, die den zu liefernden INSPIRE-Objektlassen entsprechen.

Völlig analog können wir aus dem (einen) Datenpool auch andere Umweltberichterstattungen bedienen.

D. h. aber auch, dass die Datenhaltung tatsächlich unabhängig von den bereitzustellenden Daten organisiert werden kann. Wir können die Daten verwalten, wie wir wollen (dann natürlich sehr nachhaltig verwendbar – z. B. entsprechend dem Doppeltetraeder-Modell) und können alle (!) Bereitstellungen recht einfach liefern. – Also eine Datenhaltung und viele Berichterstattungen. INSPIRE inklusive! (Vergleiche mit [1], Abschnitt 16.)

8 Quellenverzeichnis

[1] Rudolf, H.
Umweltdatenmanagement. – Eine Geo-Inspiration
http://www.moss.de/deutsch/aktuelles/veroeffentlichungen/veroeffentlichungen.html
abgerufen am 30.06.2015

[2] Müller, U.; Rudolf, H.
Umweltdaten & INSPIRE – umgesetzt an einem Beispiel der Wasserwirtschaft in Sachsen
In: WasserWirtschaft 3/2012, S.31-34

[3] Rudolf, H.; Zulkowski, M.
Umsetzung von INSPIRE in einer neuen Geodateninfrastruktur für die Wasserwirtschaft in Sachsen
In: Strobl, J.; Blaschke, T.; Griesebner, G. (Hrsg.)
Angewandte Geoinformatik 2012 – Beiträge zum 24. AGIT-Symposium Salzburg, Wichmann, 2012

[4] Rudolf, H.
Umweltmanagement mit envVision – Eine nachhaltige GIS-gestützte Lösung für die Umsetzung der EU-Umgebungslärm- und Luftqualitätsrichtlinie
EnviroInfo, Bonn, 08.10.2010

INSPIRE Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE)
In: ABl. der EU, 2007, L 108, S.1-14

INSPIRE-GCM INSPIRE Generic Conceptual Model

ISO 191xx Standard-Serie zu Geoinformationen

Entwicklung eines prototypischen Client-/Serversystems zur mobilen Datenerfassung von VOC-haltigen Materialströmen in Automobillackieranlagen am Beispiel der Volkswagen AG

Ahmad Banna, Volker Wohlgemuth, Hochschule für Technik und Wirtschaft Berlin,
volker.wohlgemuth@htw-berlin.de

Steffen Witte, Roman Meininghaus, Volkswagen AG,
roman.meininghaus@volkswagen.de

Abstract

Automotive paint systems include very environmental relevant processes and are complex systems, which complicate the data acquisition for material flow analyses. With the help of a new developed client/server system, VOC including material data can be captured systematically and locally to build a better knowledge of these consumption data for future balances and calculations. The practicality of the developed IT solution was verified.

1 Einleitung

Im Vorfeld dieser Untersuchung wurden Massenbilanzierungen analysiert, die Lackieranlagen als größte Emittenten von Volatile Organic Compounds-Emissionen (VOC-Emissionen) in der Automobilbranche aufdeckten. Ihr Anteil an der Nutzung VOC-haltiger Lösemittel liegt bei ca. 90 % innerhalb des gesamten Fahrzeugherstellungs-prozesses [Meininghaus, 2014]. Automobillackieranlagen stellen daher besonders umweltrelevante Prozessschritte innerhalb der Automobilfertigung dar und sind von hoher Komplexität. Die Prozessschritte einer Lackierung sind zahlreich und die Materialversorgung ist von den Applikationsrobotern einige Meter entfernt, worunter u. a. die Datenqualität und Datenverfügbarkeit der Materialverbrauchsmengenangaben aus der Betriebsdatenerfassung (BDE) leiden. Gründe dafür sind u. a. fehlende

Angaben über die genaue Stelle der entstandenen Daten. Weitere Qualitätsmängel von Daten entstehen zudem durch Medienbrüche, wie das Ausdrucken von Tabellen, die handschriftliche Datenaufnahme und die manuelle Übertragung dieser in die betrieblichen IT-Systeme. Ein weiterer Grund sind fehlende Informationen über die Qualität, Herkunft und Ableitung der Daten [Vgl. Banna, 2014].

Emissionen, die durch den Einsatz von Lösemitteln verursacht werden, spielen in vielen Regionen der Welt eine wichtige Rolle bei der Entstehung des Sommersmogs [Vgl. EPA, 2015]. Da es organische Lösemittel gibt, die die Gesundheit beeinträchtigen können, ist der Gesetzgeber aus diesen Gründen an einer Einhaltung von Richtwerten interessiert [Vgl. Banna, 2014].

Die 31. Bundesimmissionsschutzverordnung (BImSchV) legt für Deutschland einen VOC-Emissionsgrenzwert von 35 g/m² fest. VOC-Emissionen entstehen, weil Lacke und andere Betriebsstoffe Lösemittel und Additive enthalten, die zum Materialinput dieser Prozesse gehören. Während der Nutzung dieser Materialien werden VOC in die Luft emittiert.

Eine gute Datenqualität ist die Basis aussagekräftiger Stoffstromanalysen. Um diese zu erreichen, ist eine gut durchdachte Datenerfassung von Nöten. Die Datenerfassung erfordert somit u. a. die Qualifizierung der Mitarbeiter und die hinreichende Planung hierfür. Darunter fallen die Einrichtung ablesbarer Messstellen und die Festlegung einer Erfassungsstrategie. Die Erfassungsstrategie setzt sich aus den Datentypen, dem Datenquellentypen und der daraus resultierenden Erfassungsmethode zusammen. Für qualitative Informationen müssen diese drei Parameter bekannt sein. Die folgende Abbildung verdeutlicht diese Notwendigkeit der Parameter.

Abbildung 1: Erfassungsstrategien [Banna, 2014]

Der Datentyp lässt sich in zwei Kategorien einordnen, nämlich in Primär- und Sekundärdaten. Zu den Primärdaten gehören direkt beobachtbare Variablen, wie zum

Beispiel das Ablesen eines Thermometers und direkt vorgenommene Messungen am Prozess. Sekundärdaten werden aus den Primärdaten abgeleitet. Hierzu gehören statistisch abgeleitete, berechnete, zusammengesetzte und simulierte Daten.

Datenquellen, wie Messstationen, können unterschiedliche Daten erhalten. Dies können sowohl berechnete als auch gemessene Daten sein. Die Kenntnis über den Datentypen trägt somit nicht zur Steigerung sondern auch zur Sicherstellung der Datenqualität bei.

Zentrales Ziel der vorliegenden Arbeit war es, ein neues Verfahren für die Datenerfassung der Materialeingangsströme zu entwickeln. Damit sollen die Datenqualität und -verfügbarkeit für Lackieranlagen gesteigert werden und somit transparentere, qualitativere Materialeingangsinformationen für u. a. zukünftige Analysen im Rahmen des Stoffstrommanagements gewährleistet werden. Mit Hilfe dieser neugewonnenen Informationen soll zudem ein neues Fundament für Analysen über den Materialverbrauch und deren Lösemittelemissionen gebildet werden [Banna, 2014].

2 Konzeption und Architektur

2.1 Analyse der Lackierprozesse in der Automobilbranche

Die Entwicklung eines softwarebasierten Verfahrens zur Datenerfassung erfordert im Vorfeld die Analyse der Prozesse verschiedener Lackieranlagentypen. Man kann hierbei zwischen Karosserie- und Komponentenlackieranlagen unterscheiden. Für die Studie wurden eine sehr große, komplexe, historisch gewachsene Karosserielackieranlage und eine moderne, kleine Komponentenlackieranlage am Standort Wolfsburg als Referenzmodelle herangezogen, um mit ihnen möglichst alle Varianten abzudecken. Trotz dieser technologischen Differenzen und der zu lackierenden Produkte unterscheiden sich die einzelnen Prozessschritte kaum voneinander. Die Prozesskette von Karosserielackieranlagen, in der VOC-haltige Lösemittel enthalten sind, besteht vereinfacht aus fünf abstrahierten Prozessschritten, die in der folgenden Abbildung 2 dunkel dargestellt sind. Der vorangestellte Prozessschritt Vorbehandlung (VBH) und die beiden letzten Schritte Dachverstärkung/Dämpfung (DVD) und Hohlraumkonservie-

rung (HRK) sind zu vernachlässigen, da diese Prozessschritte nahezu keine Materialien verwenden, die organische Lösemittel beinhalten [Vgl. Banna, 2014].

Abbildung 2: Prozessschritte einer Lackierung [Banna, 2014]

Die dargestellten Prozesse werden sequentiell abgearbeitet, wobei die Auftragung des Füllers in moderneren Lackierverfahren wegfallen kann. Die funktionellen Eigenschaften werden in diesem Fall von einem angepassten Basislack übernommen. In der kathodischen Tauchlackierung (KTL) wird der Karosserie, nach der Zinkphosphatschicht, die erste Lackschicht aufgetragen und dient als Haftungsgrundlage für die nachfolgenden Lackschichten [Vgl. Witte & Boehnke, 2012]. Anschließend erfolgt die Abdichtung des Unterbodens an den Schwellern und Blechkanten mit PVC-Material, welches ebenfalls VOC-emittierend ist. Dies fällt hier unter den Namen Unterbodenschutz (UBS) und Nahtabdichtung (NAD) [Quoll, 2014]. Der Füller hatte ursprünglich u. a. die Aufgabe, Unebenheiten auf der Karosserie auszugleichen, um eine möglichst glatte Fläche für den farbgebenden Basislack (engl. Base Coat BC) bereitzustellen. Nach der BC-Applikation erfolgt die Applikation des Klarlackes (engl. Clear Coat CC) [Vgl. Banna, 2014].

2.2 Konzeption

Der entwickelnde Lösungsansatz ist ein verteiltes Softwaresystem, das aus einem mobilen Client, mit dem die Materialeingangsströme systematisch und vor Ort erfasst werden können, einem Desktop-Client zur anschließenden Analyse und Auswertung am Arbeitsplatz und einem Server-Client zur Persistierung der gesammelten Daten [siehe Abbildung 3] besteht. Ein Teil der Daten des Desktop-Clients dient weiterhin als Input für ein weiteres Volkswagen eigenes VOC-Bilanzierungstool, das auf Microsoft Excel basiert. Die benötigten Daten umfassen u. a. tabellarische Jahresübersichten über die Materialverbrauchsmengen, deren VOC-Anteile und die Größe der lackierten Flächen.

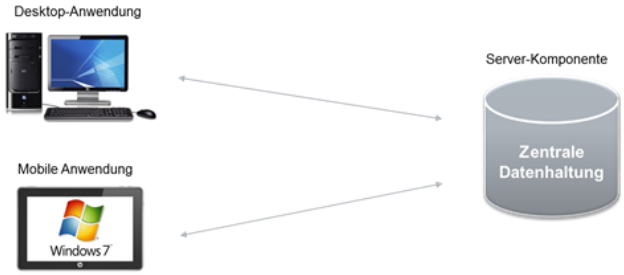


Abbildung 3: Konzept der IT-Lösung [Banna, 2014]

2.3 Architektur

Das entwickelte Client-/Serversystem basiert technologisch primär auf dem OpenResKit-Ansatz, der an der Hochschule für Technik und Wirtschaft Berlin entwickelt worden ist [Vgl. Wohlgemuth et. al, 2014]. OpenResKit stellt einen Softwarebaukasten mit Werkzeugen und Methoden bereit, der es ermöglicht, Open-Source basierte Anwendungen zur Unterstützung der Ressourceneffizienzproblematik schneller zu entwickeln [Vgl. Wohlgemuth et. al, 2014]. Die Vision des OpenResKit-Ansatzes ist es, mittels der Client-/Serverarchitektur den einzelnen Bereichen eines Unternehmens die entsprechenden kleinen Tools zur Verfügung zu stellen, die über eine zentrale Schnittstelle miteinander kommunizieren. So können Clients für die jeweiligen Anwendungsdomänen erstellt werden und deren Informationen zusammenfließen, sodass deren Datenquelle dieselbe bleibt.

Das Konzept dieser Client-/Serverarchitektur des OpenResKit- Ansatzes erläutert die folgende Abbildung 4. Der hier aufgeführte OpenResKit Hub stellt die zentrale Schnittstelle aller beteiligten Anwendungen dar und verwaltet die Datenbank. Sowohl mobile als auch Desktop- und Web-Clients sind an den Hub anbindbar, wobei die Kommunikation mit mobilen Endgeräten asynchron anstatt synchron vonstattengeht. Dies gilt jedoch nur für mobile Endgeräte, die auf den Betriebssystemen Android, iOS oder Windows Phone 8 basieren.

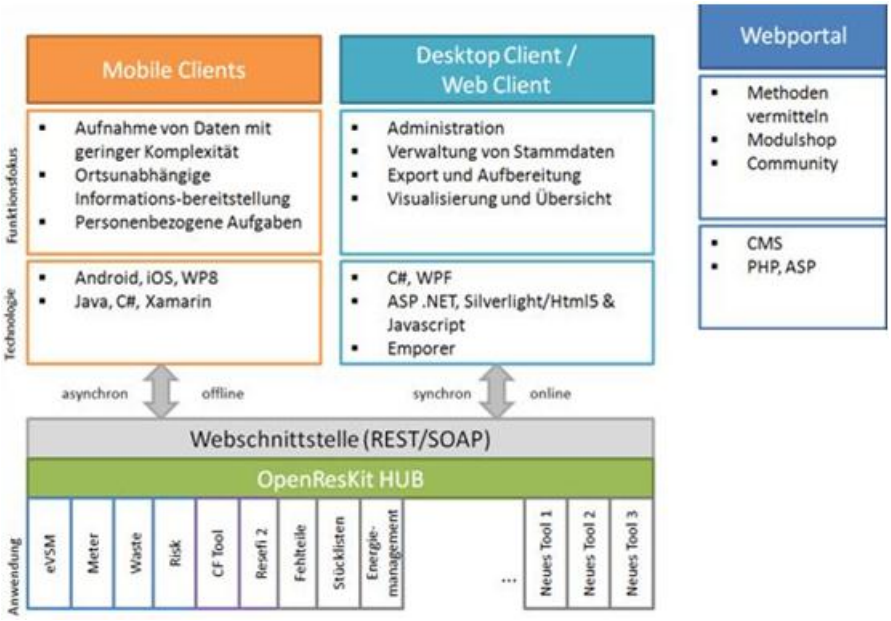


Abbildung 4: Konzept des OpenResKit-Projektes [Wohlgemuth et. al, 2014]

Das zu entwickelnde Softwaresystem basiert auf der Architektur der Abbildung 4. Da jedoch solche mobilen Endgeräte, wie sie eben genannt worden sind, im Volkswagen-Konzern nicht genutzt werden und aktuell keine Web-Clients existieren, wurde das Konzept so wie in Abbildung 5 gezeigt angepasst:

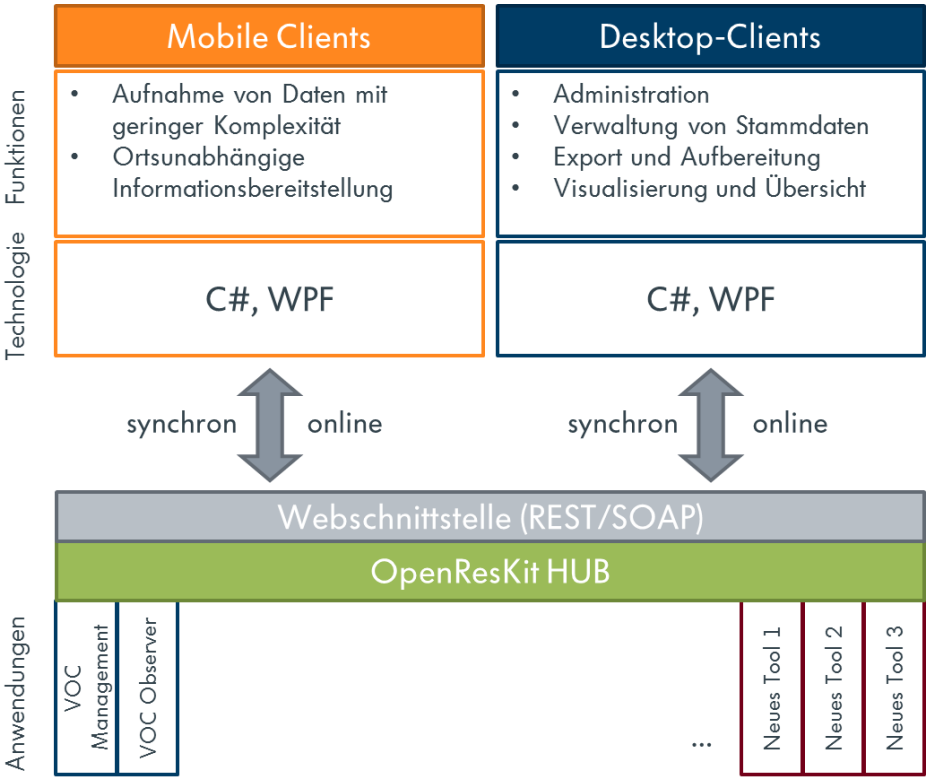


Abbildung 5: Konzept des Volkswagen-Hubs [Vgl. Wohlgemuth et. al, 2014]

Dadurch, dass die mobile Anwendung auf Tablet-PCs angewandt werden soll, auf denen Windows-Betriebssysteme betrieben werden, wurde die mobile Anwendung, genau wie die Desktop-Anwendung, in der Programmiersprache C# unter Verwendung der Graphik-Bibliothek Windows Presentation Foundation (WPF) entwickelt. In der vorliegenden Studie war es deshalb möglich, die asynchrone Kommunikation zwischen dem Hub und mobilen Anwendungen durch eine synchrone Kommunikation zu ersetzen. Das Webportal entfällt, da eine Kommunikation mit der Öffentlichkeit nicht vorgesehen ist. Passgenaue Domänenmodelle für die Desktop-Anwendung „VOC Management“ und die mobile Anwendung „VOC Observer“ wurden für den OpenResKit HUB konzipiert, angelegt und entwickelt. Der mobile Client dient hierbei der systematischen ortsunabhängigen Datenaufnahme während der Anlagenbegehungen. Der Desktop-Client ist die Administrationsanwendung, in der die Stammdaten verwaltet und die Anlagenbegehungen erstellt, bearbeitet und ausgewertet werden können. Der Server-Client (OpenResKit-HUB) beinhaltet die Datenbank, auf dem die entsprechenden Clients per sog. REST-Schnittstelle zugreifen (Vgl. Abbildung 5).

3 Entwicklung

3.1 Genutzte Entwurfsmuster

Während der Entwicklung wurde auf die konsequente Nutzung von Entwurfsmustern Wert gelegt. Entwurfsmuster beugen bekannten Fehler in der Softwarearchitektur und -entwicklung vor und stellen bewährte Ansätze dar. Dafür werden Erkenntnisse aus Erfahrungen in wiederverwendbarer Form zusammengefasst, die in Software-Projekten das Risiko von Entwurfsfehlern deutlich senken [Eilebrecht & Starke, 2013]. Entwurfsmuster fördern weiterhin die Kommunikation. Dadurch, dass sie Namen besitzen, können Struktur und Funktionsweise eines Programmes besser kommuniziert werden. Durch die klaren Strukturen, die Entwurfsmuster bieten, wird die Einfindungszeit in fremde Software-Projekte erheblich reduziert.

Zur Entwicklungsunterstützung des Softwaresystems wurden mehrere Entwurfsmuster eingesetzt, nämlich das Factory-Method-, das Repository- und das Model-View-ViewModel-Entwurfsmuster (MVVM) [Vgl. Rheinwerk, 2014].

Das Factory-Method-Entwurfsmuster dient der Entkopplung des Clients von der konkreten Instanziierung einer Klasse. Die Instanziierungsalgorithmen werden in einer eigenen Factory-Klasse abstrakt ausgelagert. Diese delegiert die konkrete Objektinstanziierung wiederum an die verwendenden Klassen. Erst die nutzende Klasse spezifiziert das abstrakte Objekt. Eine solche Auslagerung der Objektinstanziierungen steigert die Wartbarkeit und Flexibilität bei Veränderungen eines Objekttyps [Vgl. Hauer, 2005].

Das Repository-Entwurfsmuster beinhaltet eine zentrale Datenhaltung, die den verschiedenen Clients die entsprechenden Daten gebunden zur Verfügung stellt.

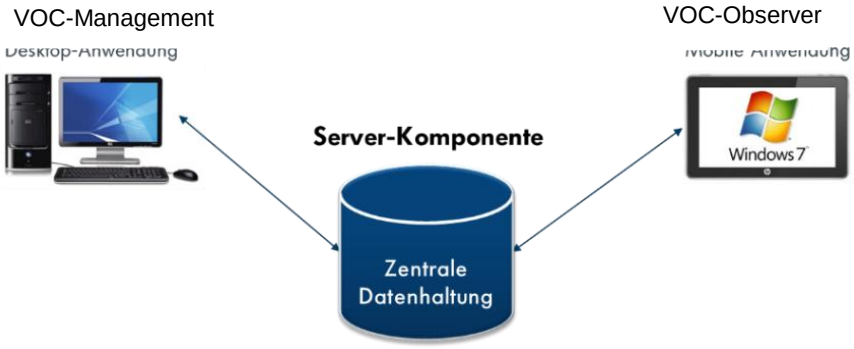


Abbildung 6: Datenzentralisierung [Banna, 2014]

Die Abbildung 6 zeigt zwei Clients, die untereinander keine Wechselwirkungen besitzen. Die Serverkomponente, welche das zentrale Repository beinhaltet, koordiniert den Informationsaustausch. Die Nutzung des Repository-Entwurfsmusters wird durch das gemeinsame Anwenden großer Datenmengen der Clients und dasselbe Format begründet. Der Vorteil durch die zentrale Datenhaltung ist zugleich eine Schwäche des Repository-Entwurfsmusters, denn durch den Transfer sämtlicher Daten sinkt die Performanz des Systems [Eilebrecht & Starke, 2013]. Um diesem Phänomen weitestgehend entgegenzuwirken, werden die zu übermittelnden Daten in den jeweiligen Repository-Klassen der entsprechenden Anwendungen definiert.

Für die Implementierung des Systems wird weiterhin das Entwurfsmuster MVVM gewählt, da es strikt die Programmlogik von der grafischen Benutzungsoberfläche (engl. Graphical User Interface, sog. GUI) trennt. Der größte Vorteil daraus ist der Erhalt der Übersichtlichkeit der zu entwickelnden Anwendung für zukünftige Wartungen. Einen weiteren Vorteil bildet die Entkopplung der Programmlogik von der GUI selbst, da Programmierer und GUI-Designer unabhängig voneinander arbeiten können. Die Implementierung des MVVM-Entwurfsmuster wird durch das Caliburn.Micro-Framework unterstützt, welches im nächsten Abschnitt näher erläutert wird.

3.2 Genutzte Frameworks und Bibliotheken

Ein Framework ist eine Teilanwendung, die wiederverwendet und an eigene Bedürfnisse angepasst werden kann. Damit das funktionieren kann, sind während der Entwicklung solcher Teilanwendungen die Wiederwendbarkeit und Erweiterbarkeit ihrer Klassen von höchster Bedeutung [Vgl. Brügge & Dutoit, 2004]. Klassenbibliotheken (engl. Class Library) dagegen sind passiv, das bedeutet, dass Klassenbibliotheken den Kontrollfluss einer Anwendung nicht beeinflussen.

Das entwickelte System basiert primär auf dem bereits erwähnten OpenResKit-Framework, und auf dem .Net-Framework von Microsoft. Diese Frameworks können in der Entwicklungsumgebung Visual Studio unter der Programmiersprache C# genutzt werden. Außer der Architektur des OpenResKit-Frameworks wurden weiterhin Klassen, wie die Kalender-Klasse zur Planung und Erstellung der Anlagenbegehungen, die Map-Klasse zur Begehungsunterstützung in der Lokalisierung der Datenquellen, die Settings-Klasse zur Verbindungsherstellung zum Server-Client und das OpenResKit-Theme zur Bereitstellung mehrerer Visualisierungsmöglichkeiten wiederverwendet. WPF ist eine Bibliothek des .Net-Frameworks, die für die Umsetzung intensiv eingesetzt wird. WPF-Anwendungen werden zur Visualisierung von der eXtensible Application Markup Language (XAML) unterstützt. Um die Nutzung des MVVM-Entwurfsmusters zu vereinfachen, wird das Caliburn.Micro-Framework herangezogen. Durch seine strenge Namenskonvention wird eine .cs-Klasse einer XAML-Datei zugeordnet [Vgl. Banna, 2014].

Die Datenbank für die Serverkomponente wird unter Verwendung des EF (Entity Framework) implementiert. Dieses Framework unterstützt die Code First-Konvention. Die Code First-Konvention ist zugleich eine der jüngsten Ansätze, Datenbanken zu erstellen und eine Vorgehensweise, die entwicklerorientiert ist. Sie ermöglicht Software-Entwicklern eine Datenbank in ihrer gewohnten Umgebung mittels einer Programmiersprache zu implementieren. Mit Hilfe des EF und der Code First-Konvention muss der Großteil des Datenzugriffscodes, welchen Entwickler ohnehin programmieren, folglich kein weiteres Mal geschrieben werden [Vgl. Grech, 2014].

Emgu CV ist eine Klassenbibliothek, die für die Bilderkennung herangezogen wird, um die Datenquellen eindeutig voneinander zu differenzieren. Sie stellt Methoden zur Verfügung, die es ermöglicht, mit einfachen Mitteln geräte- und herstellerunabhängig verschiedene Kameras anzusprechen und Aufnahmen darzustellen. Durch die ZXing-Klassenbibliothek können Zeichenketten zu QR-Codes kodiert und wieder dekodiert werden. So kann durch die kombinierte Nutzung der ZXing- und der Emgu CV-Klassenbibliotheken ein durch die Kamera erfasster QR-Code dekodiert werden [Vgl. Banna, 2014].

4 Beispielanwendung

Das erstellte Client-/Serversystem wurde bei der Volkswagen AG zur Erfassung von VOC-relevanten Stoffströmen in zwei Lackierereien mithilfe eines typischen Anwendungsbeispiels praktisch und prototypisch erprobt, um die Betriebstauglichkeit nachzuweisen.

Der reguläre Workflow mit dem Softwaresystem beginnt in der Desktopanwendung nach dem Anlegen der Stammdaten. Zu diesen gehören u. a. die Werke, Gewerke, Materialien und die zu lackierenden Produkte. Unter dem Reiter „Planung“ (siehe Abbildung 7) können Begehungen für die entsprechenden Lackieranlagen und Materialien generiert werden. Um das Anlegen der Begehungen effizient zu gestalten, ist es möglich, Regeltermine für mehrere Materialien zugleich zu erstellen. Diese

Begehungen sind ebenso in der mobilen Anwendung einsehbar, um die jeweiligen Begehungen durchzuführen [Vgl. Banna, 2014].

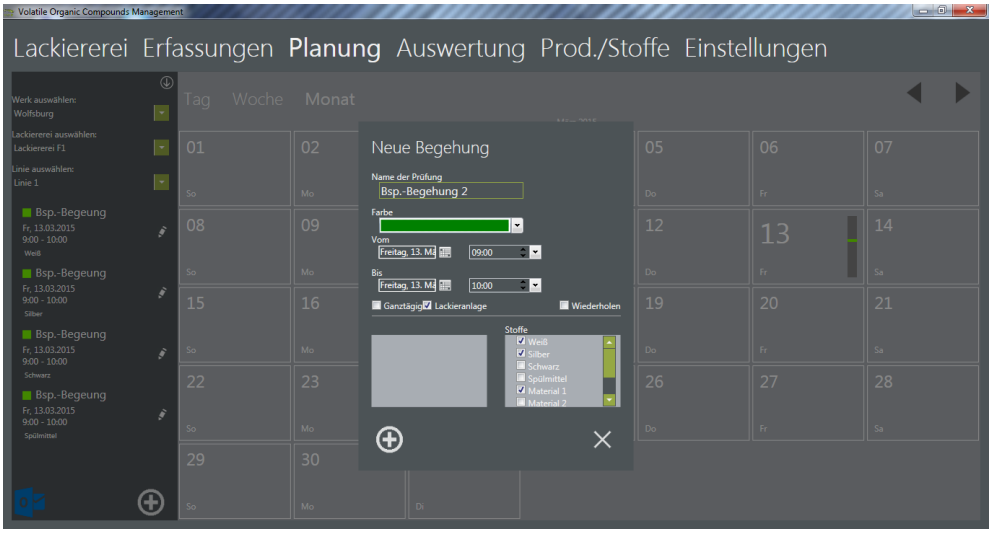


Abbildung 7: Begehungen anlegen [Vgl. Banna, 2014]

Zur Unterstützung des Begehenden können Datenquellen mittels Kartenaufrufen lokalisiert werden. Die nachfolgende Abbildung zeigt solch eine Karte im VOC-Observer, auf der zwei Datenquellen durch Punkte dargestellt werden.

