Geomatik - Kolloquium Sommersemester 2013

Bettina Schnor
Institute of Computer Science
University of Potsdam
schnor@cs.uni-potsdam.de

und

Joachim Wächter
Helmholtz-Zentrum Potsdam
Deutsches GeoForschungsZentrum GFZ
joachim.waechter@gfz-potsdam.de

Berichte Institut für Informatik, UP, ISSN 0946-7580, TR-2013-2

Zusammenfassung

Die dynamischen Veränderungen des Systems Erde verursachen große globale Herausforderungen für die gesellschaftliche und wirtschaftliche Entwicklung. Das Monitoring sowie die Analyse und Auswertung von Geoprozessen ist ein hochrelevantes Forschungsfeld, welches die Entwicklung von flexiblen Monitoringsystemen, leistungsfähigen Prozessierungs- und Analysekomponenten, aber auch von Werkzeugen zur Entscheidungsunterstützung und Planung erfordert. Der engen Verbindung und effektiven Vernetzung von Informatik und Geowissenschaften kommt bei der Bewältigung dieser Herausforderungen eine Schlüsselrolle zu. Im Forschungsseminar werden sowohl aktuelle Problemstellungen und Anwendungsfelder aus den Geowissenschaften als auch mögliche Lösungsansätze aus der Informatik diskutiert.

Inhaltsverzeichnis

1 FOSSLAB - Free and Open Source Laboratory im GFZ
Joachim Wächter, Martin Hammitzsch, GeoForschungsZentrum 1

2 Energy Management of Android
Bettina Schnor, Institut für Informatik 3

3 Indoor-Lokalisierung
Sebastian Fudickar, Institut für Informatik 5

4 Scientific Workflows: Dealing with a World full of Services and their Varying Compositions
Tiziana Margaria-Steffen, Anna-Lena Lamprecht, Institut für Informatik 6

5 Sensor-Integrationsplattform - Connector-Plugins und Mapping zu O&M / SWE-Interfaces
Stephan Gensch, GeoForschungsZentrum 8

6 Maschinelles Lernen zur Modellbildung in den Naturwissenschaften
Niels Landwehr, Institut für Informatik 9

7 Geomatik & Human-Computer Interfaces
Peter Löwe, Léonie Schäfer, GeoForschungsZentrum 11

8 Big Data: Anwendungsfelder in der Wissenschaft und Technologien
André Luckow, Institut für Informatik, BMW Group 12

9 Interaktive Routenvisualisierung auf einem taktilen Flächendisplay
Ulrike Lucke, Mihail Ivanchev, Institut für Informatik 13

10 Information Modelling in the Geological Sciences
Jens Klump, CeGIT, German Research Centre for Geosciences GFZ 14

11 Projektportale für raumbezogene Daten - Architektur und Entwicklung
Vivien Stender, Matthias Schroeder, GeoForschungsZentrum 16

1 FOSSLAB - Free and Open Source Laboratory im GFZ

Joachim Wächter, Martin Hammitzsch, GeoForschungsZentrum

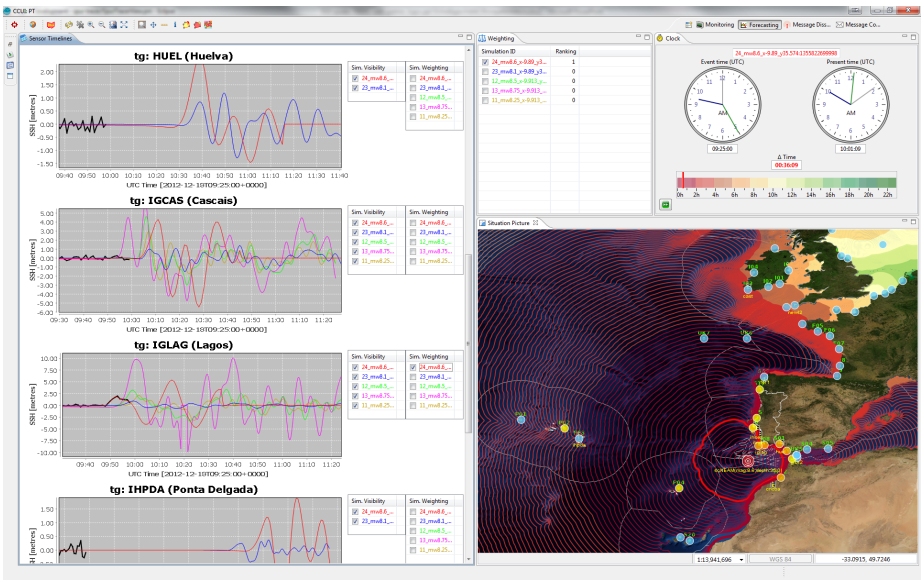


Abbildung 1: Das Command and Control User Interface (CCUI) des TRIDEC Systems ermöglicht den Vergleich von gemessenen Wasserständen (schwarze Linien) mit berechneten Wasserständen aus verschiedenen Simulationen (bunte Linien) in einem Diagramm pro Pegelstation. Simulationen können gewertet werden, um die passendste auszuwählen und auf einer Karte die dazugehörige Tsunamiwellenausbreitung anzuzeigen

FOSSLAB bietet eine Plattform für die Entwicklung von Open Source Software Projekten, die einen Bezug zur Geoinformatik haben. Alle bisherigen Entwicklungen in FOSSLAB sind im Rahmen von Projekten des Deutschen GeoForschungsZentrums GFZ entstanden und basieren zum gröten Teil auf FOSS-Komponenten und Industrie-Standards. Schwerpunkt war und ist die Nutzung von Open Source Technologien, die ausgereifte Modelle für Softwaresystemarchitekturen unterstützen. FOSSLAB soll sich künftig als Plattform für die Entwicklung von FOSS-Projekten etablieren, die mit Geodaten arbeiten. Ihre Mitglieder sollen nicht nur die Möglichkeit haben Ereignisse über FOSSLAB zu sichern und zur Nachnutzung anzubieten, sondern sollen auch auf Ergebnisse aus früheren und laufenden Projektaktivitäten zugreifen, diese wiederverwenden und