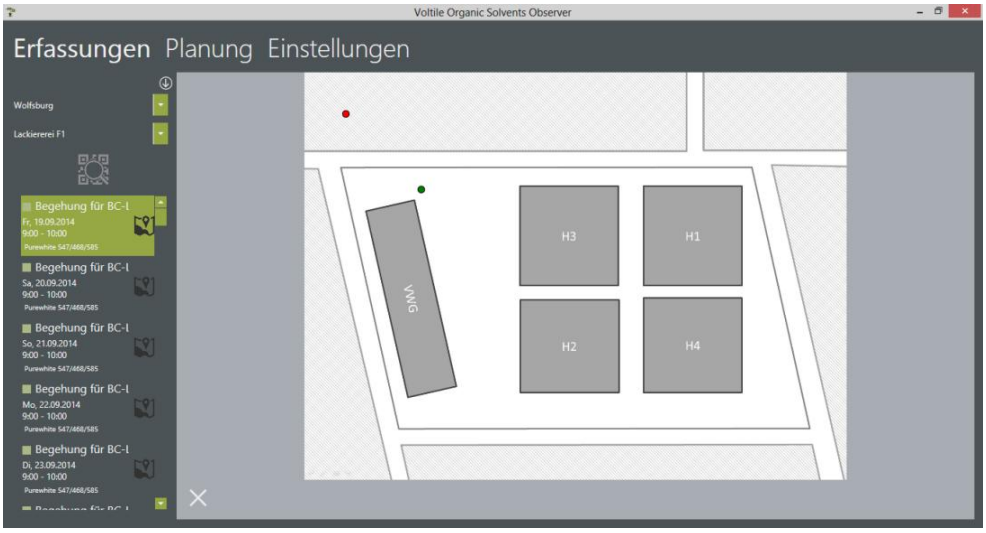


Abbildung 8: Lokalisierung [Banna, 2014]

Da Datenquellen wie Farbbehälter optisch meist kaum voneinander unterscheidbar sind, wurde eine Identifizierungsfunktion über QR-Codes implementiert. Wird der entsprechende QR-Code über die Kamera erkannt, werden die passenden Eingabefelder aufgerufen, sodass die dazugehörigen Informationen eingegeben werden können. Alternativ kann die entsprechende Begehung auch manuell ausgewählt

werden. Mit dem Speichern dieses Erfassungsbogens werden die Informationen an den Server-Client gesandt und dort persistiert [Vgl. Banna, 2014].

Anschließend können die erfassten Daten in der Desktop-Anwendung nachbereitet werden. Unter dem Reiter „Auswertung“ in der Desktopanwendung existieren zwei tabellarische und eine grafische Auswertungen (siehe Abbildung 9). Die Tabellen dienen dem Export für das VOC-Bilanzierungstool der Volkswagen AG. Die grafische Auswertung ist lediglich dafür gedacht, einen schnellen visuellen Überblick über die Materialverbrauchsverhältnisse zu erhalten. Einen solchen Überblick zeigt die folgende Abbildung mithilfe von Beispieldaten [Vgl. Banna, 2014].

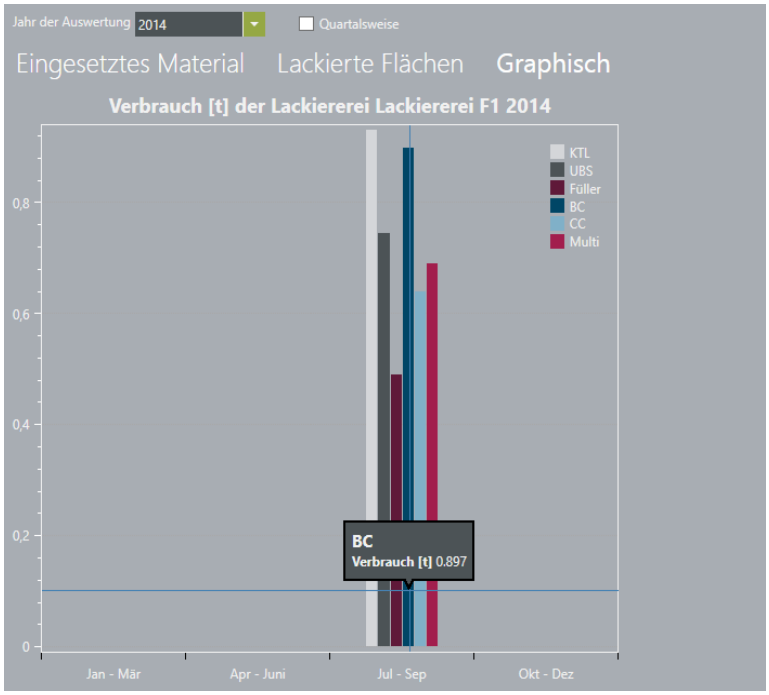


Abbildung 9: Grafische Auswertung [Vgl. Banna, 2014]

Das Client-/Serversystem wurde durch eine Pilotierung getestet, um einen realen Einsatz zu simulieren und somit vorhandenes Wartungspotential aufzudecken und zu nutzen. Unter anderem wurden die Testvarianten Usabilitytest durch Rapid-Prototyping, Akzeptanztest und Regressionstest intensiv angewandt [Banna, 2014].

Das Rapid-Prototyping ist eine Methode, die zur Verbesserung der Usability beiträgt [Vgl. Onlinemarketing-Praxis, 2014].

Sie findet ihren Einsatz in den Entwurfs- und Implementierungsphasen. Prototypen der Software-Anwendungen sind zu Beginn nur bedingt funktional, da sie immer nur auf gerade aktuell benötigte Aspekte fokussieren. Das Rapid-Prototyping wurde regelmäßig mit potentiellen Benutzern der jeweiligen Clients angewandt, um Interaktionen zwischen den Benutzern und der Software durch deren Feedbacks besser zu designen. Zu den getesteten Eigenschaften gehörten unter anderem der Workflow, die Anordnung und Gestaltung der Steuerelemente und ob deren Bezeichnungen und Symbolik verständlich und intuitiv sind.

Der Akzeptanztest erfolgte in Form der Pilotierung im Volkswagenkonzern. Nach dem Akzeptanztest wurden aufgedeckte Wartungspotentiale ausgeschöpft und vergangene Testfälle wiederholt, um Kausalitäten von Korrekturen auszuschließen. Diese Phase des Testens wird als Regressionstest bezeichnet [Vgl. Liggesmeyer, 2009].

Um eine Steigerung der Datenqualität feststellen zu können, wurden mittels des entwickelten Client-/Serversystems die Verbrauchsmengen der Kunststoffteilelackiererei des Oktobers 2014 ermittelt, um diese den durch die BDE ermittelten Verbrauchsmengen gegenüberzustellen. Insgesamt beträgt der Unterschied zwischen den beiden Systemen ca. 14,5 %. Diese Abweichungen könnten möglicherweise durch die mangelhafte Dokumentation und der fehlenden Pilotierung am Wochenende hervorgerufen worden sein. Lediglich bei zwei Decklacken erstrecken sich die Abweichungen um über 100 % [Vgl. Banna, 2014].

In der Karosserielackieranlage sind die erfassten Informationen mit den eingepflegten Informationen der BDE identisch, was auf die Korrektheit des entwickelten Systems hindeutet und verifiziert, dass die Datenqualität der Materialverbrauchsmengen, zumindest für die zwei Decklacke, in der Kunststoffteilelackiererei gesteigert werden konnte [Vgl. Banna, 2014].

Zur Vollständigkeit der Daten gehören die Typen der lackierten Produkte und deren Mengen in den jeweiligen Zeitintervallen, die nicht an den Datenquellen vor Ort zu erfassen sind, sondern am Arbeitsplatz aus der vom Hersteller mitgelieferten BDE ausgelesen werden müssen. Dieser Aufwand ist hoch und kann nicht ohne weiteres

reduziert werden. Lackierereien verwenden verschiedene BDE-Anwendungen aufgrund der unterschiedlichen Hersteller, wodurch diese Problematik auch nicht durch eine einheitliche Schnittstellendefinition aufgelöst werden könnte. Wie lange diese Informationen einem Produktionssystem zu entnehmen sind, ist aus diesem Grund auch unklar. Diese Informationen sind in der BDE der Kunststoffteilelackiererei beispielsweise lediglich nur einen Monat vorhanden. Daher empfiehlt es sich in diesem Falle, die Regeltermine der Begehungen auf drei Wochen zu begrenzen.

5 Zusammenfassung und Ausblick

Entstanden ist ein Client-/Serversystem, das die erforderlichen Eingangsinformationen für die Bilanzierung von VOC-Emissionen liefert. Die dafür relevanten Eingangsinformationen werden mit Hilfe des mobilen Clients vor Ort systematisch erfasst, indem der Erfassende durch einen strukturierten Arbeitsprozess geleitet wird. Die Erreichung des Zieles einer hohen Gebrauchstauglichkeit wurde durch das Rapid Prototyping verwirklicht. Das Ziel, den Zeitaufwand für die Begehungen zu reduzieren, konnte ebenfalls erreicht werden. Dieser wurde, zumindest in der Kunststoffteilelackieranlage, um ca. 50 % gesenkt. Dieser Wert wurde mithilfe einer Durchschnittszeit aus Begehungen ermittelt, die wiederum aus einem Stichprobenumfang von 5 Begehungen ermittelt wurde [Vgl. Banna, 2014].

Weiterhin werden mit diesem Verfahren die erwähnten Medienbrüche vermieden, wodurch das Fehlerpotential während der Datenaufnahme reduziert werden kann.

Die Diversität der eingesetzten Materialien ist höher als während der Konzeptionsphase gedacht. Dadurch leidet die Übersichtlichkeit in der Auflistung und der Zeitaufwand zum Anlegen dieser Stammdaten ist sehr hoch. Ein Excel-Import und Filterfunktionen würden diesen Aufwand entscheidend reduzieren [Vgl. Banna, 2014].

Eine weitere Ergänzung für das System, ist die Erweiterung um die Klasse „ResponsibleSubject“ aus dem OpenResKit-Framework. Mit ihr könnten Anlagenbegehungen Erfassern zugeordnet werden, um im Nachhinein nachvollzogen zu werden. Diese Zuordnung würde einen größeren Mehrwert des Client-/Serversystems für

Lackieranlagenbetreiber bilden, da diese Zuordnung für den realen Einsatz in der Praxis notwendig ist [Vgl. Banna, 2014].

Insgesamt bleibt festzuhalten, dass eine verteilte, mobile Unterstützung der Datenerfassung für das Stoffstrommanagement bei der Volkswagen AG eine erfolgversprechende Möglichkeit darstellt, die Datenqualität relevanter Stoffströme deutlich zu verbessern und damit sowohl ökonomische als auch ökologische Einsparpotentiale aufzuzeigen [Vgl. Banna, 2014].

6 Literaturverzeichnis

[Banna, 2014]

Banna, A., 2014. Entwicklung eines Verfahrens zur Vor-Ort-Erfassung der Materialströme in Lackieranlagen eines Automobilherstellers für die Bilanzierung von VOC-Emissionen; Masterarbeit an der Hochschule für Technik und Wirtschaft Berlin. Seite 3, 12, 17, 19, 22, 31, 67, 75, 77, 79-83

[Brügge & Dutoit, 2004]

Brügge, B. & Dutoit, A. H., 2004. Objektorientierte Softwaretechnik mit UML, Entwurfsmustern und Java. München; Boston [u.a.]: Pearson Studium. Seite 364-366

[EPA, 2015]

United States Environmental Protection Agency. [Online]; Available at: http://www.epa.gov/groundlevelozone/basic.html [Zugriff am 16.10.2014].

[Eilebrecht & Starke, 2013]

Eilebrecht, K. & Starke, G., 2013. Patterns kompakt. 4. Aufl. Hrsg. Berlin: Springer Vieweg. Seite 293

[Rheinwerk, 2014]

Rheinwerk Verlag GmbH, 2014. [Online]; Available at: http://openbook.rheinwerk-verlag.de/visual_csharp_2012/1997_28_005.html [Zugriff am 16.10.2014].

[Grech, 2014]

Grech, S., 2014. Model First vs Code First. [Online]; Available at: http://www.incredible-web.com/blog/model-first-vs-code-first [Zugriff am 16.10.2014].

[Hansmann, 2013]

Hansmann, K., 2013. Bundes-Immissionsschutzgesetz. 31. Aufl. Hrsg. Baden-Baden: Nomos. Seite 578, 579, 603

[Hauer, 2005]

Hauer, P., 2005. Das Factory Method Design Pattern. [Online]; Available at: http://www.philipphauer.de/study/se/design-pattern/factory-method.php [Zugriff am 31.10.2014].

[Liggesmeyer, 2009]

Liggesmeyer, P., 2009. Software-Qualität: Testen, Analysieren und Verifizieren von Software. Heidelberg: Spektrum Akademischer Verlag. Seite 192-194

[Meininghaus, 2014]

Meininghaus, R., 2014. Der neue think Blue. Factory. Baustein für die Reduzierung der Lösemittelemissionen (VOC) in den Standorten. Wolfsburg: K-EFUW/I - Volkswagen AG [Unveröffentlichter Vortrag vom 28.07.2014].

[Onlinemarketing-Praxis, 2014]

Onlinemarketing-Praxis, 2014. Definition Rapid Prototyping. [Online]; Available at: http://www.onlinemarketing-praxis.de/glossar/rapid-prototyping; [Zugriff am 31.10.2014].

[Quoll, 2014]

Quoll, M., 2014. Lackieranlagenprozesse. Wolfsburg. [Interview]

[Witte & Boehnke, 2012]

Witte, S. & Boehnke, B., 2012. Parametrisiertes benutzergeführtes Simulationsmodell einer idealisierten Automobillackiererei. In: V. Wohlgemuth, Hrsg. Einsatz der Software UMBERTO in angewandter Forschung und Praxis. Aachen: Shaker Verlag. Seite 103-104

[Wohlgemuth et. al, 2014]

Wohlgemuth, V., Krehan, P., Ziep, T., Schiemann, L., 2014. Entwicklung eines Open-Source basierten Baukastens zur Unterstützung und Etablierung der Ressourceneffizienz in produzierenden KMU. Konzepte, Anwendungen und Entwicklungstendenzen von betrieblichen Umweltinformationssystemen (BUIS). Hrsg. Wohlgemuth, Lang, Gómez. Aachen: Shaker Verlag. Seite 41-57

Praxisbericht: Business Intelligence mit Geovisualiserung

Frank Reußner, DB Mobility Logistics AG, Frank.Reussner@deutschebahn.com
Benedikt Süßmann, Hochschule RheinMain, Benedikt.Suessmann@googlemail.com
Christian Ullerich, DB Mobility Logistics AG, Christian.Ullerich@deutschebahn.com
Michael Schoch, DB Mobility Logistics AG, Michael.Schoch@deutschebahn.com

Abstract

Analyzing data with business intelligence systems (BI systems) by means of visual analytics is increasingly the focus of companies. These systems are very powerful regarding the interaction between user and data. Data can be processed as charts or tables, which are then combined into dashboards or cockpits. However, a weakness of BI systems is an analysis of data with focus to its localization. Even if standard BI systems support geo-visualization of data it can only be used for rather simple analysis. For this reason, we have explored the market to find a BI system that can be extended with additional components to have an advanced geo-visualization. This article describes the searching for a BI system and the extending of a BI system with geo-visualization.

Zusammenfassung

Mit Business-Intelligence-Systemen (BI-System) Daten mittels Visual Analytics zu analysieren, ist immer stärker im Fokus von Unternehmen. Die BI-Systeme sind sehr leistungsstark bezüglich der Interaktion zwischen Benutzern und Daten. Daten können als Charts oder Tabellen aufbereitet werden, die dann zu Dashboards oder Cockpits zusammengefügt werden. Eine Schwäche der BI-Systeme ist aber die Auswertung der Daten hinsichtlich ihrer Lokalisierung. Eine weitergehende Geovisualisierung ist im

Standard der Produkte nicht enthalten. Aus diesem Grund haben wir den Markt sondiert, um ein BI-System zu finden, das mit weiteren Komponenten zu einem BI-System mit Geovisualisierung erweitert werden kann. Dieser Aritkel beschreibt zum einen die Suche nach dem BI-System. Zum anderen wird die Erweiterung mit Geovisualisierung beschrieben.

1 Einleitung

Aus Daten Wissen zu generieren ist ein Grundbestreben jeder Datenerhebung. Um dies zu unterstützen, werden Daten in unterschiedlicher Art aggregiert. Für die Aufbereitung der Daten in eine aggregierte Form hat sich der Begriff Business Intelligence (BI) etabliert. War ehemals BI hauptsächlich auf den Datenkollektions- und Aggregierungsprozess bezogen, ist heute auch der Datenauswertungsprozess Teil von BI. Zudem ist ein Wandel zu beobachten, dass BI-Systeme zunehmend nicht mehr von einzelnen Spezialisten genutzt werden, sondern einer breiteren Masse von Nutzern als Werkzeug zur Seite stehen.

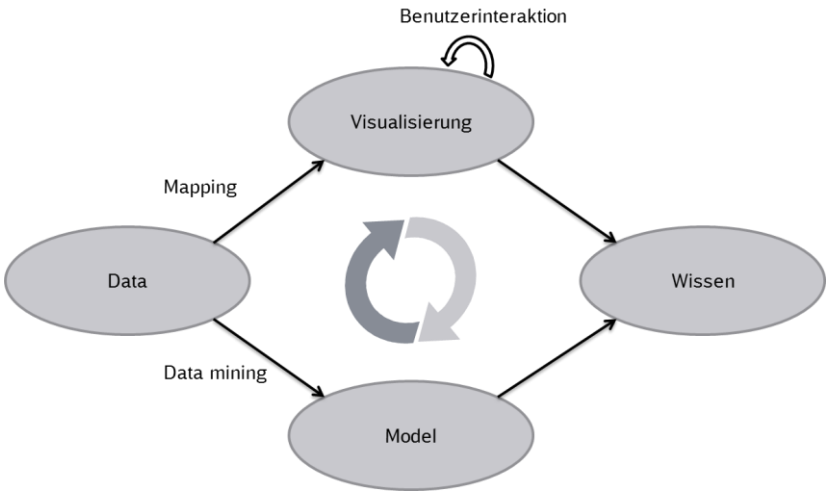


Abbildung 1: Visual Analytics (nach [„What is Visual Analytics?“ 2015])

Aus Anwendersicht haben sich Begriffe wie Visual Analytics (siehe Abbildung) und Self-Service-BI entwickelt. Letzteres bezieht sich darauf, dass der Nutzer nicht nur Berichte konsumiert, sondern aktiv bzw. interaktiv die Analysen erleben kann. Mit Visual Analyitics werden diese interaktiven Analysen durch grafische Auswertungen

sichtbar gemacht, so dass Sachverhalte leichter verständlich aufbereitet werden können [Thomas und Cook 2005].

Ein Einstieg in diese Welt der Datenanalyse mit BI ist für Umweltinformationssysteme sehr reizvoll. In der Vergangenheit sind diese Bestrebungen aber meist daran gescheitert, dass die anfallenden Daten überwiegend geografisch verortet sind und dementsprechend auch geovisualisiert werden sollten. Geeignete Werkzeuge hierfür waren allerdings kaum vorhanden. In den letzten Jahren hat sich dies geändert, so dass einfache Kartenvisualisierungen mit den BI-Werkzeugen möglich geworden sind.

Die Geovisualisierung von Massendaten ist für uns schon immer ein wichtiger Aspekt, was uns zur Suche nach geeigneten Lösungen zur Geovisualisierung im BI-Umfeld geführt hat. Im vergangenen Jahr haben wir deshalb den Markt dahingehend sondiert, ob Standardkartenvisualisierungen im BI-Umfeld für unsere Zwecke geeignet sind oder ob geeignete spezialiserte Lösungen unsere Anforderungen abdecken. Dieser Artikel beschreibt das daraus entstandene BI-System mit Geovisualisierung.

2 Ausgangslage

Der bisherige Prozess der Datenaufbereitung und Analyse kann in vier Phasen unterteilt werden:

- 1. Sammlung,
- 2. Modellierung/Aufbereitung,
- 3. Visualisierung und
- 4. Analyse.

Die Daten werden aus unterschiedlichen Systemen bezogen, sind aber in der Regel immer strukturiert. So werden die Daten meist in SQL-Server-Datenbanken persistiert – sowohl als relationale Datenbank als auch als OLAP-Cube. Ferner werden auch Excel-Dateien oder andere strukturierte Dateien (z.B. XML) als Quellen genutzt.

Beim Prozess der Aufbereitung kommen unterschiedliche Werkzeuge zum Einsatz. Herauszuheben ist dabei die geografische Aufbereitung, die meist mit Hilfe des

Programms PTV Visum erstellt wird. PTV Visum ist eine auf Logistik und Transport spezialisierte Software für Verkehrsanalysen, Verkehrsprognosen und eine GIS-orientierte Datenverwaltung.

Die Nutzung der aufbereiteten Daten (Prozess Visualisierung und Analyse) wird zurzeit mit der Erstellung eines Berichts abgeschlossen – meist in Form eines PDFs. Die Daten sind dabei nicht interaktiv auswertbar. Daneben gibt es den Zugriff auf die aufbereiteten Daten nur mit Hilfe der Spezialwerkzeuge. Für die geovisualisierten Daten heißt dies, dass eine Analyse nur mit dem Programm PTV Visum getätigt werden kann und dementsprechend Schulungs- und Anschaffungskosten anfallen.

3 Business Intelligence mit Geovisualisierung

Die Schwäche des bisherigen Prozesses ist vor allem die schwierige Weitergabe der Analyseergebnisse. Diese sind statisch und können nicht interaktiv genutzt werden. Eine interaktive Auswertung des Nutzers selbst ist aber ein großer Mehrwert. Zudem ist durch den Einsatz einfacherer Werkzeuge im Auswertungs- und Analyseprozess ein Freisetzen von Ressourcen innerhalb bestehender Auswertungs- und Analyseteams zu erwarten.

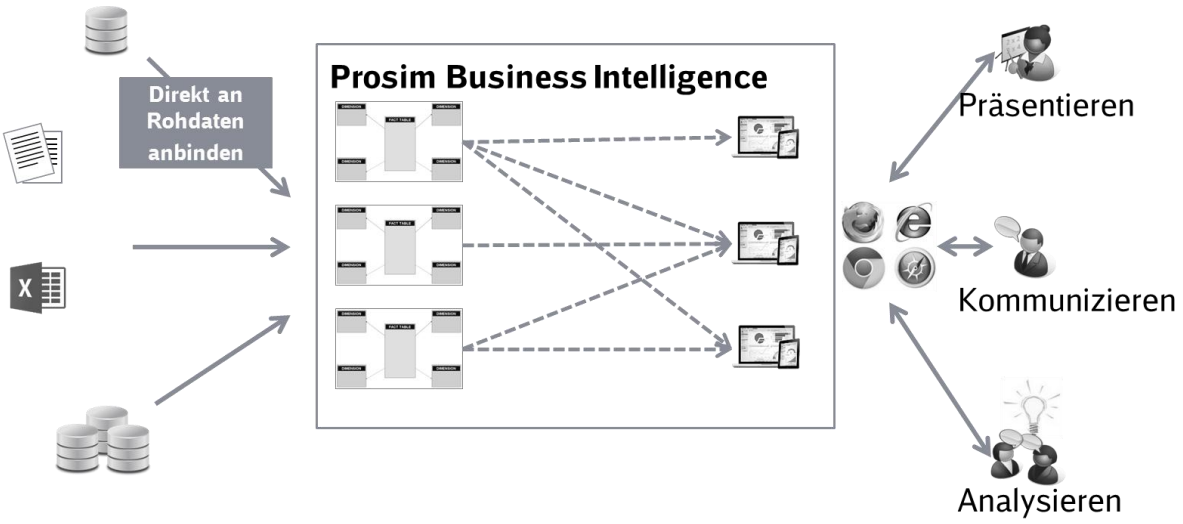


Abbildung 2: Aufgebautes Business Intelligence System

Es gibt einen viel zitierten Satz, dass über 80 Prozent aller Daten geografisch verortet werden können. Sehr wahrscheinlich geht dieser Satz auf ein Paper aus dem Jahr 1987

[Williams 1987] zurück und ist wohl in dem jetzigen Rahmen nicht mehr aussagekräftig. Aktueller ist da die Untersuchung aus dem Jahr 2013 [Hahmann und Burghardt 2013], die immerhin bei Wikipedia Artikeln etwa 60 Prozent der Daten als potentiell geoverortbar sieht. In unserem Anwendungsrahmen ist zu erwarten, dass wir deutlich über diesen 60 Prozent liegen. Aus diesem Grund haben wir im vergangenen Jahr ein geeignetes BI-Werkzeug gesucht, das auch eine Geovisualisierung der Daten ermöglicht. Bei diesem Evaluierungsprozess haben wir insbesondere die führenden Anbieter betrachtet und besonders auf die Fähigkeit der Geovisualisierung geachtet. Dabei wurde früh erkennbar, dass es keine integrierte Lösung gibt, sondern zwei unterschiedliche Lösungen nötig sind – ein BI-Werkzeug, das die Standardfunktionalitäten umfasst, und ein zusätzliches Modul zur Geovisualisierung.

Bei der Evaluierung für die BI-Werkzeuge lagen unsere Schwerpunkte auf

- visuelle Aufbereitung – Visual analytics [Visual Analytics 2015],
- Self-Service-BI,
- Erweiterbarkeit (insbesondere HTML5-Schnittstelle) sowie
- On-Premise-Lösung.

Die Geovisualierungskomponente sollte

- massendatentauglich sein,
- Punkt-, Region- und Streckeninformationen abbilden können sowie
- eine On-Premise-Lösung sein.

Einen guten Start beim Auswahlprozess lieferte uns der Gartner-Report für BI-Produkte [Sallam u. a. 2014]. Die Auswahl konnte dann auf wenige Anbieter eingeschränkt werden. Unsere Wahl fiel letztendlich auf ein Produkt von Qlik als BI-Werkzeug und NPGeoMap als Komponente zur Geovisualisierung [NPGeoMap 2015].

Abbildung 2 zeigt das von uns aufgebaute System als Skizze. Der Projektname ist Prosim View. Das BI-System ist eine Serveranwendung, auf die Nutzer mit marktüblichen Web-Browser direkt auf die Analysen zugreifen können. Zurzeit haben wir zwei Produkte von Qlik im Einsatz (QlikView und Qlik Sense). Für die Geovisualisierung

nutzen wir die Qlik-Extension NPGeoMap. Diese basiert auf HTML5 und Javascript. Als Basis für die Kartendarstellung wird dabei das Open-Source-Projekt „OpenLayers“ genutzt [OpenLayers 3, 2015]. Parallel zu dem erworbenen Kartenmodul haben wir eine eigene Kartenkomponente entwickelt. Auch als Qlik-Extension konzipiert, basiert diese auf dem Open-Source-Projekt „leaflet“ und ebenfalls auf Javascript [Leaflet 2015]. Dies war für uns von Vorteil, da wir damit die Möglichkeit haben, spezialisierte Kartenauswertungen auch selbst darstellen zu können – beispielsweise Streckeninformationen in Hin- und Rückrichtung als gestapelte Balkeninformationen je Richtung auf einer Karte. Abbildung 3 zeigt ein aktuelles Beispiel einer Anwendung.

3.1 Architektur BI-System mit Geovisualisierung

Abbildung 4 zeigt die Stufen der Geovisualisierung im Projekt Prosim View. In der ersten Stufe werden die Daten verortet (soweit zuvor nicht geschehen). Die Verknüpfung der Daten kann beispielsweise über eine einfache ID erstellt werden. Die zweite Stufe ist die eigentliche Visualisierung der Daten. Mit Hilfe des Kartenmodule werden die Daten dargestellt, die aber durch ihre direkte Anbindung an das BI-System mit allen anderen Charts interagieren. So wirkt sich beispielsweise eine Selektion/Filterung der Daten in der BI-Anwendung unmittelbar auf die dargestellten Daten in der Karte aus. Auch werden die Selektionen innerhalb der Karte, direkt an die BI-Anwendung weitergegeben.

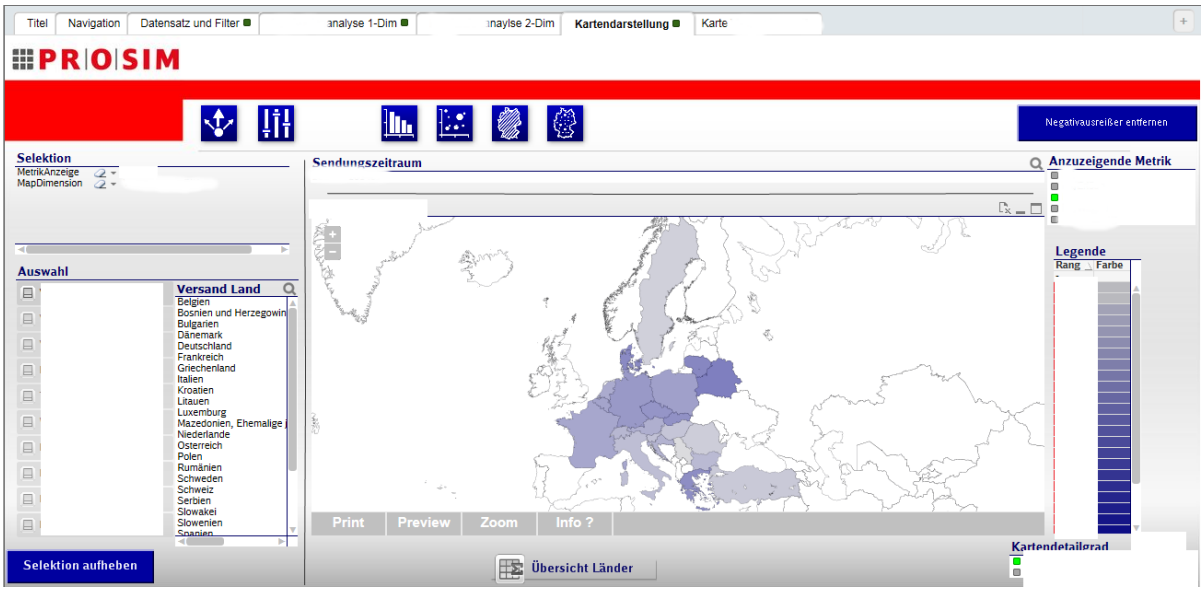


Abbildung 3: Beispiel einer Prosim-View-Applikation (anonymisiert)

Die eigentliche Anbindung der Geometriedaten wird dabei nicht über einen GIS-Server verwirklicht, sondern über die Anbindung von Kacheln (Tiles) und Shape-Dateien (siehe Abbildung 5). Dies hat den Vorteil, dass das BI-System mit Geovisualisierung schlank bleibt und der Client prinzipiell auch offline arbeiten kann. Auch sind die Anschaffungskosten gegenüber einem System mit GIS-Server deutlich günstiger. Wenn umfangreiche grafische Operatoren wie Verschneidung o. Ä. nicht im Vordergrund stehen, kann auf einen GIS-Server verzichtet werden.

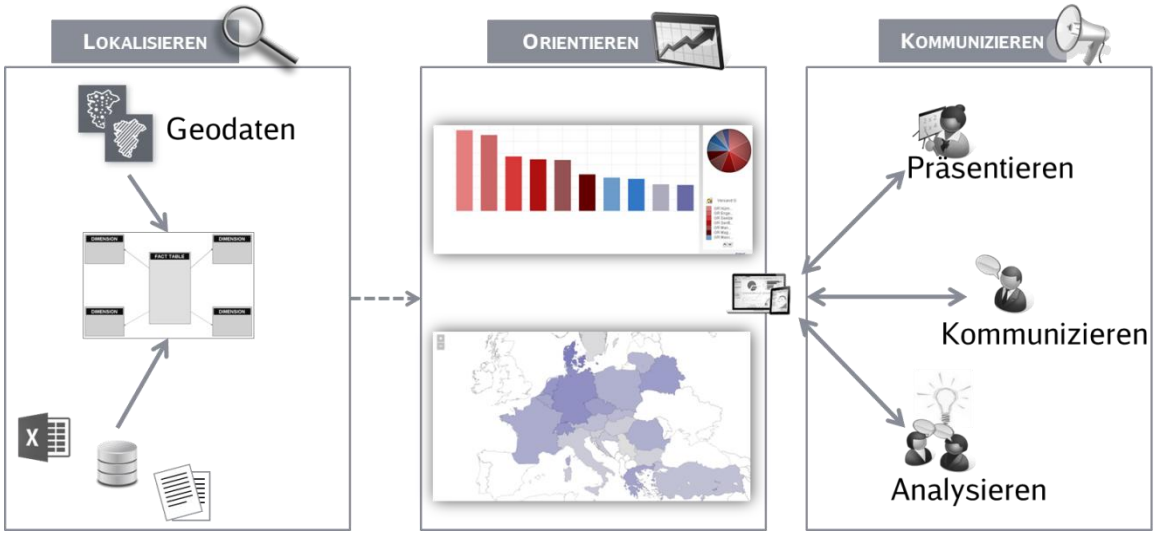


Abbildung 4: Stufen der Geovisualisierung

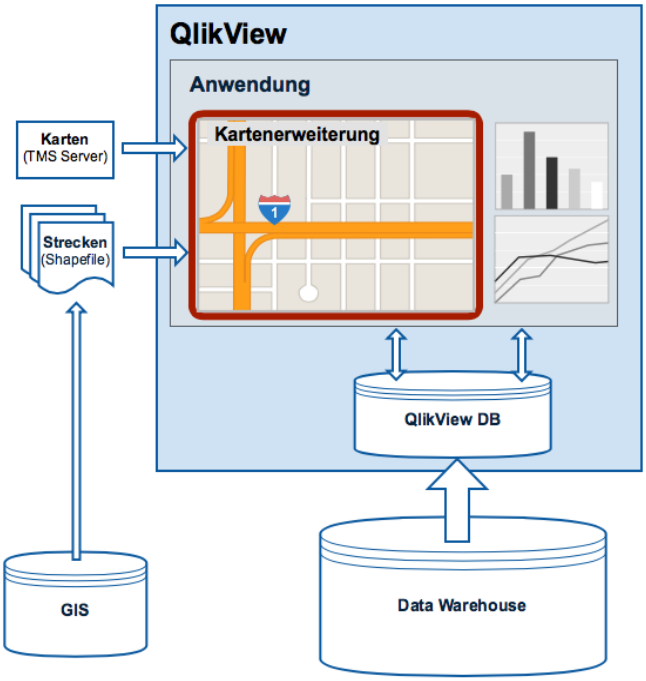


Abbildung 5: Datenanbindung und Serverkomponenten [Benedikt Süßmann 2015]

3.2 Erweiterbarkeit

Ein wesentliches Kriterium bei der Entscheidungsfindung war die Erweiterbarkeit des Systems. Zum einen hinsichtlich von Drittanbietern, zum anderen hinsichtlich der eigenen Anpassung. Insbesondere letzteres spielte eine große Rolle, da im Evaluierungsprozess schon früh klar wurde, dass z.B. der Anwendungsfall der gestapelten Balken auf Strecken von keinem Anbieter erfüllt werden kann.

Ein weiterer Aspekt war die Technologie zur Interaktion mittels Webseiten. Da das Projekt neu aufgesetzt wurde, waren wir nicht an eine bestimmte Technologie gebunden. Deshalb ist für uns HTML5 als Standard als geeigneter eingeschätzt worden, als beispielsweise eine Lösung, die auf Flash basiert. Ein Grund dafür ist, dass Lösungen mit HTML5 eine einfache Brücke zu Javascript ermöglichen, um dort eigene Erweiterungen vornehmen zu können.

Mit Qlik haben wir einen Anbieter gewählt, der mit HTML5 arbeitet. Für eigene Erweiterungen gibt es eine definierte Javascript-API mit der man eigene QlikView-Extensions (bzw. Qlik-Sense-Extensions) erstellen kann. Zudem gibt es einen Markt für Extensions von Drittanbietern.

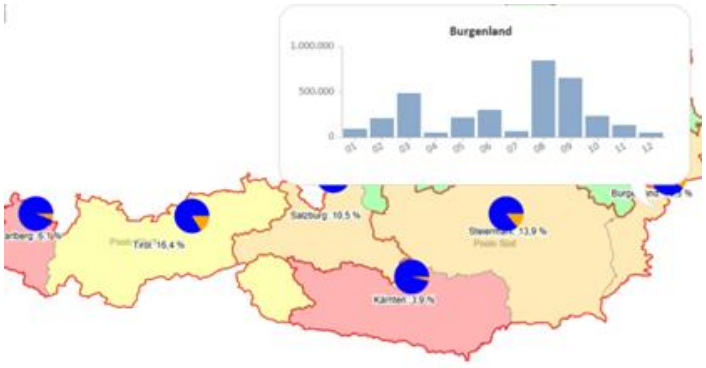


Abbildung 6: Charts in Karten

3.3 NPGeoMap

Für die Visualisierung der Karten, die nicht so speziell sind, dass sie durch eigenentwickelte Extensions eingebunden werden müssen (siehe Abschnitt 3.4), haben wir

uns für die Karten-Extension NPGeoMap entscheiden [NPGeoMap 2015]. Mit diesem Produkt können wir Flächen, Punkte und Strecken visualisieren (siehe Abbildung 6 und Abbildung 7). Ein Hauptgrund für die Nutzung von NPGeoMap ist die performante Verarbeitung von Massendaten. Viele evaluierte Karten-Extensions kamen nicht mit einer Anzahl von Geometrien mit mehr als 5000 Elementen zurecht (oder liefen nicht mehr performant). Das Schienennetz für Europa wird bei uns aber mit ca. 60.000 Strecken abgebildet wird. NPGeoMap kann diese Anzahl überhaupt anzeigen und dazu noch in einer anwenderfreundlichen Geschwindigkeit.

3.4 Eigene Entwicklung Karten-Extension

Eine wichtige geovisualisierte Darstellungsform im Transport und Logistikbereich sind Informationen, die einer Strecke (geometrisch ein Linienzug) zugeordnet sind.

Abbildung 7: Quelle-Ziel-Beziehungen visualisieren

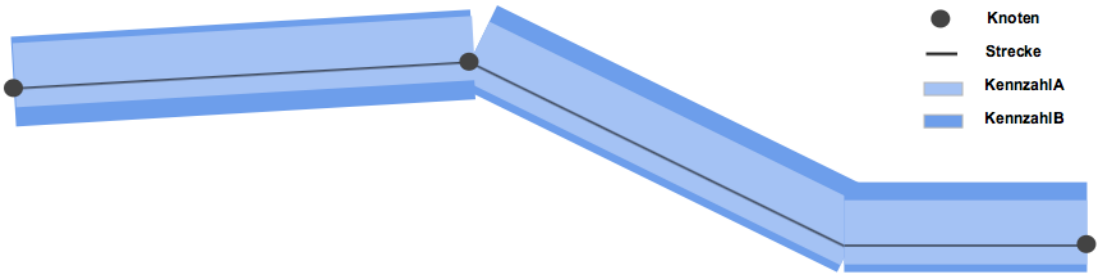


Abbildung 8: Was sind gestapelte Balken auf einer Strecke?

Eine einfache Form ist einen Parameter mittels Farbe zu visualisieren. Eine nur wenig komplexere Form ist die Visualisierung mittels Strichstärke. Für unseren Anwendungsbereich sind diese Formen allerdings nicht ausreichend. An einer Strecke müssen

gleichzeitig mehrere Informationen angezeigt werden können. Ferner wird die Darstellung um ein vielfaches komplizierter, wenn die Kennzahlen spezifisch für die Hin- und Rückrichtung einer Strecke anzuzeigen sind. Ein Beispiel dafür ist eine Staukarte für das Autobahnnetz. Ein Stau existiert selten sowohl in Hin- als auch in Rückrichtung. Eine gängige Visualisierung ist eine asymmetrische Darstellung mit Strichstärken. So wird jeweils Rechts vom Streckenmittel die Kennzahl als langezogenes Balkendiagramm dargestellt. Kennzahlen können dann als gestapeltes Balkendiagramm über die gesamte Strecke angezeigt werden. Sind die Kennzahlen gleichartige Segmente, so wird zudem die Summe aller Kennzahlen sichtbar – beispielsweise „Fahrten Diesel“ und „Fahrten Elektrisch“ zeigen zusammen „Fahrten gesamt“ an (siehe Abbildung 8).

Abbildung 9: Gestapelte Balken auf Strecken

Kein Anbieter konnte uns eine solche Visualisierung anbieten, so dass wir eine eigene Erweiterung für Qlik entwickelt haben. Basierend auf der Javabibliothek Leaflet haben wir dies umgesetzt [Leaflet 2015]. Auch hier ist eine der Anforderungen, mehr als 5000 Elemente anzeigen zu können. Abbildung 9 zeigt das Ergebnis dieser Entwicklung.

4 Fazit

Die Suche nach einem geeigneten BI-Werkzeug mit Geovisualisierung für Massendaten war erfolgreich. Durch den Aufbau des beschriebenen Systems ist ein großer Mehrwert

entstanden. So können zum einen die Analysen interaktiv durchgeführt werden. Dadurch können wir neue Perspektiven auf unseren Daten einnehmen und damit verbunden neue Erkenntnisse gewinnen (Data Discovery). Zum zweiten wird der Anwender durch Visual Analytics unterstützt, was z.B. eine Suche nach Ausreißern durch grafische Unterstützung (Charts und Karten) wesentlich vereinfacht. Zum dritten konnten wir durch den Einsatz von Kartenmodulen die Geovisualisierung umsetzen, die massendatentauglich ist, und damit georeferenzierte Datenanalysen durchführen, die aufgrund der einfachen Handhabung nicht mehr nur von Experten vorzunehmen sind.

5 Literaturverzeichnis

[Benedikt Süßmann 2015]

Benedikt Süßmann: Erweiterung eines Business Intelligence Systems zur kartengestützten Visualisierung und Interaktion mit Massendaten basierend auf Verkehrsströmen. *Hochschule RheinMain, (in progress)* 2015.

[Hahmann und Burghardt 2013]

Hahmann, Stefan und Burghardt, Dirk: How much information is geospatially referenced? Networks and cognition. *International Journal of Geographical Information Science* 27 (6): 1171–89, 2013.

[Sallam u. a. 2014]

Sallam; Rita, L.; Tapadinhas, J.; Parenteau, J.; Yuen, D.; Hostmann, B.: Magic quadrant for business intelligence and analytics platforms. *Gartner RAS core research notes. Gartner, Stamford, CT,* 2014.

[Thomas und Cook 2005]

Thomas, J.J. und Cook, K.A., Hrsg.: *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE CS Press. http://nvac.pnl.gov/agenda.stm, 2005.

[Williams 1987]

Williams, Robert E.: Selling a geographical information system to government policy makers. In *URISA*, 3:150–56, 1987.

[Leaflet 2015]

Webseite: Leaflet — an open-source JavaScript library for interactive maps. http://leafletjs.com/, 2015.

[NPGeoMap 2015]

Webseite: NPGeoMap QlikView Geo-Visualization extension. http://www.npgeomap.com/index.php/de/ , 2015.

[OpenLayers 3, 2015]

Webseite: OpenLayers 3. http://openlayers.org/ , 2015.

[Visual Analytics 2015]

Webseite: What is Visual Analytics?. Zugegriffen Juni 15. http://www.visual-analytics.eu/faq/, 2015

Offenlegung und Verarbeitung von Lärmbelastungsdaten zur Bürgerinformation und Entscheidungsunterstützung am Flughafen Frankfurt

Günter Lanz, Gemeinnützige Umwelthaus GmbH, guenter.lanz@umwelthaus.org
Wassilios Kazakos, Disy Informationssysteme GmbH, kazakos@disy.net

Abstract / Zusammenfassung

In a project of Disy Infomationssysteme GmbH for Gemeinnützige Umwelthaus GmbH Kelsterbach, on one hand, environmental noise data stemming from air traffic at the Frankfurt International Airport, have been published through an interactive Web tool. On the other hand, the combination of GIS and database technologies was used to compare different models for changing the air traffic schedules with respect to the maximum number of people who would benefit from less noise during the night.

Im Rahmen eines Projekts der Disy Informationssysteme GmbH für die Gemeinnützige Umwelthaus GmbH Kelsterbach wurden einerseits Lärmbelastungsdaten aufgrund des Luftverkehrs am Flughafen Frankfurt mithilfe von Disy Cadenza veröffentlicht. Andererseits wurden durch den Einsatz von GIS- und Datenbanktechnologien verschiedene Modelle zur Flugbewegungsverlagerung bzw. -reduktion dahingehend bewertet, dass das Modell mit der größten Anzahl entlasteter Bürger identifiziert werden konnte. Dies stellt die Basis für den Sommerflugplan 2015 dar.

1 Einleitung

Mehr Ruhe für die Anwohner des Frankfurter Flughafens – das ist erklärtes politisches Ziel der Landesregierung in Hessen. Im Auftrag des Umwelt- und Nachbarschaftshaus in Kelsterbach hat die Disy Informationssysteme GmbH fünf Modelle zur unterschiedlichen Nutzung der Start- und Landebahnen mit GIS- und Datenbanktechnologien

analysiert und dabei ermittelt, bei welchem Modell die meisten Betroffenen eine Lärmpause haben.

2 Fakten zum Flughafen Frankfurt

Über 50 Millionen Passagiere und 2 Millionen Tonnen Fracht jährlich, bis zu 1.500 Starts und Landungen am Tag auf vier bis zu 4,6 km langen Bahnen, etwa 80.000 Beschäftigte vor Ort – das ist der Frankfurter Flughafen[27,28]. Das Start- und Landebahnsystem des Flughafens besteht aus vier Bahnen, wobei drei parallel in Ost-West-Richtung ausgerichtet sind und eine in Nord-Süd-Richtung (Abbildung 2).

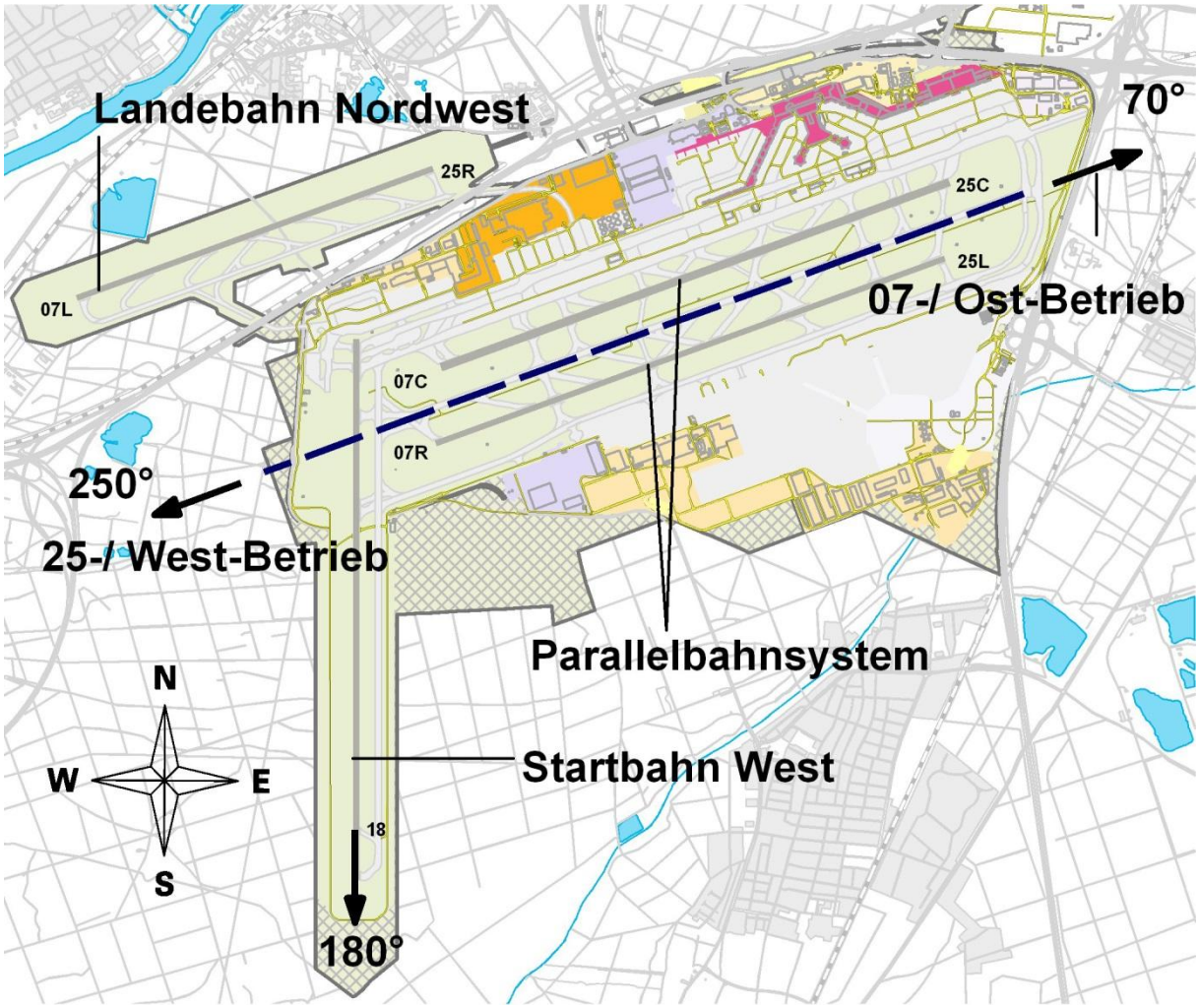


Abbildung 1: Übersicht über die Start- und Landebahnen
am Flughafen Frankfurt

[27] http://www.fraport.de/zahlendatenfakten/

[28] http://www.forum-flughafen-region.de/themen/basiswissen/zahlen-datenfakten-des-flughafens-frankfurt/

Im Jahr 2010 wurden ca. 83 Flugbewegungen pro Stunde abgewickelt, was 464.432 Bewegungen im Jahr bedeutet. Die Flugbewegungen nehmen weiter zu, Flugzeugmix sowie An- und Abflugströme haben sich geändert. Der Ausbau, begonnen mit der neuen Nordwest-Landebahn, ist für eine Kapazität von 701.000 Flugbewegungen ausgelegt – eine weitere Steigerung um ca. 40 %. Neben der enormen wirtschaftlichen Bedeutung, die dieser drittgrößte Flughafen Europas hat, birgt er aber auch viel Konfliktpotenzial, denn der Luftverkehr hat erhebliche Auswirkungen auf die Umwelt und die in der Umgebung lebenden Menschen.

3 Rolle des Umwelt- und Nachbarschaftshaus Kelsterbach

Deshalb hat das Land Hessen gemeinsam mit weiteren Akteuren, den Kommunen und der Luftverkehrswirtschaft, das Forum Flughafen und Region geschaffen. Ein Teil davon ist das Umwelt- und Nachbarschaftshaus (UNH)[29] in Kelsterbach, eine Informationsstelle, die Fakten zum Flugverkehr rund um den Frankfurter Flughafen sammelt, auswertet und der Öffentlichkeit sowie dem Betreiber und den Nutzern des Flughafens zur Verfügung stellt.

Das UNH ist eine Einrichtung des Landes Hessen, welches neben dem Konvent und dem Expertengremium Aktiver Schallschutz die dritte Säule des Forums Flughafen und Region bildet. Entstanden ist das Ganze aus dem Regionalen Dialogforum mit dem Ziel, bei besonders konfliktträchtigen Situationen rund um den Frankfurter Flughafen zu vermitteln. Die Hauptaufgabe liegt darin, den Dialog über die Wechselwirkungen des Flughafens, der Region und einzelner Akteure im Rhein-Main-Gebiet zu moderieren und zu verbessern. Beim UNH laufen Informationen aus unterschiedlichen Bereichen sowie Organisationen zusammen und werden für ein unabhängiges Monitoring zusammengeführt. Laut Satzung ist das UNH verpflichtet, mit diesen Daten neutral und transparent umzugehen.

[29] http://www.forum-flughafen-region.de/ueber-uns/umwelt-und-nachbarschaftshaus/aufgabe/

4 UNH schafft transparente Informationswege

Damit das UNH aus den verschiedenen Datenquellen transparente Informationen und damit Mehrwerte für die Betreiber, Nutzer und Anwohner des Flughafens schaffen kann, benötigte es eine geeignete Software. Deshalb setzt das UNH seit 2013 für den Abruf, die Aufbereitung und Präsentation der Daten die GIS- und Reportingsoftware Cadenza[30] von Disy ein sowie eine zentrale Oracle-Datenbank, in der alle Informationen gesammelt und vorgehalten werden. Unter anderem wird im UNH mit Cadenza eine interaktive Präsentation zahlreicher Umweltdaten in einer Dauerausstellung[31] realisiert, wodurch jeder Bürger flexibel auf die vielfältigen Datensätze zugreifen kann. Abrufbar sind beispielsweise die Fluglärmkonturenkarten. Diese zeigen, wie hoch der Lärm in verschiedenen Zonen rings um den Flughafen tagsüber und nachts war. Neben Daten aus den Jahren 2007 bis 2013 werden auch Prognosen für das Jahr 2020 gegeben. Zusätzlich ist das auf Cadenza Web basierende Online-Portal[32] gestartet, in dem alle interessierten Personen auf interaktiven Karten, Tabellen oder Diagrammen verschiedene Angaben zu Flughafenbetrieb und seinen Auswirkungen finden können.

5 Politische Motivation für die Einführung von Lärmpausen

Aufgrund der hohen Lärmbelästigung in der Flughafenumgebung hat sich die neue hessische Landesregierung u. a. das Ziel gesetzt, das geltende sechsstündige Nachtflugverbot durch die Einführung einer zusätzlichen Lärmpause zu verlängern. Damit möchte man den Betroffenen morgens und abends im Wechsel eine Stunde mehr Ruhe geben und die nächtliche Lärmpause auf sieben Stunden verlängern. Damit die Anwohner eine Chance auf einen ausreichenden und erholsamen Nachtschlaf haben, sollen im Wechsel von 22 bis 5 Uhr oder von 23 bis 6 Uhr in der Umgebung des Frankfurter Flughafens in keinem Schlafzimmer mit gekipptem Fenster mehr als sechs Schallereignisse von mehr als 58 Dezibel (dB(A)) außen auftreten. Dies entspricht

[30] http://www.disy.net/produkte/cadenza.html
[31] http://www.informationszentrumumwelthaus.org/
[32] http://cadenza.umwelthaus.org/pages/home/welcome.xhtml

innen etwa 43 dB (A)[33] und erlaubt einem durchschnittlich empfindlichen Menschen ungestörtes Weiterschlafen.

6 Entwicklung geeigneter Modelle zur Umsetzung der Lärmpause

Von dieser zusätzlichen Lärmpause sollen einerseits möglichst viele Betroffene profitieren, andererseits darf dadurch nicht die Wettbewerbsfähigkeit des Flughafens gefährdet werden. Deshalb hat das hessische Verkehrsministerium zusammen mit der Deutschen Flugsicherung (DFS), dem Flughafenbetreiber Fraport und der Deutschen Lufthansa insgesamt 256 theoretisch denkbare Varianten für die unterschiedliche Nutzung der Start- und Landebahnen entwickelt und zunächst unter sicherheitstechnischen, wirtschaftlichen und betrieblichen Aspekten geprüft und bewertet. Von diesen vielen Varianten, die wechselnde Nutzungen der Start- und Landebahnen in den Randstunden vorsahen und alle zunächst aus reinen Kapazitätsgründen realisierbar gewesen wären, erwiesen sich fünf Modelle[34] als möglich für einen Probebetrieb. Diese fünf Modelle wurden dem UNH vorgelegt um zu prüfen, wie viele Menschen bei den jeweils vorgeschlagenen Bahnnutzungen eine geringere Lärmbelastung, wie viele eine höhere Lärmbelastung erfahren und bei wie vielen sich keine Änderung ergibt.

7 Analyse der Modelle mit GIS und Datenbanktechnologien

Im Auftrag des UNH hat Disy fünf Modelle zur unterschiedlichen Nutzung der Start- und Landebahnen mit GIS- und Datenbanktechnologien analysiert und dabei ermittelt, bei welchem Modell die meisten Betroffenen eine Lärmpause haben. Bedingt durch die Betriebsrichtung Ost bzw. West und die Unterscheidung in Morgen- und Abendrandstunde ergaben sich 20 zu untersuchende Varianten. Dazu hat Disy – unter Einbeziehung von Cadenza sowie der Datenbanktechnologie von Oracle Spatial – Gemeindegrenzen und soziodemographische Daten mit Ergebnissen der Fluglärmberechnung auf Datenbankebene verschnitten und diese mit der Bevölkerungsdichte unterlegt. Diesen datenbankbasierten Lösungsansatz hat Disy schon in vielen Projekten erfolgreich

[33] https://www.hessen.de/sites/default/files/media/hmwvl/15-02-04_charts_laermpausen-pk.pdf
[34] https://www.hessen.de/sites/default/files/media/hmwvl/modelle_laermpausen_uebersicht.pdf

angewendet und im Laufe der Zeit immer weiter ausgebaut. Dabei kann auf Erfahrung aus mehreren großen Infrastrukturprojekten zurückgeblickt werden, in deren Rahmen sowohl die Algorithmen immer weiter verfeinert als auch eine entsprechende hausinterne Datenbankinfrastruktur aufgebaut wurde. Im GIS-Bereich zählt Disy zu den wenigen Firmen weltweit, die die Technologien von Oracle Spatial in diesem Umfang und in dieser Tiefe einsetzen.

Auch im Lärmpausenprojekt hat dieser Ansatz sich bewährt, weil damit große Datenmengen in Verbindung mit komplexen Verschneidungsalgorithmen in kurzer Zeit effektiv verarbeitet werden konnten. Die projektspezifischen Algorithmen greifen dabei auf Funktionen aus der Disy-Standardbibliothek zurück. Für die Berechnungen im Zusammenhang mit dem Projekt Lärmpause wurden in einem ersten Schritt der Ist-Zustand und die berechneten Modellzustände in Bezug auf die Lärmbelastung, die als ESRI-Shapefile zur Verfügung standen, mit Hilfe der Software Talend in eine Oracle Datenbank importiert. Bei Talend handelt es sich um ein ETL-Werkzeug (ETL = Extract, Transform, Load) zum Umwandeln und Importieren von Datenbeständen in verschiedenste Formate. Bei Disy wird derzeit die freie Version Talend Open Studio mit der Erweiterung „Spatial extension for Talend“[35] eingesetzt. Oracle wurde für dieses Projekt in der Version 11g Release 2 verwendet. In der Datenbank wurde jeweils die Basis (= IST-Zustand) mit den jeweiligen Modellen verschnitten. Daraus ergaben sich jeweils drei Bereiche (Abbildung 2).

[35] https://www.talend.com/, http://talend-spatial.github.io

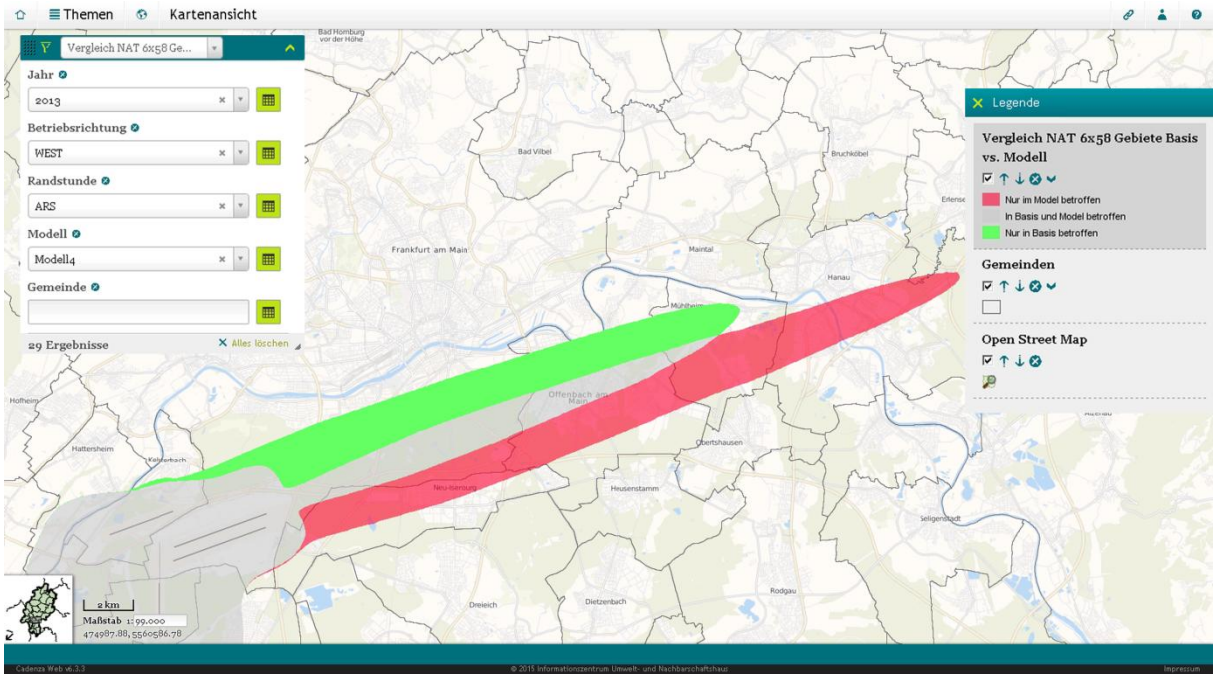


Abbildung 2: Hier wurde für die Abendrandstunde der Maximalpegel Basis gegen das Modell verschnitten

Zum Einen gab es einen Bereich, der in der Basis und im Modell betroffen ist. Für diesen Bereich gab es weder im IST-Zustand noch im simulierten Modell eine Verbesserung. Ein weiteres Gebiet ist nur in der Basis betroffen und würde somit vom Modell profitieren. Im Gegensatz dazu gibt es einen weiteren Bereich, der nur im Modell betroffen ist, so dass diese Variante für die darin wohnenden Einwohner eine Verschlechterung bedeutet.

Im nächsten Schritt, der auf Grund der großen Datenmenge nur sinnvoll in der Datenbank durchgeführt werden konnte, wurden die berechneten Flächen mit den Bevölkerungsdaten, die in einem 125m-Raster vorlagen, verschnitten (Abbildung 3).

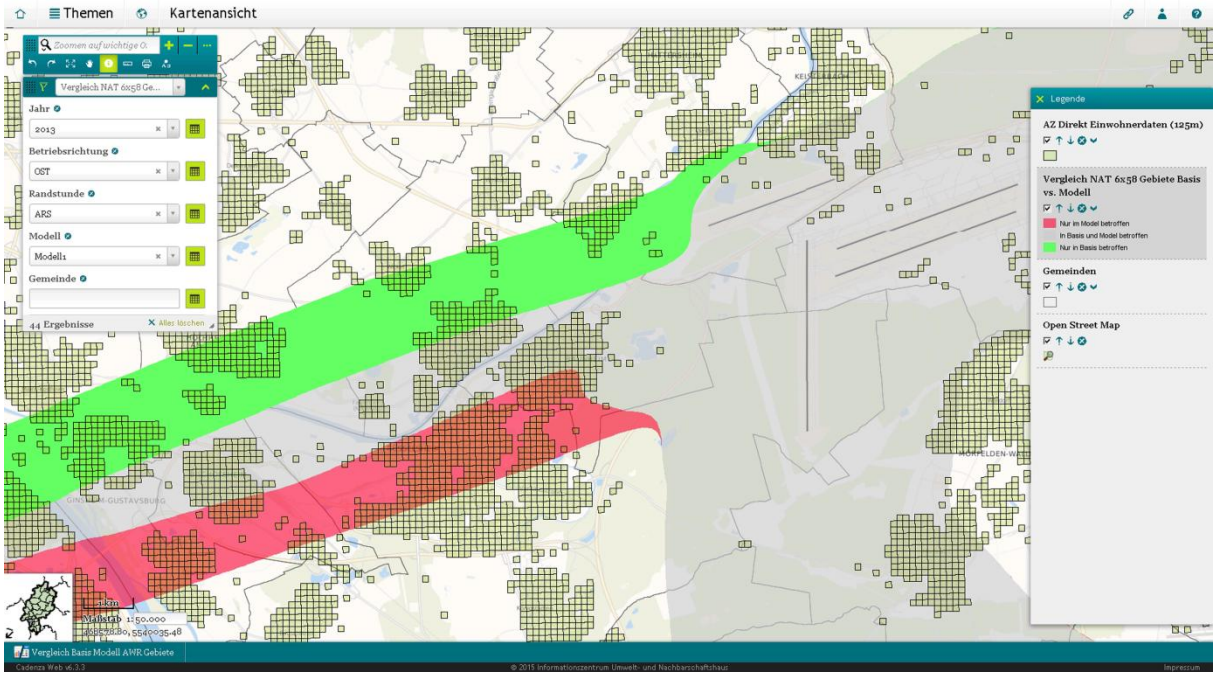


Abbildung 3: Hier wurde für die Abendrandstunde der Maximalpegel Basis gegen das Modell verschnitten und mit einem Bevölkerungsraster unterlegt

Trotz der zuvor beschriebenen Optimierung und der Durchführung der Berechnung in einer Datenbank in einer leistungsfähigen Umgebung mit einem hochperformanten Storage-Backend dauerte dieser automatisierte Berechnungsprozess mehrere Tage. Ein weiterer Vorteil dieser Vorgehensweise war die automatisierte Backup-Strategie mit Snapshots, die auf solchen zentralen Datenbankservern bei Disy durchgeführt wird. Damit war es jederzeit möglich, auf vorherige Bearbeitungsstände zurückzugehen. Eine manuelle Bearbeitung mit einer klassischen GIS-Software hätte jedoch mehrere Wochen bis Monate Zeit in Anspruch genommen. Des Weiteren konnte mit diesem Ansatz die Berechnung je nach Fragestellung angepasst werden, so dass sehr schnell neue Daten für den Entscheidungsprozesse der Politik zur Verfügung gestellt werden konnten. Außerdem war es durch die Verwendung eines einheitlichen Datenmodells jederzeit möglich, neue Daten zu ergänzen und diese bei den weiteren Auswertungen zu berücksichtigen.

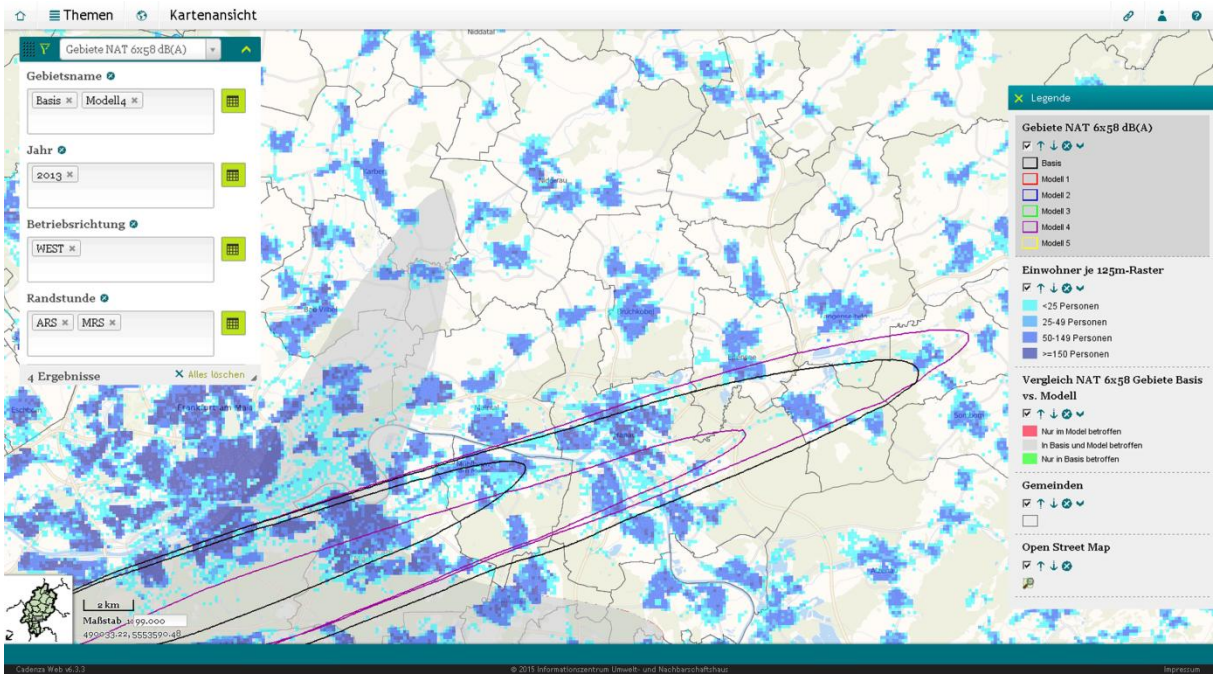


Abbildung 4: Für die Abend- und Morgenrandstunde wurde der Maximalpegel Basis gegen das Modell überlagert; der Kartenausschnitt ist mit einem klassifizierten Bevölkerungsraster unterlegt

Nach dieser Verschneidung waren nicht nur Aussagen über die Größe und Lage der betroffenen Gebiete, sondern auch über die Anzahl der Menschen möglich, die in diesen betroffenen Gebieten leben, weil mithilfe der klassifizierten Bevölkerungsdichte analysiert wurde (Abbildung 4).

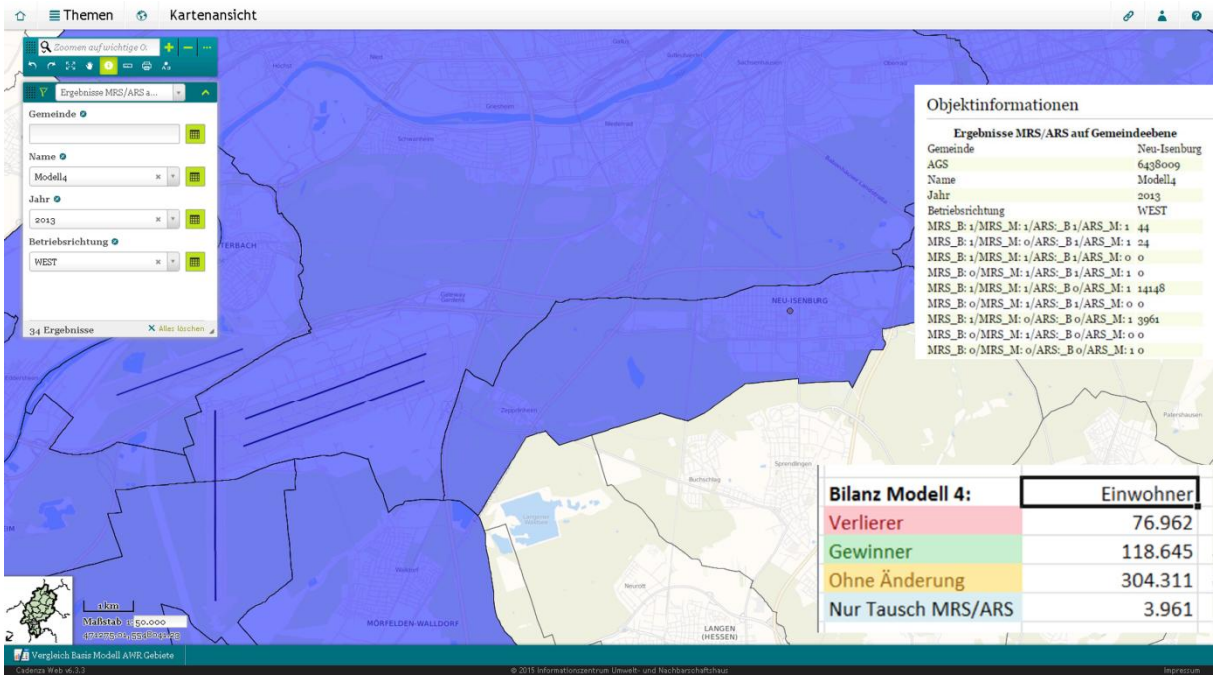
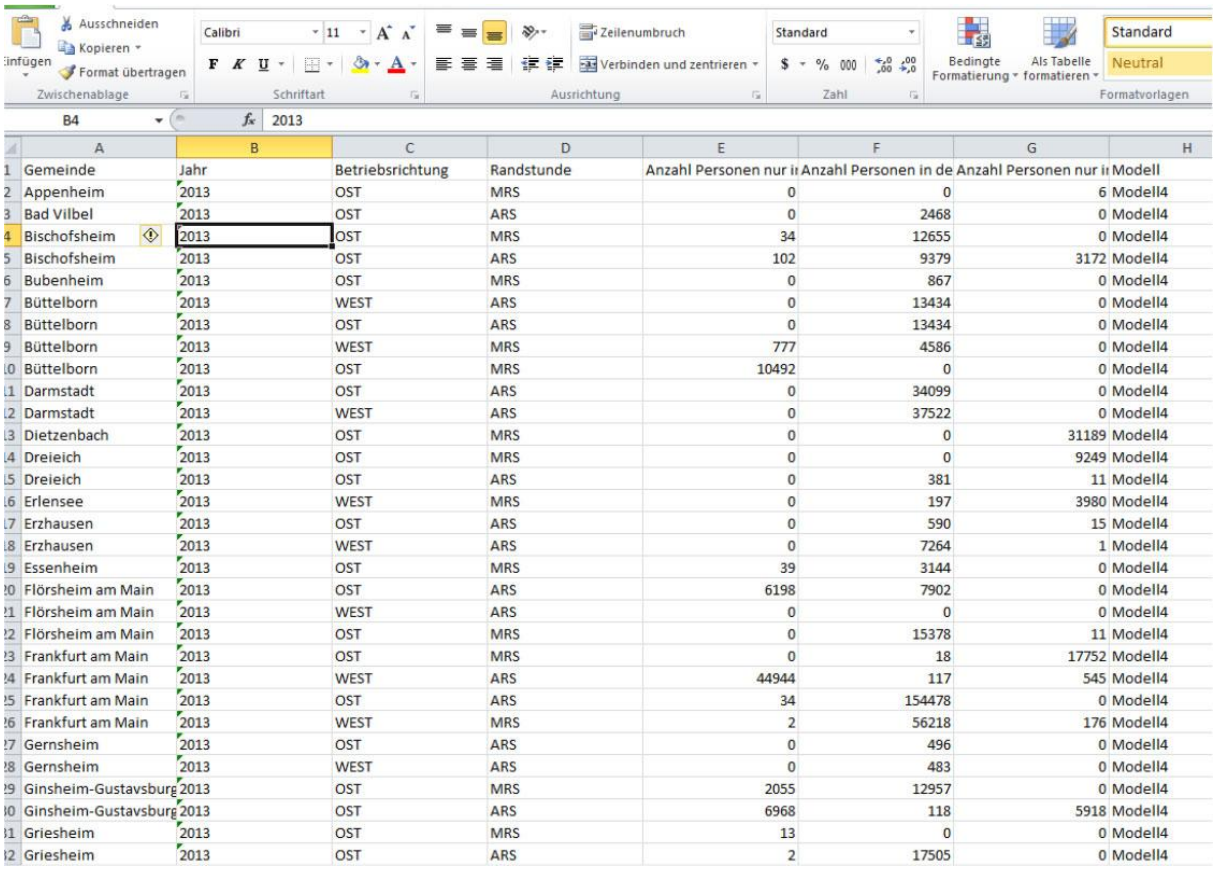


Abbildung 5: Durch die eindeutige Verschneidung auf Datenbank-Ebene konnten im Vergleich die Gewinner und Verlierer ermittelt werden

Da bei den Verschneidungen auch die Gemeindegrenzen berücksichtigt wurden, können diverse Aggregationsebenen über Cadenza Web betrachtet werden (Abbildung 5).

So lassen sich die Anzahl der Gewinner und Verlierer für jedes Modell bilanzieren und insgesamt oder auf Gemeindeebene betrachten.

Durch die integrierte Auswertung der Sach- und Geodaten mit Cadenza konnten die Ergebnisse nicht nur kartographisch aufbereitet und analysiert werden, sondern es konnte am Ende des Analyseprozesses auch in einer Tabelle kommunenscharf bilanziert werden, wie viele Menschen je Modellvariante entlastet, belastet oder nicht betroffen sind (Abbildung 6).



	A	B	C	D	E	F	G	H
1	Gemeinde	Jahr	Betriebsrichtung	Randstunde	Anzahl Personen nur i	Anzahl Personen in de	Anzahl Personen nur i	Modell
2	Appenheim	2013	OST	MRS	0	0	6	Modell4
3	Bad Vilbel	2013	OST	ARS	0	2468	0	Modell4
4	Bischofsheim	2013	OST	MRS	34	12655	0	Modell4
5	Bischofsheim	2013	OST	ARS	102	9379	3172	Modell4
6	Bubenheim	2013	OST	MRS	0	867	0	Modell4
7	Büttelborn	2013	WEST	ARS	0	13434	0	Modell4
8	Büttelborn	2013	OST	ARS	0	13434	0	Modell4
9	Büttelborn	2013	WEST	MRS	777	4586	0	Modell4
10	Büttelborn	2013	OST	MRS	10492	0	0	Modell4
11	Darmstadt	2013	OST	ARS	0	34099	0	Modell4
12	Darmstadt	2013	WEST	ARS	0	37522	0	Modell4
13	Dietzenbach	2013	OST	MRS	0	0	31189	Modell4
14	Dreieich	2013	OST	MRS	0	0	9249	Modell4
15	Dreieich	2013	OST	ARS	0	381	11	Modell4
16	Erlensee	2013	WEST	MRS	0	197	3980	Modell4
17	Erzhausen	2013	OST	ARS	0	590	15	Modell4
18	Erzhausen	2013	WEST	ARS	0	7264	1	Modell4
19	Essenheim	2013	OST	MRS	39	3144	0	Modell4
20	Flörsheim am Main	2013	OST	ARS	6198	7902	0	Modell4
21	Flörsheim am Main	2013	WEST	ARS	0	0	0	Modell4
22	Flörsheim am Main	2013	OST	MRS	0	15378	11	Modell4
23	Frankfurt am Main	2013	OST	MRS	0	18	17752	Modell4
24	Frankfurt am Main	2013	WEST	ARS	44944	117	545	Modell4
25	Frankfurt am Main	2013	OST	ARS	34	154478	0	Modell4
26	Frankfurt am Main	2013	WEST	MRS	2	56218	176	Modell4
27	Gernsheim	2013	OST	ARS	0	496	0	Modell4
28	Gernsheim	2013	WEST	ARS	0	483	0	Modell4
29	Ginsheim-Gustavsburg	2013	OST	MRS	2055	12957	0	Modell4
30	Ginsheim-Gustavsburg	2013	OST	ARS	6968	118	5918	Modell4
31	Griesheim	2013	OST	MRS	13	0	0	Modell4
32	Griesheim	2013	OST	ARS	2	17505	0	Modell4

Abbildung 6: Tabellarische Bilanz der ermittelten Gewinner und Verlierer; durch die Betriebsrichtung Ost bzw. West und die Randstunde am Morgen bzw. am Abend ergaben sich für die 5 Modelle insgesamt 20 Varianten für die Untersuchung

Ein weiterer, nicht zu unterschätzender Vorteil war die grafische Ergebnisdarstellung mittels in Cadenza erzeugter Verschneidungskarten. Erst sie machen es möglich, die tabellarischen Ergebnisse optisch nachzuvollziehen und zu verstehen.

8 Das Ergebnis der Analyse

Die von Disy in enger Abstimmung mit dem UNH durchgeführte Analyse ergab, dass beim Modell 4 die meisten Betroffenen entlastet werden (Abbildung 7).

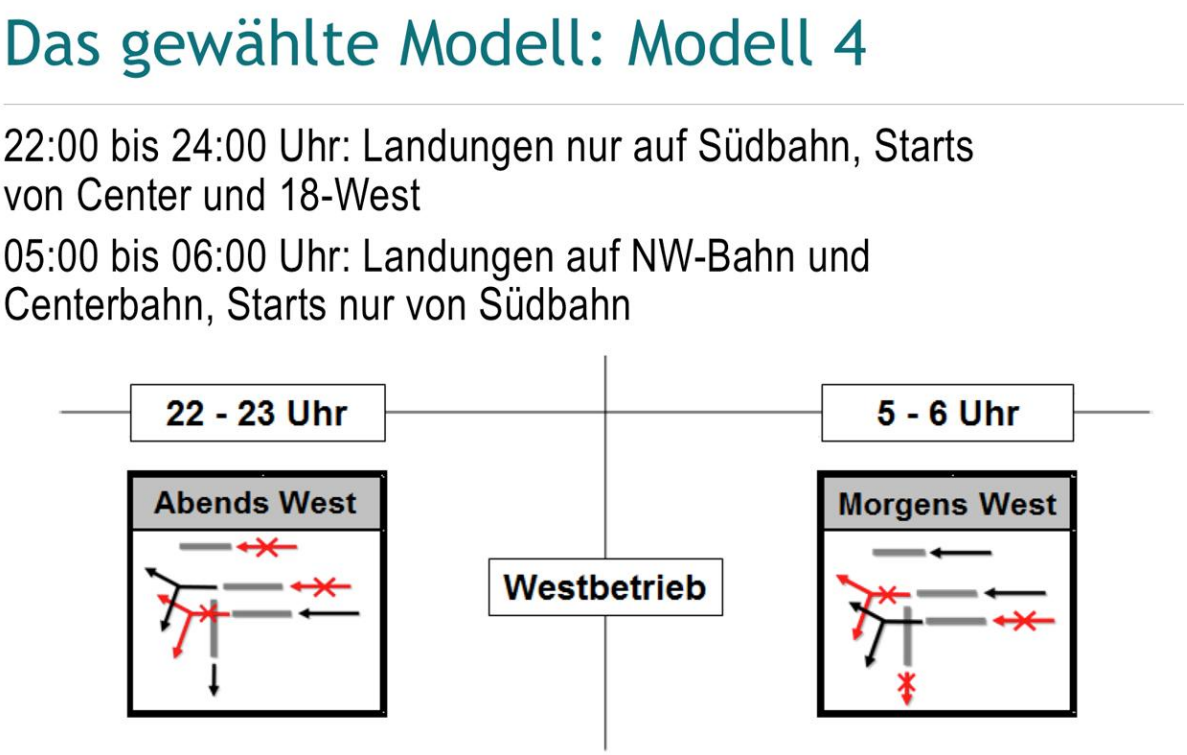


Abbildung 7: Geplante Nutzung der Start- und Landebahnen beim Lärmpausen Modell 4

Dieses Modell sieht vor, dass bei Westwind, der an drei von vier Tagen weht, während der Lärmpause einzelne Bahnen vermehrt, andere dafür gar nicht genutzt werden. Dadurch werden im Rhein-Main-Gebiet 105.000 Menschen entlastet und 65.000 Menschen belastet. Das ist ein Nettogewinn von 40.000 Menschen, die nun nach der Definitionen eine Lärmpause erhalten (Abbildung 8).

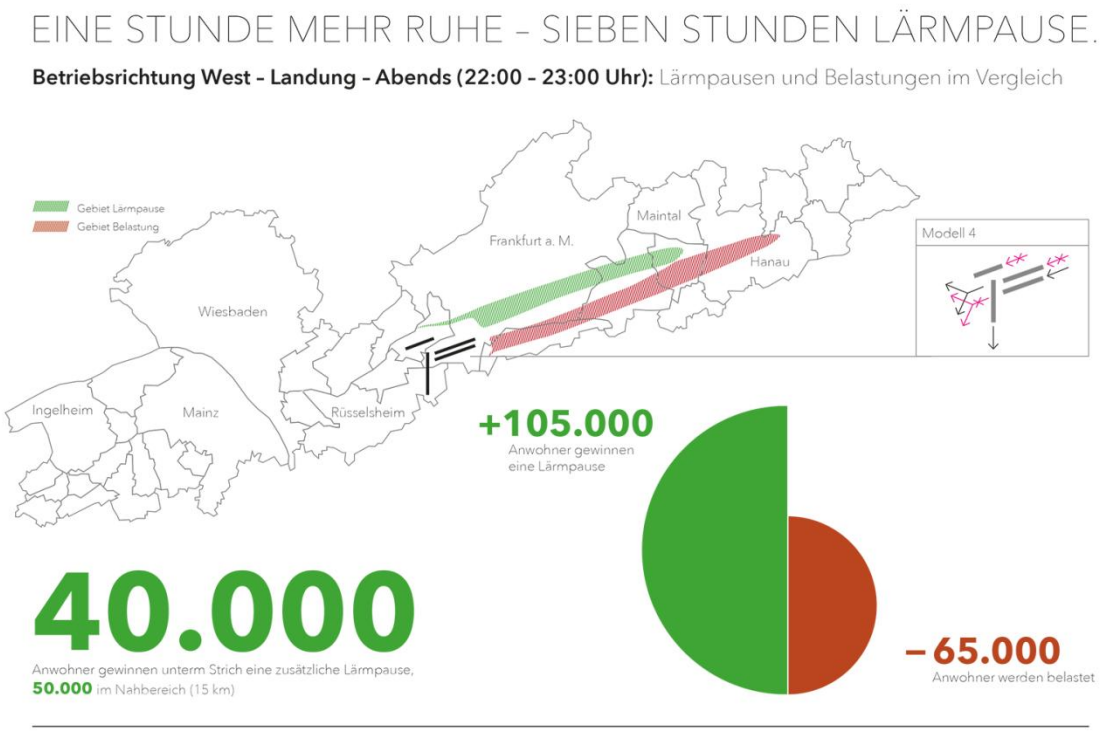


Abbildung 8: Modell 4 – Anzahl der Betroffenen, die be- und entlastet werden

Durch die Analyse dieser komplexen Fragestellung mit GIS- und Datenbanktechnologien konnte am Ende des Analyseprozesses kommunenscharf bilanziert werden, wie viele Menschen entlastet, belastet oder nicht betroffen sind. Durch die Ergebnisdarstellung mittels in Cadenza erzeugten Verschneidungskarten. Erst sie machen die tabellarischen Ergebnisse optisch nachvollziehbar und damit verständlich für die politischen Entscheidungsträger. Auf Grundlage dieser Ergebnisse haben sie entschieden, dass mit Beginn des Sommerflugplanes am 23. April 2015 die Lärmpause für den einjährigen Probebetrieb eingeführt wird.

Systemkonzept für die Einbindung freiwilliger Helfer zur verbesserten Krisenbewältigung

Michael Jendreck, michael.jendreck@fokus.fraunhofer.de
Stefan Pfennigschmidt, stefan.pfennigschmidt@fokus.fraunhofer.de
Markus Hardt, markus.hardt@fokus.fraunhofer.de
Dr. Ulrich Meissen, ulrich.meissen@fokus.fraunbhofer.de
Fraunhofer FOKUS
Prof. Dr. Frank Fuchs-Kittowski, frank.fuchs-kittowski@htw-berlin.de
Fraunhofer FOKUS & HTW Berlin

Abstract

Crises and emergencies require a rapid support from aid helpers / emergency assistants. The ENSURE project is developing concepts and applications for a stronger citizen involvement for civil protection to guarantee a sufficient amount of volunteer emergency assistants in the future. One of the main objectives of this project is the development of an innovative concept for a technical system that can generally be applied for the recruitment, administration, activation, and coordination of emergency assistants and is summarized in this paper.

1 Einführung

Krisen und Katastrophen erfordern den schnellen Einsatz von Helfern. Um auch zukünftig eine ausreichende Zahl an freiwilligen Helfern sicherstellen zu können, erarbeitet das Projekt ENSURE Konzepte für eine stärkere Einbeziehung der Bevölkerung in den Katastrophenschutz. Vor allem Bürgerinnen und Bürger, die aufgrund ihrer professionellen Kenntnisse, z.B. als Hausmeister, Sicherheitsbeauftragte in

Unternehmen oder Ersthelfer, die Einsatzkräfte spezifisch unterstützen bzw. Soforthilfe einleiten können, stehen hierbei im Fokus.

Die Frage nach der technischen Umsetzung dieser Einbeziehung ist eine der zentralen innerhalb des ENSURE-Projekts und bildet die Grundlage für den vorliegenden Beitrag. Ziel ist es, ein innovatives und generell anwendbares Konzept für ein technisches System zu entwickeln, das die Rekrutierung, Verwaltung, Aktivierung und Koordinierung von Helfern im urbanen Raum u.a. bei Großschadenslagen unterstützt und dabei den besonderen Anforderungen hinsichtlich Performanz, Skalierbarkeit, Datenschutz/Datensicherheit etc. genügt.

Der Beitrag gliedert sich dabei in vier Abschnitte. Zunächst werden zentrale Probleme gegenwärtiger Ansätze zur Einbindung bürgerlichen Engagements analysiert und Anforderungen an eine geeignete technische Unterstützung daraus abgeleitet. Die wichtigsten Funktionen einer solchen Unterstützung werden im Anschluss dargestellt, um dann auf die Architektur eines konkreten Systems einzugehen und dessen Kernkomponenten zu beschreiben. Zum Abschluss wird ein Ausblick auf zukünftige Entwicklungen gegeben.

2 Probleme des bürgerlichen Engagements im Katastrophenschutz

Das bürgerliche Engagement hat sich mit der rasanten Verbreitung der sozialen Medien neu erfunden. Sogenannte ungebundene Helfer spielen eine immer größere Rolle bei der Bekämpfung von (Groß-) Schadensereignissen. Beim Hochwasser 2013 haben sich diese Helfer meist der sozialen Medien zur Absprache und Organisation bedient, um so die Betroffenen zu unterstützen. Gerade Facebook und Twitter mit Unterstützung anderer Dienste wie Google Maps waren Hauptmedien dieser wertvollen Helferleistung (Kaufhold & Reuter, 2014). Neben der Vielzahl von positiven Beiträgen entstanden jedoch gerade bei der Aktivierung und Koordinierung der ungebundenen Helfer Probleme.

So kann die schier unüberschaubare Menge der Hilfegesuche auf diesen Plattformen zu einer Unschlüssigkeit führen. Freiwillige Helfer irren regelrecht von einem Einsatzort

zum nächsten, auf der Suche nach einem noch nicht überlaufenen oder noch aktiven Einsatzort[36]. Da die Wahl des Einsatzortes meist keinerlei festen Kriterien folgt, sondern entweder vom Zufall oder den persönlichen Präferenzen der Helfer abhängig ist, kam es häufig zu einem lokalen Überangebot von Helfern, während diese anderenorts dringend benötigt wurden[37]. Gerade in diesen Situationen kann die Stimmung schnell in Frustration bei allen Beteiligten umschlagen (Kircher, 2014). Die Verbreitung von Falschinformationen (Quednau & Weber, 2014) trägt das ihrige dazu bei. Problematisch stellt sich auch die Arbeitssicherheit bzw. der Arbeitsschutz der Helfer dar (Kaufhold & Reuter, 2014). Ungebundene Helfer, die keine Anleitung bzw. Belehrung durch die Einsatzkräfte erhalten, gefährden sich und andere Helfer. So wurden beispielsweise beim Elbehochwasser 2013 Dämme errichtet, die aufgrund mangelhafter Konstruktion später in sich zusammenfielen (Deutsches Rotes Kreuz e. V., 2014). Ein massives Problem für die Einsatzplanung vor Ort ist die offensichtliche Unplanbarkeit der Helferaktivitäten. Aufgrund der vielen Aufrufe mit unterschiedlichen Einsatzorten in den sozialen Netzwerken (private und/oder öffentliche Kanäle) ist es schwierig vorherzusagen, wo und vor allem mit wie vielen freiwilligen Helfern (sicher) geplant werden kann (Deutsches Rotes Kreuz e. V., 2014). Weitere Problemstellungen, die sich nicht nur auf einen länger geplanten Einsatz beziehen, sondern eher einen ad-hoc-Charakter aufweisen, sind teils unzureichende Ortskenntnisse oder fehlende Schlüsselgewalt der Einsatzkräfte vor Ort.

Um diesen Problemen zu begegnen, sind eine Vielzahl von Anforderungen vom System zu erfüllen. So müssen Mechanismen zur Registrierung und Verwaltung der Helfer bereitgestellt werden. Weiterhin sind Funktionalitäten bezüglich der zielgerichteten Alarmierung sowie der initialen Koordinierung zu realisieren.

[36] Vgl. Kommentar von Compufreak: 06/2013, http://www.socialmediaevolution.de/2013/06/03/die-personliche-flut-gedanken-zur-hochwasserkatastrophe-im-zeitalter-von-social-media/, Stand 17. Juni 2015

[37] Vgl. Karin Etzrodt: 06/2013, https://etzrodt.wordpress.com/2013/06/06/hochwasser-in-dresden-2013-und-die-rolle-sozialer-medien/, Stand: 23.06.2015

3 Konzeption des ENSURE-Systems

In diesem Kapitel wird die technische Konzeption des ENSURE-Systems vorgestellt, welches die Rekrutierung, Verwaltung, Aktivierung und Koordinierung von Helfern im urbanen Raum bei Großschadenslagen unterstützen soll. Hierfür werden aus Nutzersicht die wichtigsten Funktionen des Systems und deren prototypische Umsetzung beschrieben.

Folgende zentrale Funktionen sind für eine effiziente Rekrutierung, Verwaltung, Aktivierung und Koordinierung von Helfern im urbanen Raum bei Großschadenslagen erforderlich und im ENSURE-System vorgesehen:

- Registrierung der freiwilligen Helfer
- Profilierung der Helfer
- Alarmierung der Helfer (per Redaktionssystem)
- Aktivierung der Helfer (per App)

Aus Nutzersicht verteilt das ENSURE-System Hilfegesuche (Anfragen oder Alarmierungen) im Falle einer Gefahren- und/oder Schadenslage. Auf Endnutzerseite (freiwillige Helfer) werden diese Gesuche per Smartphone-App zugestellt. Prinzipiell handelt es sich bei diesem Vorgehen um einen Abonnement-basierten Ansatz, bei dem sich der Nutzer bereit erklärt, mittels Anfragen oder Alarmierungen, die seinen aktuellen Aufenthaltsort betreffen, aktiviert zu werden.

3.1 Registrierung der freiwilligen Helfer

Um eine effiziente und ortsgebundene Alarmierung zu gewährleisten, ist eine Registrierung der freiwilligen Mithelfer notwendig. Diese Registrierung erfolgt per Smartphone-App. Beim erstmaligen Öffnen der App wird dem Nutzer ein Projekt-Guide präsentiert, um ihn zum „Mitmachen" zu motivieren (siehe Abbildung).

Abbildung 1: Projekt-Guide (iOS-Gerät) – Fotos © Berliner Feuerwehr, Fotografin: C. Frickemaier

Entscheidet sich der Nutzer zur Teilnahme, ist das aus technischer Sicht gleichbedeutend mit einer Registrierung und einer impliziten Profilerstellung.

3.2 Profilierung der Helfer

Eines der grundlegenden Prinzipien im Datenschutz ist die Datenvermeidung[38]. Das ENSURE-System setzt dieses Prinzip – so weit wie möglich – um. So sind lediglich folgende Informationen im (Basis)-Profil eines Mithelfers (in anonymisierter Form) hinterlegt:

- **Aktueller Aufenthaltsort:** Da ein Hilfegesuch stets ortsgebunden ausgelöst wird, ist es zwingend notwendig, den ungefähren Aufenthaltsort aller Mithelfer im System zu halten. Aus Datenschutzgründen und im Sinne der informationellen Selbstbestimmung muss sich ein Mithelfer nach der Registrierung noch einmal aktiv dazu bereit erklären, bezüglich möglicher Anfragen bzw. Alarmierungen benachrichtigt zu werden und somit akzeptieren, dass der individuelle Aufenthaltsort dem System stets bekannt ist.
- **Fitnesszustand und soziale Kompetenz:** Um im Falle einer Gefahren- bzw. Schadenslage effektiv Mithelfer zu aktivieren (Filterung anhand von Eigenschaften), sind Angaben (subjektive Einschätzungen) bezüglich der individuellen

[38] Vgl. datenschutzbeauftragter-info.de: 05/2014, https://www.datenschutzbeauftragter-info.de/begriff-und-geschichte-des-datenschutzes/, Stand: 21.06.2015

körperliche Fitness und der sozialen Kompetenz im Basisprofil eines Mithelfers hinterlegt. Die Einschätzung erfolgt durch den Nutzer mittels Beantwortung weniger Fragen während der App-Einrichtung.

Neben dem Basisprofil ermöglicht es das System, Mithelfern sogenannte Profilerweiterungen zuzuordnen. Diese Profilerweiterungen können über verschiedene Mechanismen eingespielt werden. So können sowohl durch Dritte verifizierte Qualifikationen (u.a. Ersthelfer) als auch freiwillige Angaben des Nutzers (u.a. Führerscheinklasse, technisches Knowhow) dem System zusätzlich übergeben werden.

3.3 Alarmierung der Helfer (per Redaktionssystem)

Die vom System verschickten Hilfegesuche werden von den Einsatz-Leitstellen mittels eines webbasierten Redaktionssystems erstellt (siehe Abbildung 2). Die Filterung und Alarmierung der Mithelfer erfolgt in einem zweistufigen Verfahren:

- **Filterung:** Zunächst wird unter Berücksichtigung des Einsatzortes, der Einsatzstartzeit sowie der benötigten Anzahl an Helfern und deren Kompetenzen der Aktivierungsradius festgelegt.
- **Alarmierung:** Nachfolgend können dann weitere Angaben (Aufgaben, Einsatzdauer, Hinweise etc.) zum Einsatz aufgenommen und letztendlich den Mithelfern, die sich in der Aktivierungszone befinden, zugesandt werden (Alarmierung).

Es können im ENSURE-System zwei Formen der Alarmierung durchgeführt werden:

- Alarmierung mit Vorwarnzeit
- Ad-hoc-Alarmierung

3.3.1 Alarmierung mit Vorwarnzeit

Ist die Vorwarnzeit ausreichend lang, können die potentiellen Helfer zunächst einmal vorinformiert und somit angefragt werden, um ihre Bereitschaft zur Teilnahme abzuklären. Dieses Vorgehen ist vor allem zur besseren Planung des Einsatzes dienlich. Weiterhin bietet eine Einsatzanfrage die Möglichkeit, spezielle Kompetenzen, falls für den Einsatz nötig, mit einem Fragebogen zu ermitteln. Die Ergebnisse des Fragebogens

werden als Profilerweiterungen im System hinterlegt und dienen der Mithelfersuche dann als Filterangabe.

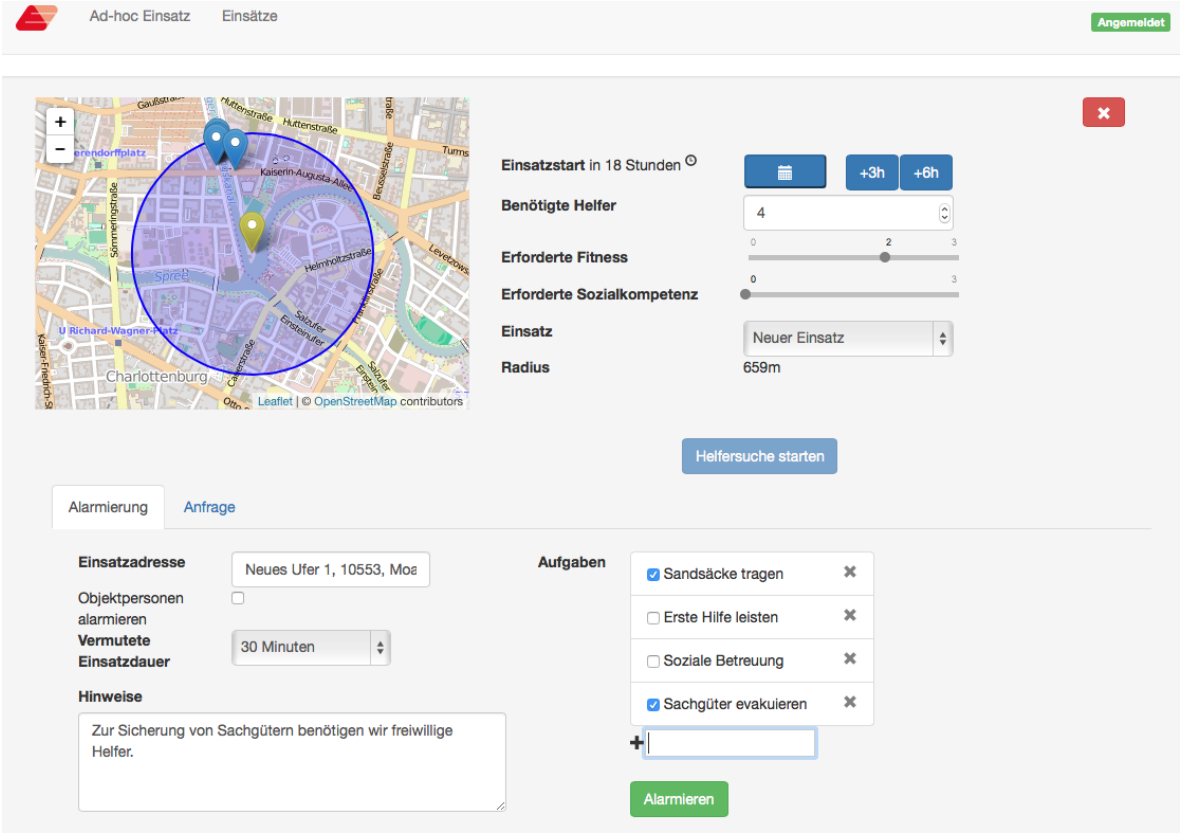


Abbildung 2: Redaktionssystem - Maske zur Alarmierung der Mithelfer mit Vorwarnzeit

Weitere Funktionalitäten, die das Redaktionssystem zur Verfügung stellt, sind:

- Das Versenden von Neuigkeiten an alle Nutzer,
- Das Versenden zusätzlicher Informationen zu einem Einsatz, die allen Mithelfern, die am jeweiligen Einsatz teilnehmen, zugestellt werden,
- Die (Mit-)Alarmierung von objektbezogenen Personen (u.a. Hausmeister, Pförtner), um der Problematik Schlüsselgewalt und fehlende Ortskenntnis entgegenzuwirken,
- Eine Detailansicht zu laufenden Einsätzen (u.a. Rückmeldungen der Mithelfer).

3.3.2 Ad-hoc-Alarmierung

Neben den Einsätzen, die eine Planungsphase voranstellen, wird systemseitig eine ad-hoc-Alarmierung ermöglicht. Die Idee besteht darin, dass gerade bei medizinischen

Einsätzen/Notfällen Mithelfer in unmittelbarer Umgebung bereits in der Isolationsphase am Einsatzort eintreffen und Erste Hilfe leisten können. Um nicht unnötig Zeit beim Ausfüllen der (zwar vereinfachten) ad-hoc-Alarmierungsmaske zu verlieren, müssen lediglich Einsatzort und Einsatzcode angegeben werden. Ist dem Einsatzort eine Adresse zugewiesen, wird diese automatisch gesetzt. Im Freitextfeld können optional weitere Einsatzdetails mitgeteilt werden. Da es sich um ein webbasiertes Redaktionssystem handelt, können mit Hilfe von URL-Parametern alle Formularfelder ausgefüllt werden. Vorgeschaltete Fachverfahren können so bereits vorliegende Informationen per URL-Link dem Redaktionssystem übergeben.

3.4 Aktivierung der Helfer (per App)

Zur Aktivierung erhält jeder ausgewählte Mithelfer eine Benachrichtigung per Push-Notification auf dem Smartphone. Dem Nutzer werden in der App selbst sämtliche Informationen zum Einsatz dargestellt, woraufhin er dann situationsbezogen entscheiden kann, ob er den Einsatz (Alarmierung/Anfrage) annimmt oder ablehnt. Weiterhin ist im Falle einer Anfrage Feedback zu nachgefragten Kompetenzen möglich.

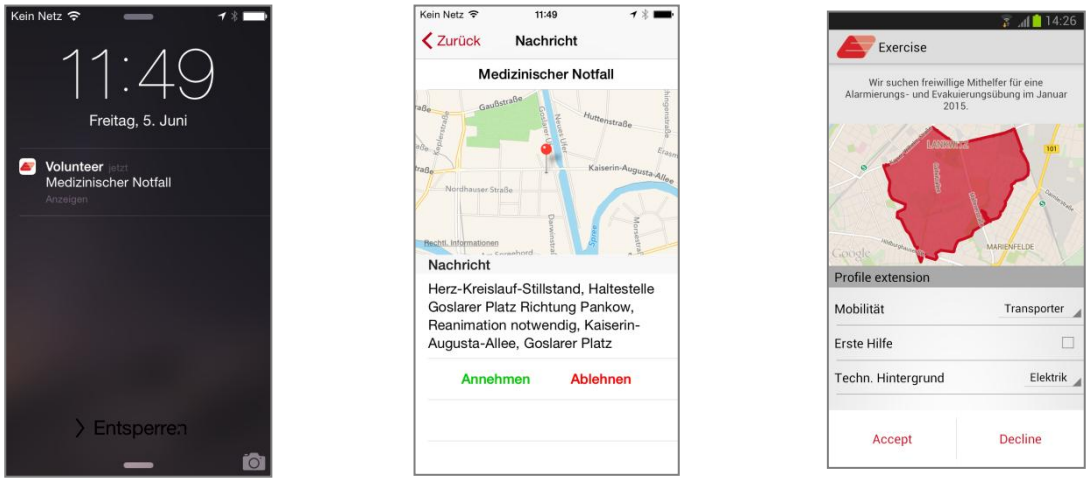


Abbildung 3: Screenshots einer eingehenden Alarmierung (iOS), einer Alarmierung (iOS) und einer Anfrage (Android)

4 Architektur des Mithelfer-Systems

In diesem Kapitel wird die Architektur des ENSURE-Systems dargestellt (siehe Abbildung 4). Die Komponenten des Helfersystems können in sieben logische Prozessbausteine unterteilt werden.

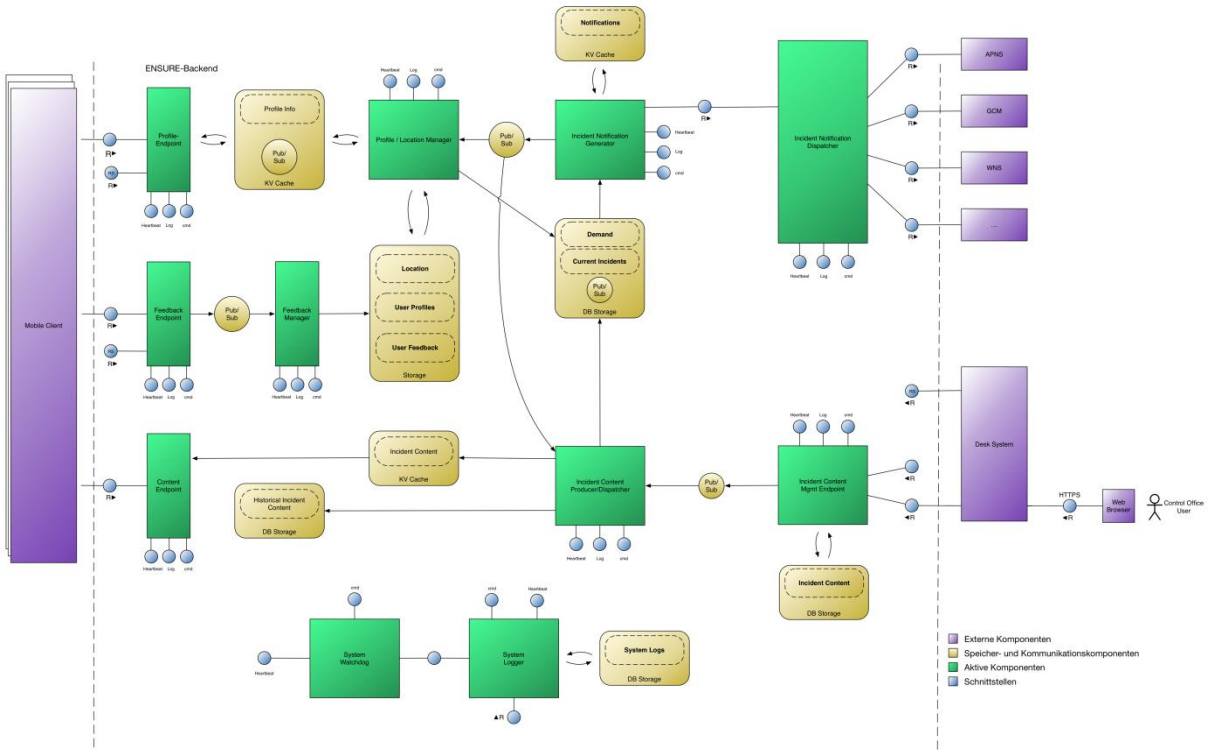


Abbildung 4: Mithelfer-System (Redaktionssystem, App und System-Backend)

4.1 Entgegennahme und Verwaltung von Ereignissen

Der Incident Content Management Endpoint ist für die Verwaltung eingehender Ereignisse (Alarmierungen bzw. Anfragen) verantwortlich. Er nimmt über unterschiedliche externe Schnittstellen Ereignismeldungen (Rohereignisse) entgegen und speichert diese persistent im Incident Content Storage. Neue oder aktualisierte Ereignisse werden an den Incident Content Producer/Dispatcher und somit an das Verteilsystem weitergeleitet.

4.2 Verteilsystem

Der *Incident Content Producer/Dispatcher* hat zwei wesentliche Aufgaben. Zum einen bereitet er die Rohdaten der eingehenden Ereignisse auf, zum anderen ist er für die Verteilung im System zuständig.

Die Aufbereitung erfolgt in Abhängigkeit zur weiteren Verteilung. So speichert er die für den Versand der Benachrichtigungen notwendigen Daten im *Demand Incident Storage* und informiert den *Incident Notification Generator* über das neue bzw. aktualisierte Ereignis. Weiterhin erzeugt er die für den Endnutzer benötigten Daten zur textuellen und grafischen Aufbereitung auf den Endgeräten und stellt diese zum Abruf über den *Incident Content Cache* bereit.

Der *Incident Content Cache* dient als Zwischenspeicher für auszuliefernde Inhalte (Ereignisse (Alarmierungen, Anfragen), Zusatzinformationen).

Der *Demand/Incident Storage* bildet Abonnements der System-Nutzer und Ereignismeldungen in einem performanten Laufzeitmodell ab, das ein effizientes Matching zwischen beiden Informationstypen ermöglicht.

4.3 Abrufen von Inhalten

Der *Content Endpoint* stellt für den Nutzer die Schnittstelle zur Verfügung, über die Inhalte abgefragt werden können. Hierzu gehören Abfragen von aufbereiteten Ereignissen und ereignisrelevanten Zusatzinformationen, die der *Content Endpoint* dem *Incident Content Cache* entnimmt. Der *Content Endpoint* ist auf maximale Skalierbarkeit ausgelegt und verfügt über eine dynamische Lastverteilung.

4.4 Verwalten von Personen- und Geräteinformationen

Der *Profile Endpoint* stellt die Schnittstelle des Systems dar, über die der Nutzer profilbezogene Informationen abfragen, hinzufügen oder ändern kann. Der *Profile Endpoint* verfügt – wie auch der *Content Endpoint* – über eine dynamische Lastverteilung, um u.a. einen Vielzahl von Profilaktualisierungen innerhalb eines kurzen Zeitraums zu ermöglichen. Lesende Anfragen beantwortet der *Profile Endpoint* mit Hilfe des *Profile Info Caches*, der häufig angefragte bzw. zuletzt geänderte profilbezogene Informationen in einem performanten Key-Value-Cache vorhält. Der *Profile/Location Manager* ist für die Verwaltung aller profilbezogenen Daten des Systems verantwortlich. Lesende Anfragen des Nutzers werden an ihn weitergeleitet, wenn sie nicht direkt aus dem vorgeschalteten *Profile Info Cache* beantwortet werden können. Alle Daten werden persistent im *Location/Profile/Feedback Storage* gespeichert. Änderungen bei

Abonnements werden darüber hinaus an den *Demand/Incident Storage* übergeben und an den *Incident Notification Generator* weitergeleitet, um die Notwendigkeit einer Benachrichtigung zu prüfen.

4.5 Benachrichtigungsdienst

Der *Incident Notification Generator* ist für den Abgleich zwischen Abonnements (aktuelle Position) und Ereignissen verantwortlich. Mit dem *Demand/Incident Storage* verfügt er über eine spezielle Laufzeitdatenbank in der diese Informationen bereits zusammengeführt sind. Die Hauptaufgabe besteht darin, die aufbereiteten Ereignismeldungen mit den Zustelladressen der betreffenden Mithelfer zusammenzuführen und diese an den *Incident Notification Dispatcher* weiterzureichen. Zusätzlich verfügt diese Komponente über einen Speicher, in dem der Benachrichtigungszustand für einzelne Geräte/Nutzer gehalten wird. Weiterhin informiert der *Incident Notification Generator* vor dem Auslösen des Versandes von Benachrichtigungen andere Systemkomponenten darüber, von welchen Geräten demnächst mit Anfragen zu rechnen ist. So können systemseitige Vorbereitungen zur besseren Verfügbarkeit von Informationen getroffen werden.

Der *Incident Notification Dispatcher* ist für den Versand von Benachrichtigungen über verschiedene Kanäle verantwortlich. Er erhält vom *Incident Notification Generator* aufbereitete Ereignismeldungen mit entsprechenden Zustelladressen und versendet diese.

4.6 Nutzer-Feedback

Über den Feedback Endpoint werden alle vom Nutzer gegebenen Rückmeldungen (Einsatz angenommen/abgelehnt, Feedback nach einem Einsatz sowie Rückmeldung bzgl. der Basisprofilierung) entgegengenommen und an den Feedback Storage weitergeleitet. Weiterhin kann das Redaktionssystem über diesen Endpoint alle Feedback-Informationen einsehen und für den Disponenten aufbereiten oder als Filterparameter (Profilerweiterungen) für die Mithelfersuche verwenden.

4.7 Systemüberwachung

Der *System Watchdog* ist eine prüfende Systeminstanz, die kontrolliert, ob alle aktiven Komponenten in einem vorgegebenen Zeitrahmen ordnungsgemäß arbeiten. Der sogenannte *Heartbeat* ist ein komplexes Objekt, welches regelmäßig von jeder aktiven Komponente an den *System Watchdog* gesendet wird. Dieses Objekt enthält u.a. Informationen zum aktuellen Status, zu aufgetretenen Fehlern, zu verarbeiteten Objekten sowie zu Umgebungsvariablen (u.a. benötigter Speicher). Mit diesen Daten kann der *System Watchdog* bei fehlerhaftem Verhalten reagieren und mit Hilfe des *Command Interfaces* steuernd eingreifen.

Der *System Logger* nimmt von allen aktiven Komponenten die zur Protokollierung wichtigen Log-Einträge entgegen, bereitet diese Informationen auf und legt sie im *System Logs Storage* ab. Weiterhin bietet er eine Request-Schnittstelle an, um – unter anderem für Supportleistungen – die Log-Einträge sowohl von internen als auch externen Komponenten abgreifen zu lassen.

4.8 Performanz, Skalierbarkeit, Datensicherheit

Zur Realisierung von Performanz und Skalierbarkeit des Systems wurden folgende Entwurfsentscheidungen innerhalb der Architektur und der späteren Implementierung konsequent umgesetzt:

- Für eine schnelle Beantwortung lesender Anfragen werden Key-Value-Caches verwendet. Diese sind zu Zwecken der Lastverteilung aus mehreren verteilbaren Instanzen aufgebaut. Die Anzahl der Instanzen kann je nach erwarteter oder tatsächlicher Last dynamisch angepasst werden. Weiterhin bieten Key-Value-Speicherinstanzen hinsichtlich der Skalierbarkeit Vorteile gegenüber relationalen Speicherinstanzen.
- Alle Daten werden persistent in Datenbanken gespeichert. Das bedeutet, dass für alle Cache-Speicher-Instanzen eine persistente Datenhaltung vorliegt. Bei einem (unerwarteten) Neustart können so die in den Caches vorliegenden Informationen wiederhergestellt werden.

Die Kommunikation zwischen den einzelnen Komponenten im System-Backend erfolgt prinzipiell asynchron.

5 Zusammenfassung /Ausblick

In dem vorliegenden Papier wurde ein Systemkonzept für die Einbindung freiwilliger Helfer zur verbesserten Krisenbewältigung vorgestellt. Da sich dieses System noch in der Entwicklungs- und Erprobungsphase befindet, sind vor allem die Beschreibungen in den Kapiteln drei und vier eine Wiedergabe des aktuellen Stands (06/2015) und nicht zwingend als final anzusehen.

Aus technischer Sicht möchte ENSURE ein zusätzliches Instrument zur strukturierteren Unterstützung durch die Bevölkerung (u.a. bei Großschadenslagen) bereitstellen. Für diese Zielstellung wird auf Grundlage bestehender Anforderungen eine modulare Systemarchitektur, deren Komponenten für eine optimale Informationsverbreitung sowie -bereitstellung abgestimmt sind, konzipiert. Dabei werden vor allem Aspekte der Performance, Skalierbarkeit sowie Datenschutz und Datensicherheit beachtet. Weiterhin ist das hier beschriebene Systemkonzept als Vorarbeit für mögliche Erweiterungen des bestehenden Warnsystems KATWARN zu sehen.

Aufgrund des iterativen Vorgehens im Gesamtprojekt, erfolgen bis zum Projektabschluss noch mehrere Teststellungen (Alarmierungs- und Meldeübungen) sowie zwei große Feldtests (organisiert und geplant durch die Berliner Feuerwehr und das Deutsche Rote Kreuz), in denen die im Projekt erarbeiteten Konzepte gesamtheitlich evaluiert werden. Erste repräsentative Ergebnisse erfolgen nach dem ersten Feldtest, der im Oktober 2015 stattfinden wird.

6 Literaturverzeichnis

[Deutsches Rotes Kreuz e. V., 2014]

Deutsches Rotes Kreuz e. V. (2014). Die Rolle von ungebundenen HelferInnen bei der Bewältigung von Schadenereignissen - Teil 1. Berlin.

[Kaufhold & Reuter, 2014]

Kaufhold, M.-A., & Reuter, C. (01 2014). *Vernetzte Selbsthilfe in Sozialen Medien am Beispiel des Hochwassers 2013*. Abgerufen am 08. 07 2015 von http://www.wiwi.uni-siegen.de/cscw/publikationen/dokumente/2014/2014_kaufholdreuter_vernetzteselbsthilfehochwasser_icom.pdf

[Kircher, 2014]

Kircher, F. (08 2014). *Ungebundene Helfer im Kathastrophenschutz*. Abgerufen am 08.07.2015 von http://www.kat-leuchtturm.de/assets/content/images/pdfs/593_597_Kircher.pdf

[Quednau & Weber, 2014]

Quednau, T., & Weber, T. (10 2014). *Zwischen Forschungsprojekt und realer Katastrophe - INKA und die Flut 2013*. Abgerufen am 08.07.2015 von http://www.b-b-e.de/fileadmin/inhalte/aktuelles/2014/05/NL10_Gastbeitrag_Quednau_Weber.pdf

Mobiles Crowdsourcing von Pegeldaten

Frank Fuchs-Kittowski, HTW Berlin, frank.fuchs-kittowski@htw-berlin.de
Bernd Pfützner, BAH Berlin, bernd.pfützner@bah-berlin.de
Tobias Preuß, HTW Berlin, tobias.preuss@htw-berlin.de
Fabian Fischer, HTW Berlin, fabian.-fischer@t-online.de
Tilman Beck, HTW Berlin, tilman.beck@htw-berlin.de
Sarah Bartusch, HTW Berlin, bartusch.sarah@gmail.com

Abstract

With increasing damage caused by natural disasters like flood events in the last decade, the implementation of effective flood management systems has become an important topic. In this paper the hydrological mobile crowd sourcing system KoPmAn is presented, that uses mobile crowdsourcing to integrate water level data collected by the public in a flood management system. The underlying concepts, the architecture, and the implementation of KoPmAn are presented.

1 Einleitung

Hochwasser gehören zu den Naturkatastrophen, die regelmäßig Menschenleben fordern und hohe ökonomische Schäden verursachen [Merz et al., 2014, Blöschl et al., 2011]. Um (vorbeugende) Schutzmaßnahmen durchführen und durch Hochwasser verursachte Schäden reduzieren zu können, benötigen die verantwortlichen Behörden, aber auch die betroffenen Bürger aktuelle Informationen über den Wasserstand (und ggf. die aktuelle Gefährdung) an Gewässern [Hornemann, 2006]. Diese müssen gerade im Hochwasserfall in höherer Dichte, Frequenz sowie in Echtzeit verfügbar sein.

Allerdings ist die Digitalisierung von (Hochwasser-Melde-) Pegeln kostenintensiv, so dass nicht alle Pegel automatisiert sind und viele Pegel – insbesondere an kleineren

Gewässern - immer noch Lattenpegel sind. An diesen Lattenpegeln wird aber der Wasserstand meist nur einmal am Tag von „ausgewählten“ Personen (Helfern) vor Ort abgelesen und an die zuständige Behörde übermittelt. Damit erfolgt die Aktualisierung der Daten – insbesondere im Hochwasserfall - zu selten. Zudem entstehen durch die Überalterung der Helfer vor Ort zunehmend personelle Lücken. Die Automatisierung der Lattenpegel (automatisierte Meldepegel) ist derzeit kaum realisierbar, da diese in Anschaffung und Betrieb oftmals zu teuer sind [Maniak, 2010].

Mobiles Crowdsourcing – die Erfassung von Daten durch Freiwillige mittels ihrer eigenen mobilen Endgeräte (Smartphones, Tablets etc.) [Fuchs-Kittowski & Faust, 2014] - scheint ein geeignetes Mittel zu sein, um die unterschiedlichen Informationsbedürfnisse in den verschiedenen Phasen des Hochwassermanagements befriedigen zu können. Es bietet das Potenzial zur zeitlichen und räumlichen Verdichtung der Datenerfassung sowie zur Kompensation der fehlenden Helfer durch „normale“ Bürger (Crowdsourcees). Darüber hinaus ermöglicht es, die Bevölkerung einzubinden und zu beteiligen und somit für den Hochwasserschutz zu sensibilisieren [Meissen & Fuchs-Kittowski, 2014].

In diesem Beitrag wird eine mobile IT-Lösung zur Erfassung von Wasserständen an Gewässern beschrieben. Diese IT-Lösung ist als Client-Server-System konzipiert, das neben dem Backend-System (Server) über einen mobilen Client (App) für die Datenerfassung und -nutzung durch die Bürger sowie einen Web-Client (Dashboard) zur Echtzeit-Entscheidungsunterstützung für das Hochwassermanagement und die Einsatzkräfte verfügt. Mit diesem System sollen sowohl die Bevölkerung frühzeitig über Hochwassergefahren informiert (aktuelle Pegelinformationen und Wasserstandwarnungen) als auch die Entscheidungsfindung der Einsatzkräfte (Hochwasserzentrale) verbessert werden.

Dieser Beitrag ist wie folgt strukturiert: Im folgenden Kapitel 2 werden die Grundlagen zum mobilen Crowdsourcing im Hochwasserschutz dargestellt, Potenziale des Crowdsourcing herausgearbeitet und verwandte Arbeiten diskutiert. Das Konzept der mobilen IT-Lösung wird in Kapitel 3 und dessen technische Realisierung in Kapitel 4

beschrieben. Der Beitrag endet mit einer Zusammenfassung und einem Ausblick in Kapitel 5.

2 Grundlagen

2.1 Mobiles Crowdsourcing

Beim mobilen Crowdsourcing werden mobile Endgeräte für die Sammlung von Daten sowie die Koordination der an der Datensammlung freiwillig Beteiligten eingesetzt. Die Kernidee beim mobilen Crowdsourcing ist, dass normale Bürger befähigt werden, Daten über sich und die sie umgebende Umwelt mit ihren eigenen mobilen Endgeräten zu sammeln und zu teilen. Die Beteiligten tragen die Daten freiwillig, zu ihrem eigenen Nutzen oder zum Nutzer einer Gemeinschaft (Community) bei. Dabei wird keine aufgabenspezifische Spezialhardware verwendet, sondern normale, für den Massenmarkt verfügbare mobile Endgeräte, wie Smartphones und Tablets. Oft handelt es sich um orts- oder raumbezogene, häufig auch zeitbezogene Messwerte, Daten oder Informationen [Fuchs-Kittowski & Faust, 2014].

Es sind bereits viele unterschiedliche mobile Crowdsourcing-Anwendungen bekannt, insb. im Natur-, Umwelt- und Katastrophenschutz. Im Katastrophenschutz kann mobiles Crowdsourcing effektiv unterstützend während und nach einer Katastrophe eingesetzt werden. Beispielsweise können während einer Katastrophe die Bewohner sowie Rettungs- und Einsatzkräfte in einem Katastrophengebiet Informationen über die aktuelle Lage sammeln, so dass Hilfs- und Rettungskräfte gezielter eingesetzt werden können. Nach einer Natur-Katastrophe (wie Hochwasser, Sturm, Starkregen) ist bspw. das Finden, Dokumentieren und Bewerten von Schäden (wie umgeknickte Bäume, überflutete Wege oder umgeknickte Strommasten) eine wichtige Aufgabe, um die begrenzten Ressourcen gezielter und schneller für die Reparatur und Schadensbehebung einsetzen zu können sowie die tatsächlich aufgetretenen Schäden aktueller und umfassender auswerten zu können [Abecker et al., 2012].

2.2 Pegel im Hochwasserschutz

Wasserstände können verhältnismäßig einfach gemessen werden. Im Hochwasserschutz werden daher Pegel zur Messung der Wasserstände und Durchflüsse eingerichtet und betrieben, um auf Basis dieser Daten die Entwicklung des Wasserstandes, d.h. eine mögliche Hochwasserentstehung bewerten zu können [Maniak, 2010].

Arten von Pegeln: In Deutschland werden zum Messen des Wasserstandes zunehmend automatisierte Pegel eingesetzt, die nicht nur automatisch den Wasserstand messen (digitale Datensammler), sondern auch über eine Datenfernübertragung (DFÜ) verfügen, damit die Daten regelmäßig (ca. alle 5-15 Minuten) abgefragt/übertragen werden können oder sich der Pegel bei Überschreiten eines Grenzwertes selbst automatisch melden kann (Aktivpegel, Pegel mit Meldefunktion). Weniger verbreitet sind inzwischen sog. Schreibpegel, die die Messwerte kontinuierliche auf Pegelbögen / Bandschriebe aufzeichnen. Die aufgezeichneten Daten werden regelmäßig von ehrenamtlichen Pegelbeobachtern der zuständigen Behörde übersandt. Lattenpegel, an denen der Wasserstand manuell abgelesen werden muss, werden nur selten zur regelmäßigen Beobachtung des Wasserstandes eingesetzt. Während automatisierte Pegel und Schreibpegel vor allem für größere Gewässer, wie die Elbe oder die Havel, eingesetzt werden, existieren aus Kostengründen an kleineren Fließgewässern keine Pegel oder nur ein Lattenpegel.

Crowdsourcing-Potenziale: Lattenpegel werden aufgrund des erforderlichen hohen personellen Aufwands nicht regelmäßig abgelesen (oftmals nur einmal täglich oder nur im Hochwasserfall). Für die Prognose von Hochwasserereignissen sind diese Daten meist unbrauchbar, denn signifikante Pegeländerungen können sich binnen weniger Stunden ergeben. In kleinen Einzugsgebieten spielen sich gar komplette Hochwasserereignisse in wenigen Stunden ab [Dyck & Peschke 1995, S. 492]. Mittels Crowdsourcing könnte die Aktualität und Verfügbarkeit der Pegeldaten an Lattenpegeln deutlich erhöht werden, indem diese regelmäßig und in höherer Frequenz von freiwilligen Bürgern abgelesen werden. Auch bei Schreibpegeln könnte die Verfügbarkeit der Daten erhöht werden, indem die aufgezeichneten Daten häufiger per Crowdsourcing übertragen werden. Bei automatisierten Pegeln ist keine Übertragung

der gemessenen Daten erforderlich. Allerdings kann auch hier Crowdsourcing sinnvoll eingesetzt werden. Bspw. ist jede Pegelstelle zur Kontrolle des Schreibpegels mit einem Lattenpegel ausgestattet. Durch Crowdsourcing-Meldungen des Wasserstandes am Lattenpegel kann die Richtigkeit der vom Schreib-Pegel gemessenen Werte bestätigt werden. Damit können Messfehler erkannt werden oder auch etwaige Anstauungen von Wasser am Pegel, die das Messergebnis des Pegels beeinflussen (z.B. durch einen umgestürzten Baum nach einem Sturm), gemeldet werden. Zudem lässt sich die Datenverfügbarkeit in der Bevölkerung erhöhen, da die Daten einiger Pegel teilweise erst nach nur monatsweise stattfindender Prüfung verfügbar gemacht werden. In dem Zeitraum bis die geprüften Daten vorliegen, könnten die über Crowdsourcing ermittelten Daten genutzt werden.

Verwandte Arbeiten: Während es zahlreiche mobile Anwendungen für die Information über Wasserstände an Gewässern gibt (z.B. Pegel[39], Pegelstand[40], Pegel-Online[41], Pegelstände[42]), existieren nur sehr wenige mobile Apps für die Erfassung der Pegelstände durch seine Nutzer. Bspw. wurden im Forschungsprojekt MAGUN mobile Anwendungen für die Erfassung aktueller Wasserstände und historischer Hochwassermarken entwickelt [Fuchs-Kittowski et al., 2012], welche aber nicht flächendeckend eingesetzt werden. Das aktuell laufende EU-Projekt „WeSenseIt - Citizen water observatories“ zielt auf die Erfassung von ganz unterschiedlichen Daten zu Gewässern durch Bürger mit ihren mobilen Endgeräten sowie zusätzlichen Sensoren ab, wobei zwar der Fokus auf der Erfassung und Sammlung der Daten für verschiedenste Einsatzzwecke liegt, allerdings eine Integration der Daten oder Prozesse in konkrete Einsatzszenarien – wie Hochwassermanagement - fehlt.

[39] https://play.google.com/store/apps/details?id=de.posts.Pegel
[40] https://play.google.com/store/apps/details?id=info.pegelstand.pegelstandnoebasic
[41] https://play.google.com/store/apps/details?id=org.cirrus.mobi.pegel
[42] https://play.google.com/store/apps/details?id=com.lifestream_creations.pegelmelder

2.3 Dashboards im Hochwassermanagement

Hochwassermanagement: Hochwasser(risiko)management ist ein fortlaufender iterativer Prozess, der prinzipiell in drei Phasen verläuft: Hochwasserbewältigung (während des Hochwasserereignis), Regeneration (danach) und Hochwasservorbeugung (davor) [Müller, 2010]. Vorbeugender Hochwasserschutz (Hochwasservorsorge) ist grundsätzlich kostensparender als die Beseitigung wiederkehrender und sich häufender Hochwasserschäden. Neben der Bauvorsorge können Schäden insbesondere durch die Verhaltensvorsorge, d.h. durch Hochwasservorhersagen, durch Hochwasserwarnungen und durch planvolles Handeln während des Hochwassers (Alarm- und Einsatzpläne) Schäden verringert werden. Besondere Bedeutung haben dabei Hochwasser-Monitoring und Vorhersagesysteme, die eine frühzeitige Gefahren- und Risiko-Erkennung erlauben und so die gezielte Einleitung von Schutzmaßnahmen ermöglichen.

Dashboards: Ein digitales Dashboard ist ein Informationssystem, das auf einem einzigen Bildschirm sichtbar und nutzbar ist. Management Dashboards bieten einen Überblick über die wichtigsten Informationen, die erforderlich sind, um strategische Maßnahmen (z.B. Zielsetzen, Planen, Vorhersagen) und die Überwachung von deren Verläufen [Few, 2006] sowie das Treffen von strategischen Entscheidungen [Hosack, et al. 2012] zu unterstützen. Zentrale Eigenschaften von Management Dashboards sind Überblick über relevante Vorgänge und Zustände (z.B. Grafiken, Key Performance Indikatoren (KPI), Vorhersagen), Anpassbarkeit an die aktuellen Ziele der Nutzer (z.B. Ein- und Ausblenden) sowie schneller Zugriff auf die zugrundeliegenden Daten (drill down) [Vasiliu, 2006].

Verwandte Arbeiten: Digitale Dashboards werden in verschiedenen Anwendungsdomänen erfolgreich eingesetzt. Im Katastrophenmanagement wurden bspw. Dashboards entwickelt, die dem Nutzer dynamisch die gerade benötigte Information anzeigen [Zheng et al., 2010] oder KPI visualisieren [Bharosa, et al. 2010]. Das Hochwasser- und Sturmflutinformationssystem Schleswig-Holstein (HSI) ist ein Informationsportal, das unterschiedliche Angaben zu Wasserständen und Hochwasser in Schleswig-Holstein sowie Hintergrundinformationen zu diesen Themen präsentiert

[Hosenfeld & Hach, 2012]. AGORA-GeoDash ist ein digitales Geo-Sensor Dashboard für das Hochwassermanagement, bei dem Echtzeitinformationen aus automatisierten Hochwasserpegeln aufbereitet und visualisiert werden, um die Entscheidungsfindung von Einsatzkräften zu unterstützen und eine Echtzeit-Hochwasserrisikobewertung zu ermöglichen [Horita et al., 2011]. Im Rahmen des WeSenseIt-Projekts wurde Kite entwickelt, das als zentraler Zugriffspunkt auf von Bürgern beigetragenen Information dient und von Fachanwendern wie Bürgern genutzt werden kann [Lanfranchi et al., 2014].

3 Konzept

In diesem Abschnitt wird die mobile IT-Lösung zur Erfassung von Wasserständen an Gewässern beschrieben. Diese IT-Lösung ist als Client-Server-System konzipiert, das neben dem Backend-System (Server) über einen mobilen Client (App) und einen Web-Client (Dashboard) verfügt.

3.1 Mobile Crowdsourcing App

Die mobile App (Client) soll von der Bevölkerung eingesetzt werden. Sie dient der Information über die vorhandenen Pegel und die jeweils aktuellen Wasserstände. Zudem erhalten die Nutzer Warnungen beim Überschreiten von Grenzwerten. Vor allem dient die App aber der Erfassung von Wasserständen an Pegeln durch die Nutzer. Dabei werden die Nutzer bei der Lokalisierung (Marker auf Karten) und Identifizierung der Pegel (QR-Code) unterstützt. Zudem können sie Fotos vom Pegel aufnehmen und übertragen, um die Überprüfung der erfassten Daten zu ermöglichen.

Abbildung 1: Grundfunktionen der mobilen App

Die mobile Client-Anwendung (App) lässt sich in folgende Grundfunktionen zerlegen:

- In einer Kartenansicht hat der Benutzer die Möglichkeit, Pegelstandorte visuell aufzufinden. Eine zusätzliche Suchfunktion soll diesen Prozess unterstützen. Die Karte zeigt neben den Standorten auch das durch eine Serverkomponente bereitgestellte Gewässernetz – also Seen und Flüsse (Abbildung 1, Screen 1).
- Durch Auswählen (Berühren-Geste) einer Pegelstation oder eines Gewässers, können hydrologische Fachinformationen abgerufen werden. Diese werden über der Karte als Informationsbereich eingeblendet (Abbildung 1, Screen 2).
- Über ein über das Menü erreichbares Eingabeformular kann der Nutzer selbst aktiv werden. Dort können der Wasserstand, Nutzerinformationen, der Zeitpunkt der Ablesung eingetragen sowie ein Foto angehängt werden. Diese Crowdsourcing-Daten werden durch Absenden an den Server übermittelt (Abbildung 1, Screen 3).
- Weiterhin lassen sich zu jeder Pegelstation bereits gesammelte Wasserstanddaten abrufen. Über das Menü gelangt der Nutzer zu einer Listen- bzw. grafischen Darstellung der zurückliegenden Werte. In der Listendarstellung (Abbildung 1, Screen 5) werden die Wasserstandwerte in chronologischer Reihenfolge angezeigt. Jeder Listeneintrag enthält neben dem eigentlichen Ablesewert auch den Namen des Erfassers, den Zeitpunkt der Ablesung sowie ein Foto von der Pegelstation. Um den Messwerteverlauf grafisch anzeigen zu lassen, kann der Nutzer über das Menü auf die Diagrammansicht umschalten. Dort werden dieselben Messwerte als Liniendiagramm mit Punkten visualisiert (Abbildung 1, Screen 4).

Abbildung 2: Warn- und Datenanfrage-Funktionen der mobilen App

Für jeden Pegelstandort bietet die App die Möglichkeit, Warnungen zu aktivieren. Beim Überschreiten des seitens der Behörde definierten Warnlevel löst die App Alarm aus. Optional kann der Nutzer eigene Warnstufen mit Wasserstandwert und Warnton definieren, die auf die gleiche Weise funktionieren (Abbildung 2, Screen 1). Diese Alarme erscheinen dann als Benachrichtigung im Statusbereich des Smartphones oder Tablets und lösen den akustischen Alarm aus (Abbildung 2, Screen 4).

3.2 Management Dashboard

Das Dashboard (Web-Anwendung) wird von der Behörde eingesetzt und dient zur Echtzeit-Entscheidungsunterstützung für das Hochwassermanagement und die Einsatzkräfte. Es ermöglicht die Überwachung der Wasserstände an den verschiedenen Pegeln und deren Verläufe (Ganglinien) sowie daraus abgeleitete Kenngrößen. Hierfür werden die gesammelten Crowdsourcing-Daten gemeinsam mit den Messwerten der Pegel übersichtlich auf einer Karte und verschiedenen Auswertungs-Darstellungen (Ganglinie, Liste etc.) dargestellt. Das Dashboard dient aber auch der Verwaltung der gesammelten Crowdsourcing-Daten (insb. Qualitätskontrolle) sowie der Aktivierung (Crowd Tasking) und Verwaltung der Nutzer.

Das Dashboard zum Management der Pegeldaten besteht aus einer interaktiven Übersichtskarte und einer Detailansicht. Ähnlich zur Android-App lassen sich Pegel auffinden und selektieren. In der Detailansicht sind die Stammdaten zur ausgewählten Station sichtbar sowie eine grafische Darstellung (Ganglinie) der Zeitreihe der Wasserstände. Diese lassen sich auch in Listenform abrufen. Einzelne Crowdsourcing-

Daten können vom Nutzer der Anwendung geprüft und ggf. anhand des mitgesendeten Bildes korrigiert oder im Fehlerfall verworfen werden.

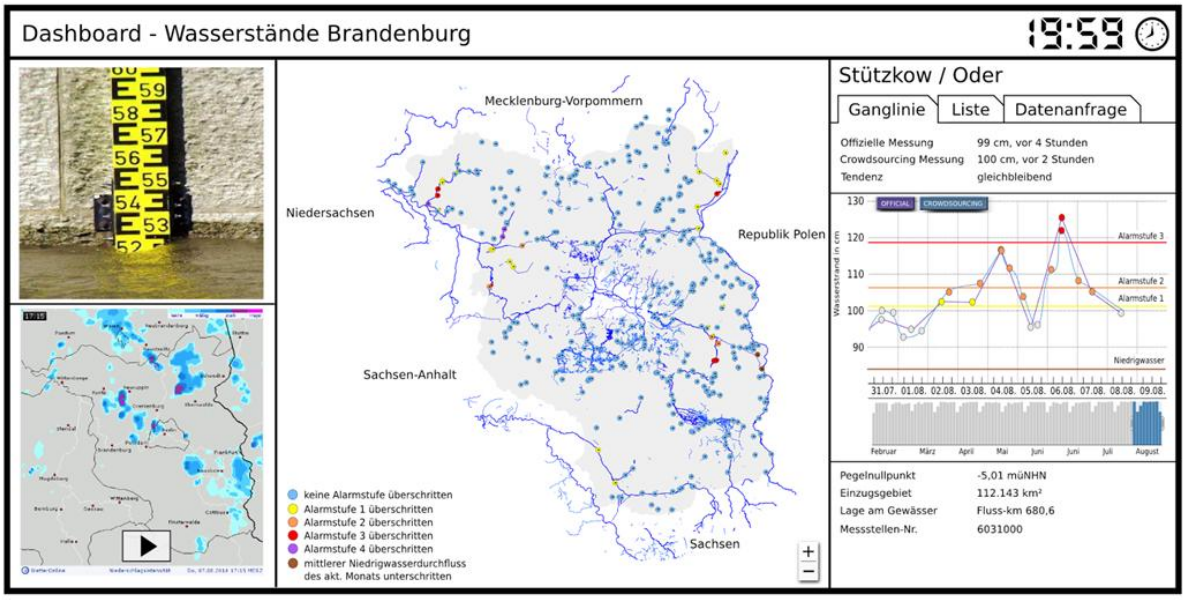


Abbildung 3: Oberflächen-Konzept des Management Dashboard

3.3 Datenanfrage-Workflow

Da die Behörde auf die Aktivität von Freiwilligen angewiesen ist, liegt eine große Herausforderung in der Motivation und Aktivierung der Nutzer. Hierfür wurde ein Konzept entwickelt, wie eine relevante Anzahl von Freiwilligen aktiviert werden kann, ohne die große Masse der Nutzer mit ungewollten Anfragen zu belästigen.

Auf Seiten der Freiwilligen soll es erst einmal grundsätzlich möglich sein, jegliche Anfragen seitens der Behörde an- bzw. abzuschalten. Damit soll die Selbstbestimmung der Nutzer garantiert werden und eine negativ empfundene Nachrichtenüberflutung vermieden werden (Abbildung 2, Screen 1 unten). Gleichzeitig kann der Nutzer Anfragen für bestimmte Pegelstandorte erlauben, und somit seine Bereitschaft zu helfen signalisieren. Diese Information wird an die Behörde übermittelt. Dadurch liegt eine eindeutige Zuordnung von Pegelstationen und potentiellen Teilnehmern einer Datenanfrage vor. Die Behörde kann dann im Bedarfsfall über das Dashboard einen bestimmten Teil der Helfer aus dieser Menge adressieren (Abbildung 3, rechts, Reiter „Datenanfrage“), worauf diese eine Anfrage-Nachricht in ihrer mobilen App erhalten (Abbildung 2, Screen 5).

3.4 Backend-System

Das Backend-System besteht aus drei zentralen Komponenten: eine Komponente zur Erfassung der Crowdsourcing-Daten, eine zur Bereitstellung der Messwerte (Crowdsourcing und andere Systeme) als Zeitreihen und eine Komponente zur Auslieferung von Geo-Daten (Informationen zum Gewässernetz, Pegelstammdaten etc.) an die Clients. Darüber hinaus bestehen Schnittstellen vom Backend zu den Pegelmesswerten in anderen Systemen.

Beim Backend-System kann man zwischen Komponenten unterscheiden, auf die aus Client-Sicht lesend bzw. schreibend zugriffen wird. Den erstgenannten ist die Komponente zur Auslieferung von Geo-Daten an die Clients zuzuordnen. Diese stellt einerseits die geometrischen Informationen zum Gewässernetz (als Kacheln durch einen sogenannten Tile Server) und anderseits die zugehörigen strukturierten Daten (z.B. Standort- und hydrologische Fachinformationen) bereit. Ebenfalls lesend wird auf die Komponente zur Bereitstellung der Messwerte als Zeitreihen zugegriffen. Über diese Schnittstelle kann die Client-Anwendung Werte für einen bestimmten Zeitraum abrufen, um diese darzustellen. Diese Komponente erfasst zudem die aktuellen Live-Daten aus den verschiedenen Datenquellen. Schreibender Zugriff erfolgt auf die Komponente zur Erfassung der Crowdsourcing-Daten. An diese REST-Schnittstelle übermittelt der mobile Client gesammelte Daten, die in einer Datenbank gespeichert werden.

4 Technische Realisierung

Zur technischen Realisierung dieses Systems werden verschiedene Technologien eingesetzt, die im Folgenden kurz skizziert werden.

4.1 Client für Smartphone und Tablet

Die mobile Client-Anwendung wird als Android-App entwickelt. Die wichtigsten Bereiche der Anwendung sind die Übersichtskarte (Abbildung 4, Screen 1) mit Details zur ausgewählten Pegelstation (Abbildung 4, Screen 2) oder einem Gewässer, das Formular zur Eingabe des beobachteten Wasserstands (Abbildung 4, Screen 3), eine

Listendarstellung der Wasserstandwerte sowie deren grafische Darstellung (Abbildung 4, Screen 4).

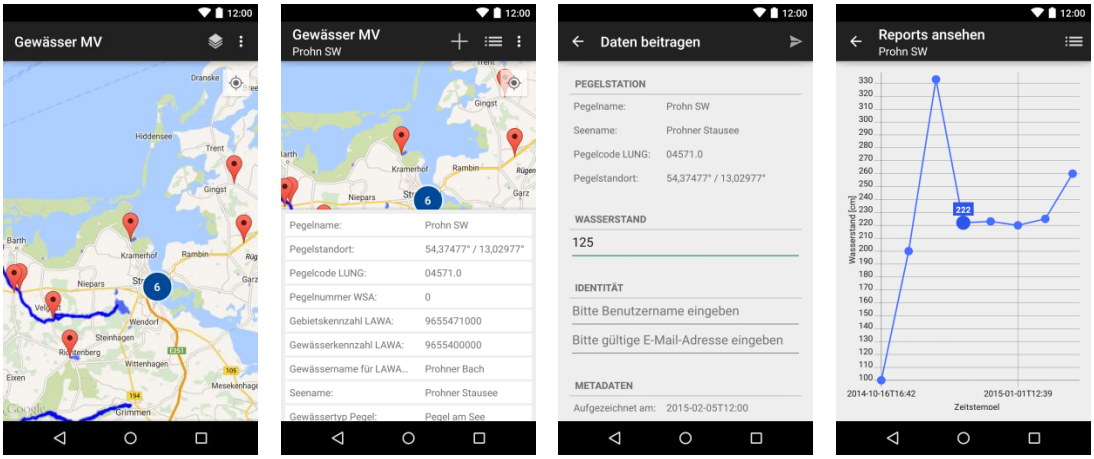


Abbildung 4: Screenshots der mobilen Crowdsourcing-App

Die Übersichtskarte dient dem Nutzer zum Auffinden der Pegelstandorte (Abbildung 4, Screen 1). Zudem stehen über das Menü eine Suchfunktion zum Auffinden von Pegeln sowie eine QR-Code-Einlese-Funktion zum Identifizieren einer Pegelstation zur Verfügung. Für das Clustering der Pegelstationen auf der Karte wird *Android Map Utils* verwendet. Alle vom und zum Server übertragenden Daten werden im JSON-Format übermittelt und auf dem Client mittels *Jackson* serialisiert bzw. deserialisiert. Als HTTP-Client liegt *OkHttp* zu Grunde. Der Nutzer kann fachspezifische, hydrologische Informationen zu einem Gewässer oder einer Pegelstation aufrufen. Diese werden initial am unteren Bildschirmrand angezeigt (Abbildung 4, Screen 2) und können durch Schieben nach oben bewegt werden. Dieses Modul wurde mit der Bibliothek *Android Sliding Up Panel* realisiert. Für einen selektierten Pegel kann der Nutzer im Formular (Abb. Abbildung 3) den Wasserstand, Kontaktdaten, den Zeitpunkt der Messung eingeben sowie ein Foto des Pegels anhängen. Für die Erstellung des Fotos wird die jeweilige Foto-App genutzt. Alternativ lassen sich die Wasserstanddaten in Listenform und als Diagramm (Abb. Abbildung.4) betrachten. Die Diagrammdarstellung wurde mit *HelloCharts* umgesetzt.

4.2 Management Dashboard

Das Management Dashboard ist eine Web-Anwendung, die auf HTML5 und CSS3 basiert und zu deren Realisierung verschiedene Frameworks eingesetzt werden (jQuery, Google Maps JS API, Flot, Bootstrap etc.).

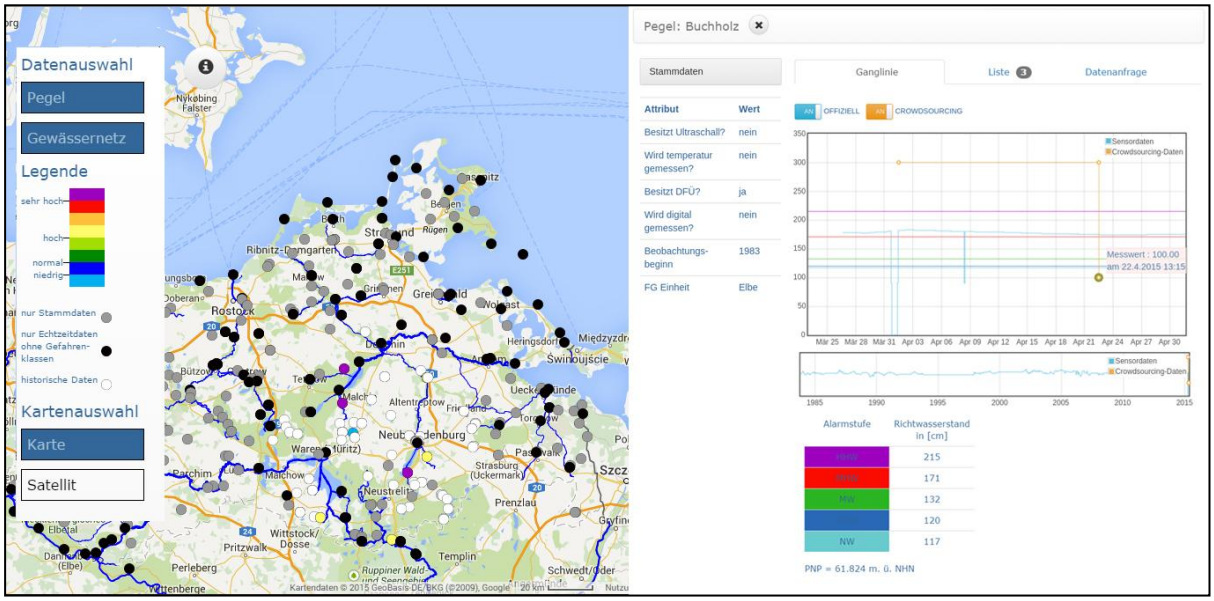


Abbildung 5: Screenshot des Management Dashboard

Über das Dashboard lassen sich Historisch-, Echtzeit- und Crowdsourcing-Pegeldaten in einem Gangliniendiagramm anzeigen und zum Pegel dazugehörige Stammdaten abrufen. Historische und aktuelle Datensätze werden in einem Diagramm zusammen dargestellt. Sind Crowdsourcing-Daten für einen Pegel verfügbar, können diese zusätzlich in das Diagramm ein- und ausgeblendet werden. Die Crowdsourcing-Datensätze lassen sich über ein Monitoring-System kontrollieren und gegebenenfalls korrigieren. Um die Performance des Management Dashboards zu steigern, werden die Crowdsourcing-Datensätze mittel „lazy-loading“ nacheinander vom Server abgerufen, so dass stets ein kleiner Anteil der Daten vorgeladen bereitsteht. Sollten aktuelle Messwerte für eine Pegelstation benötigt werden, können für diesen Pegel registrierte Nutzer benachrichtigt werden.

4.3 Backend

Auch zur Realisierung der Komponenten des Backend wurden unterschiedliche Technologien eingesetzt:

Crowdsourcing-Service: Als Server-Komponente für das Verarbeiten der Crowdsourcing-Daten wurde eine für diese Aufgabe spezifische Schnittstelle zur Datenbank konzipiert. Die API basiert auf dem *Django REST Framework*. Die Instanz umfasst Schnittstellen zum Anlegen von neuen Wasserstandsmeldungen, zum Auslesen gespeicherter Datensätze sowie zum Bewerten einzelner Meldungen und Auslesen dieser Bewertungen.

Zeitreihen-Service: Der Zeitreihen-Service importiert zum einen Messpunkte von externen Datenquellen in die Zeitreihendatenbank und stellt zum anderen eine REST-Schnittstelle zum Ausliefern von Zeitreihen bereit. Er wurde in NodeJS implementiert und orientiert sich an der OGC-Spezifikation des Sensor Observation Service (SOS)[43]. Ursprünglich war statt einer Eigen-Implementierung der Einsatz einer der verfügbaren Implementierungen für den SOS-Standard (z.B. die Java-basierte Open Source-Implementierung von 52North[44]) geplant. Da aber nur ein Bruchteil des von SOS spezifizierten Funktionsumfanges benötigt wurde, wurde doch eine auf die spezifischen Anforderungen zugeschnittene Lösung implementiert.

Tile-Service: Als Tile-Service zum Ausliefern von vorgenerierten Karten-Daten (wie z.B. Gewässernetz) kommt die GeoWebCache-Komponente des GeoServer[45] zum Einsatz.

Stammdaten-Service: Die Stammdaten (z.B. von Pegeln) werden vom integrierten holgAR-Server ausgeliefert. Dabei handelt es sich um eine Eigenentwicklung auf Basis von Java, die strukturierte Daten, in einer projektzentrierten Konfiguration, auf Basis von geographischen Abfragen in unterschiedlichen Ausgabeformaten zurückliefert [Fuchs-Kittowski et al., 2012].

[43] http://www.opengeospatial.org/standards/sos
[44] http://52north.org/communities/sensorweb/sos/
[45] http://geowebcache.org/docs/current/

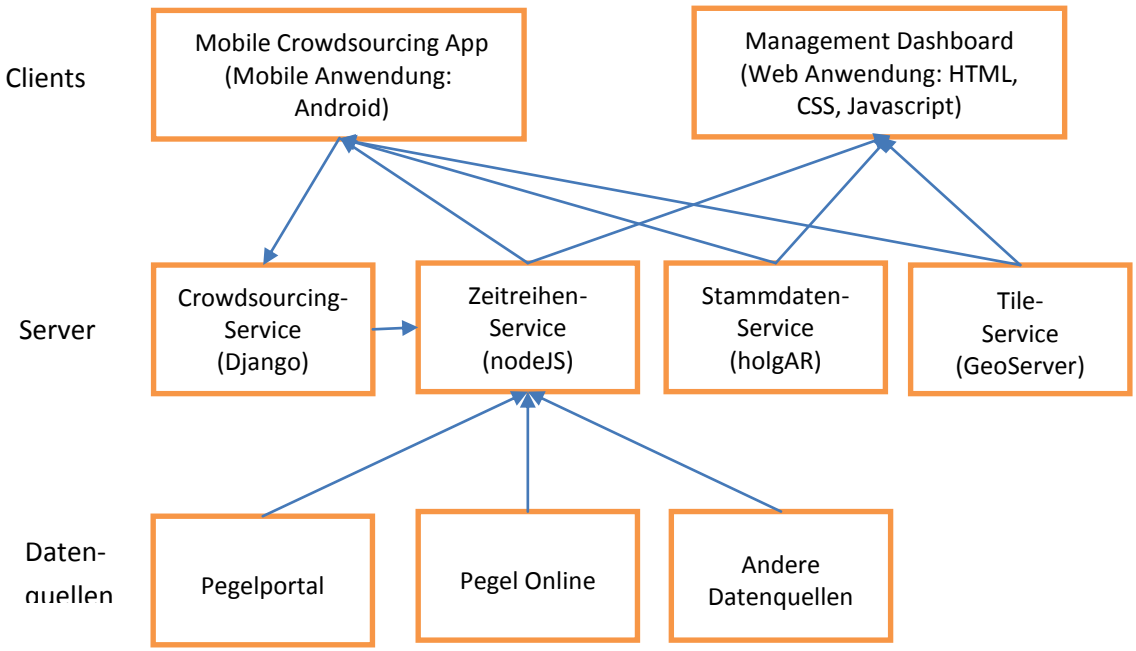


Abbildung 6: Architektur

Darüber hinaus bestehen Schnittstellen vom Backend zu anderen Datenquellen, d.h. Pegelmesswerten in anderen Systemen (u.a. für Live-Daten zu z.B. Pegelportal oder Pegel-Online[46]) in Form von Web-Services oder XML-Dateien.

5 Zusammenfassung und Ausblick

In diesem Beitrag wurde die Konzeption und die Implementierung eines Hochwassermanagementsystems vorgestellt, das darauf abzielt, sowohl die Bevölkerung frühzeitig über Hochwassergefahren zu informieren (aktuelle Pegelinformationen und Wasserstandwarnungen) als auch die Entscheidungsfindung der Einsatzkräfte (Hochwasserzentrale) durch eine höhere Informationsdichte und -aktualität zu unterstützen, die durch die Sammlung von Daten durch die Bevölkerung mittels ihrer persönlichen mobilen Endgeräte (mobiles Crowdsourcing) erreicht wird.

Neben der Erhöhung der Datendichte und -aktualität ist eine weitere wichtige Verbesserung in der Beteiligung der Bevölkerung zu sehen, was die Akzeptanz und Transparenz von Hochwasserschutz erhöhen wird. Darüber hinaus sind die durch die Bevölkerung erfassten Daten für eine Reihe weiterer hydrologische Prognosen und

[46] https://www.pegelonline.wsv.de/webservice/guideRestapi

Zwecke verwendbar, wie Erkenntnisgewinn über das Verhalten von Einzugsgebieten bei verschiedenen Niederschlagssituationen, Bemessung von Hochwasserschutzanlagen oder Simulationen von Auswirkungen von Landnutzungsänderungen.

Das mobile System zur Erfassung von Pegeldaten wird derzeit in Kooperation mit der Umweltbehörde eines Bundeslandes entwickelt und soll in einer spezifischen Region getestet und evaluiert werden.

6 Danksagung

Dieser Beitrag ist im Rahmen des Projekts „KoPmAn“ entstanden. Das Vorhaben wird gefördert aus Mitteln der Europäischen Union (Europäischer Sozialfonds). Die Verfasser danken für die Unterstützung.

7 Literaturverzeichnis

[Abecker et al., 2012]

Abecker, A., Braun, S., Kazakos, W., Zacharias, V.: Participatory Sensing for Nature Conservation and Environment Protection. In: 26th International Conference on Informatics for Environmental Protection (EnviroInfo2012), Shaker, Aachen, 2012, S. 393-401.

[Bharosa, 2010]

Bharosa, N.; Meijer, S.; Janssen, M.; Brave, F.: Are we prepared? Experiences from developing dashboards for disaster preparation. In: 7th ISCRAM Conference, Seattle, USA, 2010, S. 1-6.

[Blöschl, et al. 2011]

Blöschl, G., A. Viglione, R. Merz., J. Parajka, J. Salinas und W. Schöner (2011) Auswirkungen des Klimawandels auf Hochwasser und Niederwasser. Österreichische Wasser- und Abfallwirtschaft, 63, (1-2), S. 21- 30.

[Few, 2006]

Few, Stephen: Information Dashboard Design - the effective visual communication of data. O'Reilly Media, 2006.

[Fuchs-Kittowski et al., 2012]

Fuchs-Kittowski, Frank; Simroth, Stefan; Himberger, Sebastian; Fischer, Fabian; Schley, Maxi: A content platform for smartphone-based mobile augmented reality, Proceedings of the 26. International Conference on Informatics for Environmental Protection (EnviroInfo2012), Shaker, Aachen, 2012 S. 403-412.

[Fuchs-Kittowski & Faust, 2014]

Fuchs-Kittowski, Frank; Faust, Daniel: Architecture of Mobile Crowdsourcing Systems. In: Collaboration and Technology - Proceedings of the 20th International Conference on Collaboration and Technology (CRIWG 2014), Springer International Publishing, Berlin, Heidelberg, New York, 2014, S. 121-136.[Horita et al., 2011]

[Horita et al., 2014]

Horita, Flávio; Fava, Maria Clara; Rotava, Jairo; Ueyama, Jó; Albuquerque, João Porto de; Souza, Vladimir; Mendiondo, Eduardo Mario: AGORA-GeoDash: A Geosensor Dashboard for Real-time Flood Risk Monitoring. In: S.R. Hiltz; M.S. Pfaff; L. Plotnick; A.C. Robinson (Hrsg.): Proceedings of the 11th International ISCRAM Conference – University Park, Pennsylvania, USA, May 2014.

[Hornemann, 2006]

Hornemann, Corinna; Rechenberg, Jörg: Was Sie über den vorsorgenden Hochwasserschutz wissen sollten. Dessau: Umweltbundesamt, 2006.

[Hosenfeld & Hach, 2012]

Hosenfeld, Friedhelm; Hach, Ralf: Hochwasser- und Sturmflutinformationssystem Schleswig-Holstein (HSI-SH). In: Umweltinformationssysteme - Frühwarn- und Informationssysteme für den Hochwasserschutz, UBA-Texte 41/2012, Umweltbundesamt, Dessau, 2012, S. 9-18.

[Lanfranchi et al., 2014]

Lanfranchi, Vitaska; Wrigley, Stuart N.; Ireson, Neil; Ciravegna, Fabio; Wehn, Uta: Citizens' Observatories for Situation Awareness in Flooding. In: Proceedings of the 11th International ISCRAM Conference – University Park, Pennsylvania, USA, May 2014

[Maniak, 2010]

Maniak, Ulrich: Hydrologie und Wasserwirtschaft. 6. Neu bearb. Auflage, Springer, 2010.

[Meissen & Fuchs-Kittowski, 2014]

Meissen, Ulrich; Fuchs-Kittowski, Frank: Towards a reference architecture of crowd-sourcing integration in early warning systems. In: Proceedings of the 11th International ISCRAM Conference, Pennsylvania State University, Pennsylvania, USA, 2014, S. 339-343.

[Merz et al., 2014]

Merz, B., J. Aerts, K. Arnbjerg-Nielsen, M. Baldi, A. Becker, A. Bichet, G. Blöschl, L.M. Bouwer, A. Brauer, F. Cioffi, J.M. Delgado, M. Gocht, F. Guzzetti, S. Harrigan, K. Hirschboeck, C. Kilsby, W. Kron, H.-H. Kwon, U. Lall, R. Merz, K. Nissen, P. Salvatti, T. Swierczynski, U. Ulbrich, A. Viglione, P.J. Ward, M. Weiler, B. Wilhelm, M. Nied (2014) Floods and climate: emerging perspectives for flood risk assessment and management. Natural Hazards and Earth System Sciences, 14, 1921-1942, doi: 10.5194/nhess-14-1921-2014.

[Müller, 2010]

Müller, Uwe: Hochwasserrisikomanagement - Theorie und Praxis. Vieweg & Teubner, 2010.

Neue Technologien in der Forstwirtschaft am Beispiel der Apps WaldFliege und WaldKarte

Christine Müller, inforst UG, mueller@inforst.de

Abstract

Although forestry may seem to be a rather traditional branch with little interest in new technologies the use of android apps offers significant advantages in forestry and management of natural resources. The need of foresters, forest workers and logistic companies to collect and store data in the forest is met by android devices. The android apps “WaldFliege” and “Forest(M)ap(p)” offer a solution for optimizing timber logistic. With WaldFliege people working in the forest can collect timber data and combine it with GPS data, photos and data of forest owners. Forest(M)ap(p) shows the position of the stacks of timber in an offline map. Places can be marked by means of the GPS but also by simply tapping on the map if no satellite signal is available. Inforst also offers measuring with UAVs in forestry. This can simplify the planning of timber harvesting or the management of forest damages. The transfer of data from the multicopter directly to the offline map in the app is in project.

1 Einleitung

Die Forstwirtschaft wird als allgmein eher traditionelle, technikferne Branche angesehen. Jedoch ist der Umweltbereich sicher ein Bereich, in dem durch Android-Geräte ein großer Zusatznutzen entstehen kann. Das hat folgende Gründe:

- Als mobile Geräte sind Smartphones und Tablets in Wald und Feld verfügbar
- Es gibt Outdoor-geeignete Android-Geräte zu deutlich günstigeren Preisen als die früher üblichen mobilen Datenaufnahmegeräte

- Android-Geräte sind vielseitig einsetzbar
- GPS-Daten können aufgenommen werden
- Zahlreiche Schnittstellen sind verfügbar
- Die telefonische Erreichbarkeit der Arbeitskräfte ist sowieso vorausgesetzt

Inforst entwickelt daher Android-Apps für die Forstwirtschaft. Aus diesem Einsatzgebiet ergeben sich folgende Anforderungen:

- Die Apps müssen komplett offline lauffähig sein, weil im Wald nicht immer (ausreichend) Netzempfang ist.
- Die Apps müssen sehr flexibel sein, denn die Geschäftsabläufe in der Forstwirtschaft sind sehr vielfältig. Verschiedene Aufnahmemethoden, Begriffe und Einheiten werden verwendet.
- Die Datenbank muss gut strukturiert sein aufgrund der Strukturierung der Daten in Hieb, Los, Polter und Einzelstamm mit der Möglichkeit, einzelne Stämme und Polter in andere Lose zu verschieben. Die Verbindungen zwischen den einzelnen Tabellen müssen erhalten bleiben.
- Die Apps müssen einfach zu bedienen und auch für eher technikkritisches Publikum ansprechend sein.
- Schnittstellen zu schon vorhandener Software und in das Microsoft OfficePaket müssen vorhanden sein.

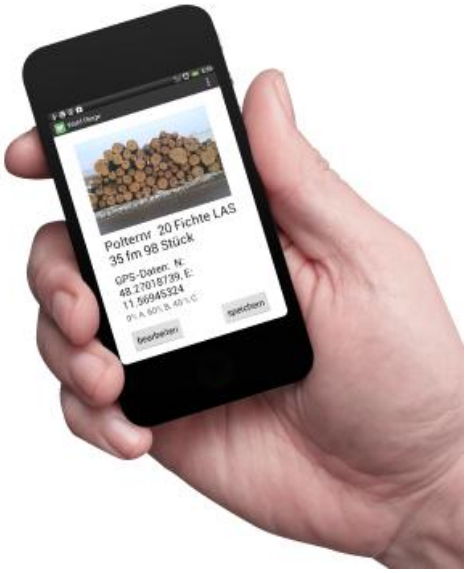


Abbildung 1: Smartphones als handliche Dateneingabegeräte in der Forstwirtschaft

2 Die App WaldFliege

WaldFliege[47] bietet ein Aufnahmeformular für Holzpolter (Holzstapel) die nach der Holzernte und -rückung an der Waldstraße bereitliegen, um dem Käufer übergeben zu werden oder von einem Frächter abgeholt und zum verarbeitenden Werk gefahren zu werden. Der Zeitraum zwischen dem Einschlag und der Ankunft im Werk ist ein kritischer Erfolgsfaktor in der Holzernte. Zu lange gelagertes Holz verliert an Wert. Mit WaldFliege kann der Rücker oder Förster die Daten der Holzpolter aufnehmen und zusammen mit GPS-Daten und Fotos in der Datenbank speichern.

47 http://www.inforst.de/de/waldfliege.html, Stand 08.Mai 2015

Abbildung 2: Screenshot der App WaldFliege – Alle Daten zum Polter auf einen Blick

Diese Daten kann er dann als Holzliste an den Kunden schicken oder aus der App heraus einen Fuhrauftrag als .gpx-Datei an den Frächter verschicken. Diese .gpx-Datei kann sich der Empfänger dann in ein Navigationsgerät einspielen oder in einer Internet-Anwendung, wie etwa dem kostenfreien Bayern-Atlas ansehen und ausdrucken. Forstwirtschaftliche Vereinigungen können zusätzlich die Daten ihrer Waldbesitzer in die App importieren. Dann können sie die Holzmengen mit dem Besitzerdaten verbinden und die Waldbesitzer bei Problemen aus der App heraus anrufen.

Die folgende Abbildung 3 zeigt die Produktbeschreibung der App Waldfliege.

Funktion	**WaldFliege**
Lauffähig ab Android-Version	2.3
Angepasst für Tablets	ja
Ermitteln und Speicher von GPS-Daten	ja
Aufnahme von Polterfotos und Speicherung in der Datenbank	bis zu 5 Fotos
Eingabe von Stückzahl und Schätzmaß	ja
Maximale Anzahl gespeicherter Polter /Einzelstämme	ohne Begrenzung
Manuell einstellbare Baumarten, Volumeneinheiten und Holzsorten	ja
Anlage von Hieben und Losen	ja

Anzeige der aufgenommen Holzmenge im Vergleich zur Sollmenge	ja
Einzelstammaufnahme	ja
Heilbronner Sortierung	ja
Sektionsverfahren	ja
Umrechnung innerhalb der Volumeneinheiten mit einstellbaren Einheiten und Faktoren	ja
Einheiten und Faktoren Export für Microsoft Excel	ja
Export als .xml-Datei im ELDAT-Format	ja
Export per E-Mail / USB	ja
Import von bis zu 2000 Waldbesitzerdaten	ja
Anrufunktion (Waldbesitzer anrufen)	ja

Abbildung 3: Produktbeschreibung Waldfliege

3 Die App WaldKarte

WaldKarte[48] ist eine Offline-Karten-App basierend auf Open-Street-Map-Karten und der Library „mapsforge“ der FU Berlin. WaldKarte interagiert mit WaldFliege, wenn beide Apps installiert sind. Es können aber auch beide Apps einzeln genutzt werden. Polterpunkte mit der Polternummer können mit Hilfe des GPS-Signals auf die Karte gesetzt werden. Man kann aber auch, wenn kein GPS verfügbar ist oder der Nutzer nicht vor Ort ist, den Lagerort auch durch Tippen mit dem Finger auf die Karte bestimmen. Auch das nachträgliche Verschieben von Polterpunkten mit den Fingern ist möglich. Aus der App heraus kann man die Karte als .jpg oder .png-Datei an die Geschäftspartner schicken. WaldKarte eignet sich auch für den Austausch von Treffpunkten.

48 http://www.androidpit.de/app/de.inforst.waldkarte, Stand 08.Mai 2015

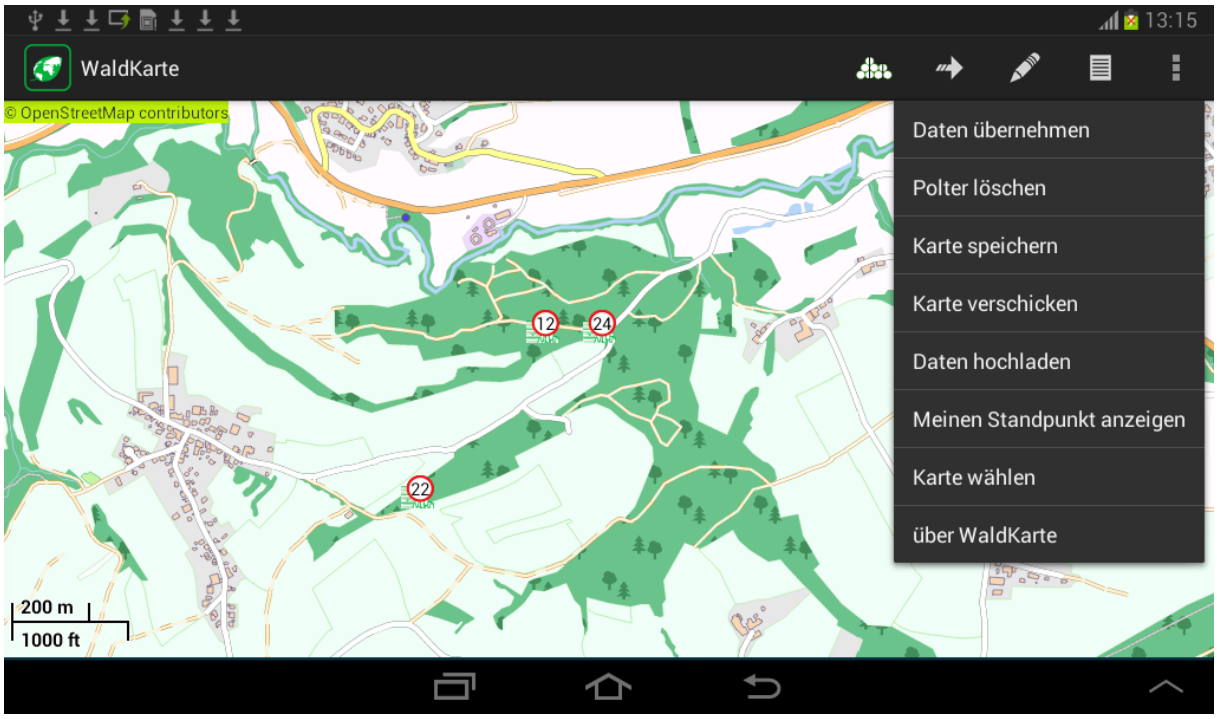


Abbildung 4: Screenshot der App WaldKarte – Anzeige der Position des Holzpolters mit Potlernummer

Die folgende Abbildung 5 zeigt die Produktbeschreibung der App Waldkarte.

Funktion	WaldKarte
Lauffähig ab Android-Version	2.3
Angepasst für Tablets	ja
Ermitteln und Speicher von GPS-Daten	ja
Darstellung der Polter in einer Offline-Karte	Ja
Manuelles Verschieben der Polter mit dem Finger	ja
Manuelles Setzen von Poltern durch Tippen auf die Karte	ja
Manuell einstellbare Baumarten, Volumeneinheiten und Holzsorten	ja
Darstellung der Polter aus WaldFliege, Bearbeitung der Polter in WaldFliege	nur bei Kauf von WaldFliege
Exportieren der Holzliste	nur über WaldFliege
Eingabe von Polternummer und Bemerkung	ja
Ermitteln und Anzeigen des Standorts	ja
Navigation	ja
Liste der gespeicherten Holzpolter	ja
Löschen von Poltern	ja

Abbildung 5: Produktbeschreibung Waldfliege

4 Multikoptereinsätze in der Forstwirtschaft

Die Auswertung photogrammetrischer Luftbilder hat in der Forstwissenschaft eine lange Tradition. Als Grundlage dienen Bilder von Befliegungen mit unbemannten Leichtflugzeugen (Unmanned Aerial Vehicle - UAV-Systemen). Baumarten, Baumhöhen, Bestockungsdichten und Gesundheitszustand der Bäume lassen sich aus solchen Bildern ableiten. Der Einsatz von UAV-Systemen bieten einen Vorteil, wenn eine lokal oder zeitlich begrenzte Befliegung gewünscht wird. Dies ist zum Beispiel der Fall:

- Beim Einsatz von Seilkränen zur Holzernte im Gebirge. Hier können Luftbilder die Trassenführung erleichtern.
- Zur Vorbeugung von Borkenkäfer-Befall, hier kann durch Kartierung von Kronenschäden das Monitoring in den besonders gefährdeten Gebieten intensiviert werden.
- Nach Sturmwurf: Hier erleichtert das Luftbild die Vermessung der Schadgebiete und die Planung der Aufarbeitung. Dadurch kann die nach Schadereignissen oft auftretende flächige Befahrung, die den Boden nachhaltig schädigt, vermieden werden.

Inforst bietet in Kooperation mit einem Ingenieurbüro für Multikoptervermessung Drohneneinsätze für die Forstwirtschaft an. Ziel ist momentan die Entwicklung einer Anwendung, mit der die Daten der Multikopter direkt auf der Offline-Karte WaldKarte dargestellt werden und die Punkte im Gelände so aufgesucht werden können.

Smart City auf Basis von 3D

Dr. Heino Rudolf, M.O.S.S. Computer Grafik Systeme GmbH, Buchenstr. 16b , 01097 Dresden, E-Mail: hrudolf@moss.de

Abstract

3D building models are currently available or in production for many urban arias. So we have the requirement to use this data for practical solutions.

M.O.S.S: developed a platform to generate and manage 3D city models. This platform offers opportunities for practical uses of the data in thematic applications. In cooperation with specialist partners applications are integrated in the platform and the city model is enriched with the results.

It is shown that the usage of 3D building models leads to a new quality in the thematic questions, e.g. the calculation of solar radiation and clouding, simulation of environmental phenomenon like noise, air pollutions or flood and the calculation of heat demand in buildings.

Zusammenfassung

3D-Stadtmodelle sind mittlerweile für viele urbane Gebiete vorhanden oder werden aktuell erzeugt. Damit steht mehr denn je die Frage: Was können wir mit diesen Daten anfangen, wie können wir sie nutzen?

M.O.S.S. Computer Grafik Systeme GmbH bringt sich in europäische und nationale Forschungsprojekte ein und befasst sich in diesem Zusammenhang damit, eine Plattform für das Daten- und Prozessmanagement zu entwickeln. Folgende Projekte sind betroffen:

- i-Scope (interoperable Smart City services through an open platform for urban ecosystems): Im Rahmen dieses Projekts wurde die Plattform erstmalig entwickelt und für verschiedene fachspezifische Services angewandt und getestet.
- SimStadt (Energiesimulation von Stadtquartieren): Erweiterung der Plattform um Wärmebedarfsberechnungen und zur Planung der Wärmeversorgung in Stadtgebieten.
- IQmulus (IQmulus-318787 – A High-volume Fusion and Analysis Platform for Geospatial Point Clouds, Coverages and Volumetric Data Sets): Programme zur Änderungserkennung von Objekten werden in die Plattform integriert.

(Links auf die Projektseiten: siehe Quellenverzeichnis)

Gemeinsam mit Fachpartnern werden Fachapplikationen in das System eingebunden, und das 3D-Stadtmodell wird um die Ergebnisse der Fachverfahren angereichert. Es zeigt sich, dass viele Fachthemen durch die Nutzung von 3D-Gebäudedaten eine völlig neue Qualität erreichen können, z. B. Berechnung von Sonneneinstrahlungen und Verschattungen, Simulation von Umweltphänomenen (wie Lärm- und Luftschadstoffausbreitungen, Hochwasser), Errechnung des Wärmebedarfs von Gebäuden.

Als Anbieter von Geographischen Informationssystemen fasst M.O.S.S. die in den genannten Projekten entwickelten Technologien zusammen, um eine Plattform zur Erzeugung, Verwaltung und fachlichen Anreicherung der 3D-Stadtmodelle anzubieten. Ziel ist dabei ein durchgängiges Datenmanagement sowie eine Dienste basierte Bereitstellung von Daten (als WFS) und Prozessen (als WPS). Die Plattform enthält Services zu folgenden Bausteinen:

- Erzeugung von 3D-Stadtmodellen (Produktion und Qualitätssicherung)
- Datenmanagement (Datenverwaltung, Importe, Exporte)
- Anbinden von Fachapplikationen
- Workflowmanagement
- Laufendhaltung des 3D-Stadtmodells.

Im Folgenden werden diese Bausteine einzeln vorgestellt.

1 Erzeugung von 3D-Stadtmodellen

1.1 Gebäudeproduktion

Die Erzeugung von 3D-Gebäudemodellen ist inzwischen kein Geheimnis mehr. Auf Basis der Gebäudegrundrisse, des Digitalen Oberflächenmodells (DOM) und des Digitalen Geländemodells (DGM) erkennen Softwareprodukte die Gebäudedachflächen und können daraus Standarddachformen zuordnen. Als Ergebnis liegen die Gebäude im CityGML-Format vor, mit dem „Level of Detail“ (LoD) 2. Gebäude, deren Dachform nicht erkannt werden konnte, werden im LoD 1 als sogenanntes „Klötzchenmodell“ erzeugt.

Abb. 1: Digitales Geländemodell, Gebäudegrundrisse

Im Rahmen der hier behandelten Projekte wurden neuartige Verfahren entwickelt, die es ermöglichen, dass der Prozess der 3D-Gebäudeproduktion als Web Processing Service (WPS) bereitgestellt wird. Die Plattform ermöglicht es, Ausgangsdaten hochzuladen, den Produktionsprozess auszuführen und die Ergebnisse im Format CityGML wieder herunterzuladen.

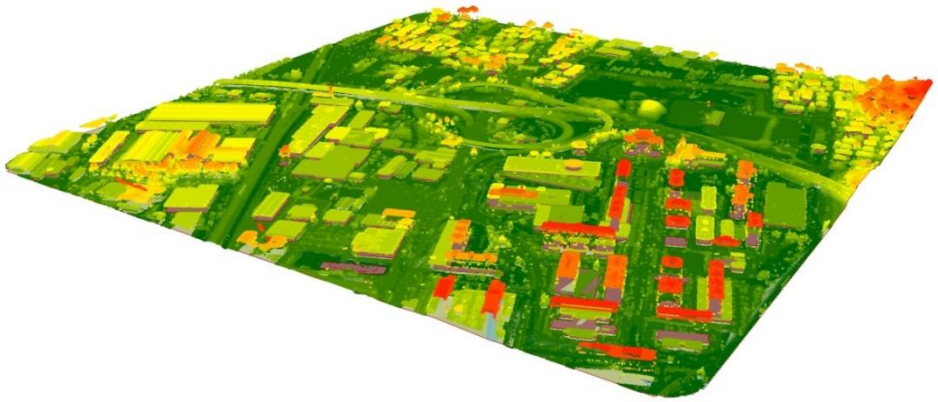


Abb. 2: Digitales Oberflächenmodell

1.2 Qualitätssicherung

Je nach Qualität der Ausgangsdaten sind manuelle Nacharbeiten der automatisch generierten Gebäude erforderlich. Dazu nutzen wir die Standard-Software Trimble SketchUp. Es war notwendig, SketchUp mittels eines eigenen Plug-Ins zu erweitern, um den verlustfreien Datenkreislauf CityGML–SketchUp–CityGML bidirektional zu gewährleisten, die generierte Gebäude-Identifikationsnummer (ID) zu erhalten (um das Aktualisieren der Daten im CityGML-Datenpool sicherzustellen) und die Erfassung von Sachdaten zum Gebäude zu ermöglichen.

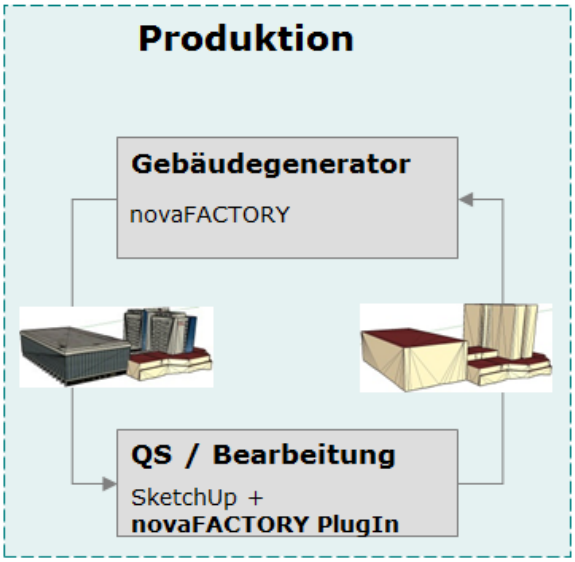


Abb. 3: Gebäudeproduktion und Qualitätssicherung

2 Datenmanagement

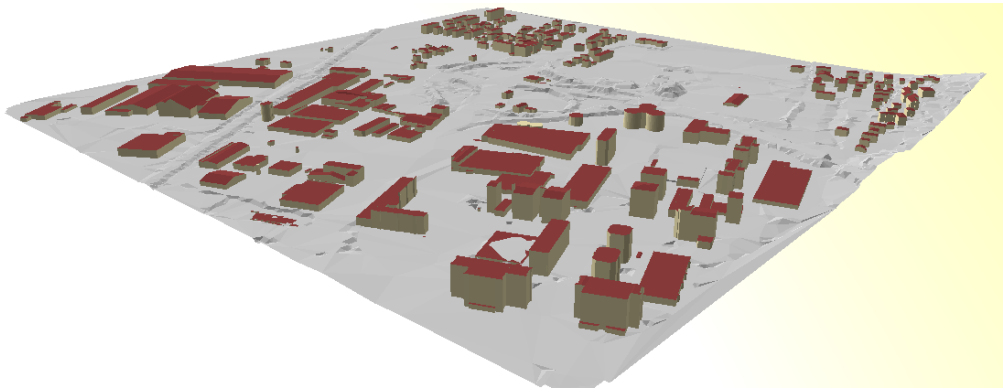


Abb. 4: 3D-Gebäudemodell

Basis für die Datenverwaltung des 3D Gebäudemodells bildet die Open Source Implementierung 3D City DB (Link siehe Quellenverzeichnis). Neben der Datenhaltung ist die Bereitstellung von Import- und Export-Services notwendig, um einerseits die Ergebnisse der Applikationen aufzunehmen und andererseits die Daten den Applikationen zur Verfügung zu stellen. Beim Import ist immer darauf zu achten, dass der Datenbestand sequentiell korrigiert und nicht komplett überschrieben wird.

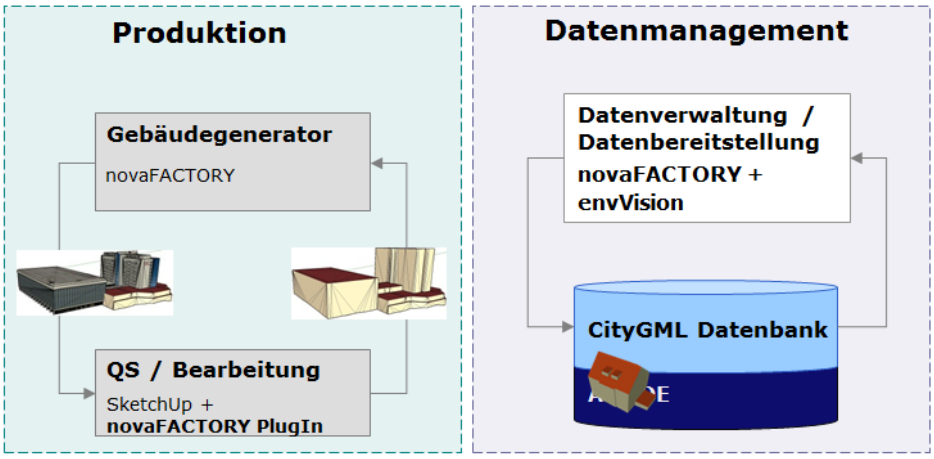


Abb. 5: Datenmanagement

3 Anbinden von Fachapplikationen

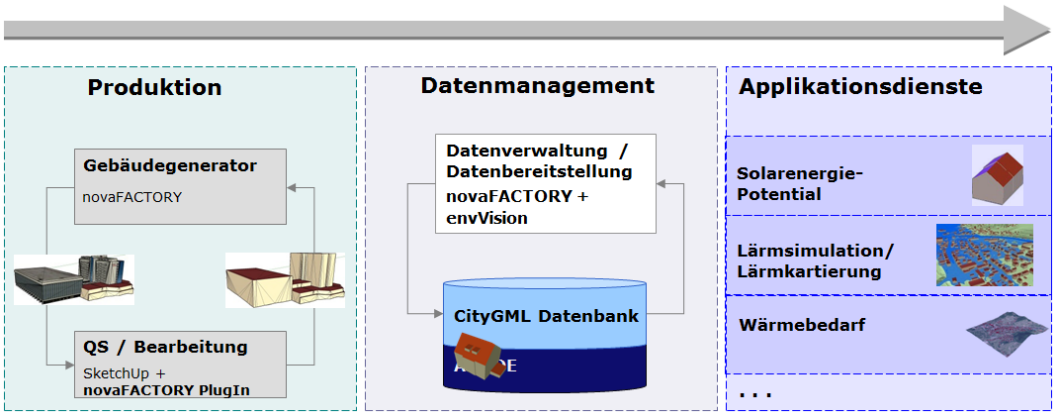


Abb. 6: Anbinden von Fachapplikationen

CityGML sieht Möglichkeiten vor, das Modell um Fachdaten zu erweitern. Einerseits ist das über sogenannte generische Objekte und Attribute möglich, andererseits können ganze Datenstrukturen kreiert werden (als sogenannte Application Domain Extension – ADE). Die Plattform ist in der Lage, die Daten beliebiger ADE zu importieren, diese als generische Objekte oder Attribute zu verwalten und sie verlustfrei wieder in der Form der ADE zu exportieren.

Damit werden die Fachdaten ergänzend (oftmals sprechen wir auch von „anreichernd“) in das Datenmanagement involviert.

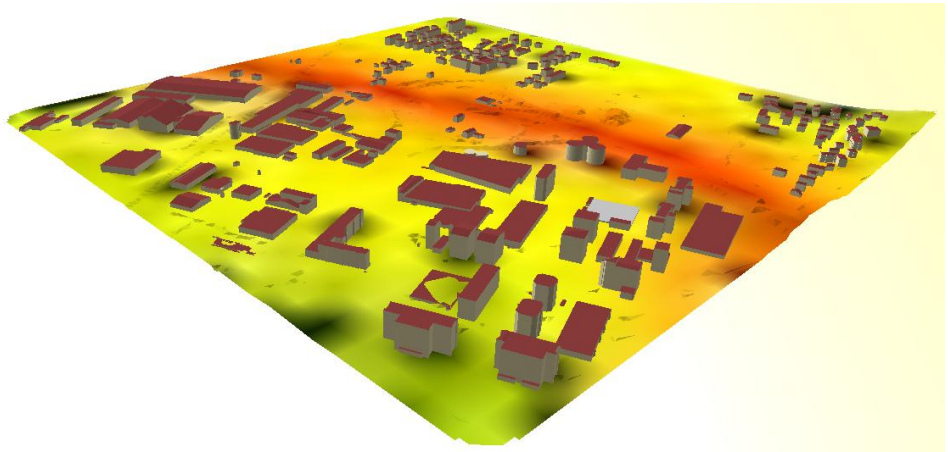


Abb. 7: Ergebnisse von Fachapplikationen, Beispiel Lärmkartierung

In den oben genannten Forschungsprojekten wurden verschiedene Fachthemen angebunden:

- Berechnung des Solarenergie-Potenzials
- Lärmausbreitungen
- Berechnung des Wärmebedarfs von Gebäuden
- Routing-Systeme.

Hier zeigt sich ein wesentliches Qualitätsmerkmal der entwickelten Plattform:

Ein und derselbe Bestand an 3D-Basisdaten wird auf diese Weise für unterschiedliche Fachanwendungen zur Verfügung gestellt und nutzbar gemacht. Mehr noch, dieser Datenbestand wird sukzessive um Ergebnisse von Fachverfahren erweitert. Und die Datenhaltung ist konsistent, begründet durch einen Themen übergreifenden Modellansatz zur Datenverwaltung.

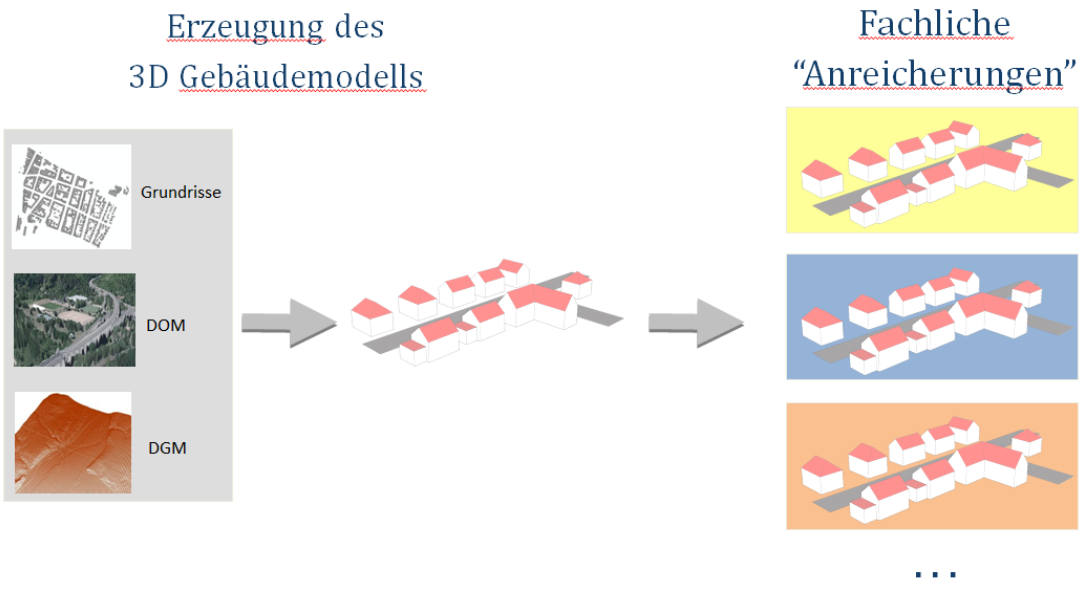


Abb. 8: „Anreicherung" mit Fachdaten

(Die unterschiedlichen Farben sollen Anreicherungen um Ergebnisdaten verschiedener Fachverfahren symbolisieren.)

4 Workflowmanagement

Sind jetzt Lösungen für das Datenmanagement in der Plattform verfügbar, so steht parallel die Aufgabe, die (fachlichen) Services im Sinne von Workflows zu managen. Es soll möglich werden, beliebige fachliche Prozesse anzustoßen, sie zu Prozessketten (Workflows) zusammenzustellen und diese Prozessketten abzuarbeiten.

Dazu wurde eine Methodik ausgearbeitet, die unter Verwendung von Petri-Netzen [PETRI 1962] gestattet, einzelne Prozesse über Ports zusammenzuschließen. Die Struktur des Prozess-Graphen kann beliebig sein (in Reihe geschaltete und parallele Prozesse) – immer, wenn ein Ausgangsport nach erfolgreicher Abarbeitung des Prozesses seine „Bereitschaft“ an den folgenden Eingangsport übergibt und alle Eingangsports eines Prozesses „bereit“ sind, wird der (Folge-) Prozess angestoßen.

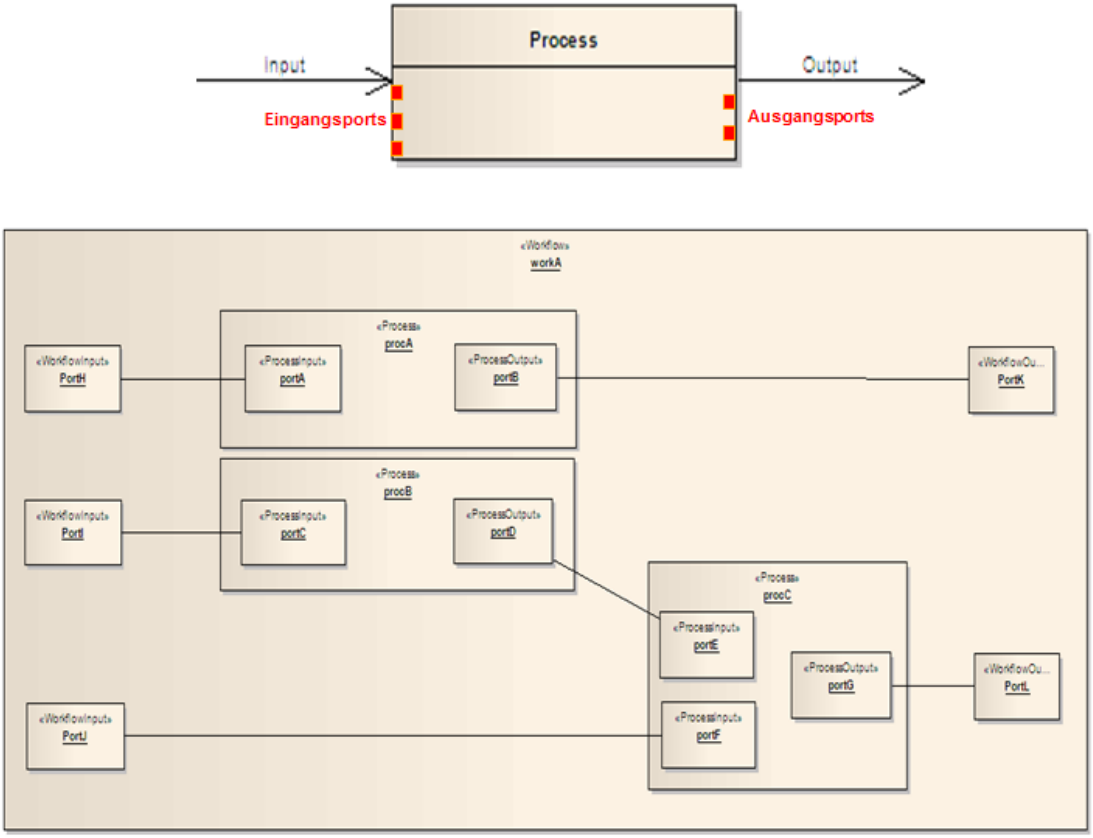


Abb. 9: Prozessausführung und Steuerung

Abbildung 9 stellt dieses Prinzip schematisch dar. Die Workflows werden in einer XML-Datei beschrieben, die dann den „Prozessketten-Manager“ steuert.

Damit ist die Plattform in der Lage, beliebige Workflows zu verarbeiten und ablaufen zu lassen. Konkret umgesetzt wurde das im Projekt IQmulus für verschiedene Objekterkennungsroutinen, insbesondere auch zur Erkennung von Änderungen des Gebäudebestands (siehe Abschnitt 5).

5 Laufendhaltung der Daten

Mit der zunehmenden Verfügbarkeit strukturiert erfasster bzw. abgeleiteter Daten gelangt immer stärker die Aktualisierung und Laufendhaltung der 3D Stadtmodelle in den Fokus der Datenmanager. Verfahren, die auf einer regelmäßigen, weitestgehend automatisierten, vollständigen Neuerfassung basieren, sind unbrauchbar, da diese die erzeugten und mit den Gebäudedaten verknüpften Zusatz- und Fachinformationen verwerfen. Intelligentere Prozesse werden somit erforderlich, einerseits um Kosten der Datenfortführung zu sparen, andererseits um die (teilweise teuren Fach-) Daten weiternutzen zu können.

Aktuell werden verschiedene Softwareprodukte bzgl. der Erkennung von Gebäuden evaluiert. Diesbezüglich wurde ein Workflow konzipiert und implementiert, mit folgenden Prozessschritten:

- Export der Ausgangsdaten
- Gebäudeerkennung unter Nutzung des „BuildingFinder“ der Fa. tridicon
- Vergleich der erkannten Gebäude mit dem Bestand
- Import der Ergebnisse: neue, geänderte, fehlende Gebäude.

Als Ergebnis werden die angenommenen neuen, geänderten und fehlenden Gebäude attributiv markiert. Damit kann im Anschluss eine gezielte manuelle Aktualisierung des 3D-Stadtmodells vorgenommen werden.

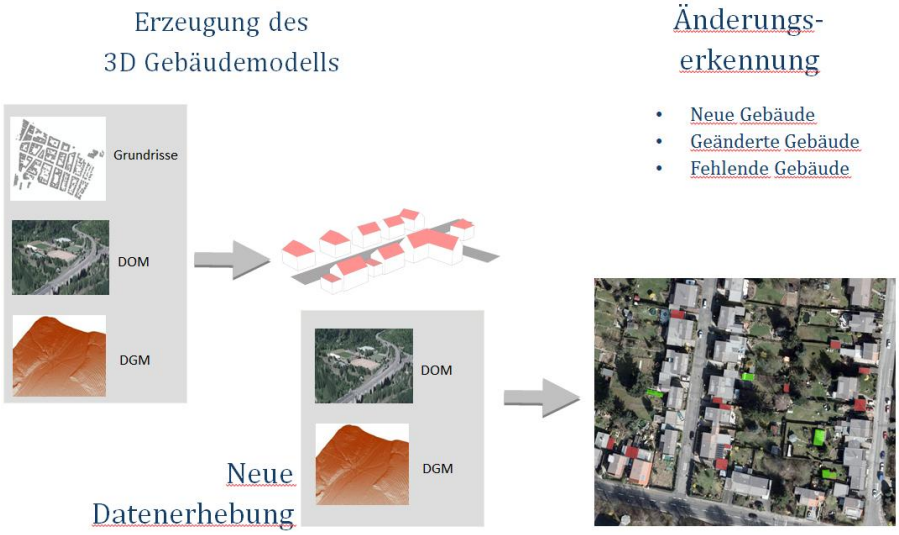


Abb. 10: Laufendhaltung der Daten: Änderungserkennung

6 Fazit und Ausblick

Für den Umgang mit 3D-Stadtmodellen steht eine Plattform zur Verfügung, die sowohl das Datenmanagement (von der Erzeugung über die Verwaltung und Bereitstellung bis hin zur fachlichen „Anreicherung“) als auch das Prozessmanagement (zur Abarbeitung beliebiger fachlicher Workflows und Prozesse) ermöglicht.

Sollte Interesse an der Erzeugung von 3D-Stadtmodellen bestehen, so können die Services getestet werden, insbesondere die zur Produktion der 3D-Gebäudedaten. Eine Plattform zur Verwaltung der 3D-Daten steht zur Verfügung und kann genutzt werden.

Programmseitig werden die sich noch im Prototypstadium befindlichen Fachanbindungen „marktreif“ implementiert. Eine Erweiterung um weitere Fachthemen ist vorgesehen, Fachpartner sind dazu willkommen.

Der Prozess der Änderungserkennung wird weiter qualifiziert und auf andere Objektklassen (wie z.B. Gewässer, Landnutzung, Wegenetz) ausgedehnt.

7 Quellenverzeichnis

I-SCOPE, i-Scope - interoperable Smart City services through an open platform for urban ecosystems, http://www.iscopeproject.net/, abgerufen: 12.01.2015

SIMSTADT, SimStadt – Energiesimulation von Stadtquartieren, http://www.simstadt.eu/de/index.html, abgerufen: 12.01.2015

IQMULUS, IQmulus-318787 – A High-volume Fusion and Analysis Platform for Geospatial Point Clouds, Coverages and Volumetric Data Sets, http://iqmulus.eu/, abgerufen: 12.01.2015

3D CITY DB, 3D City Database for CityGML, http://www.3dcitydb.org/3dcitydb/fileadmin/downloaddata/3DCityDB-Documentation-Addendum-v2_1.pdf, abgerufen 15.01.2015

CITYGML, http://www.citygml.org/, abgerufen: 12.01.2015

PETRI, C.A., Kommunikation mit Automaten. 1962, Institut für instrumentelle Mathematik der Universität Bonn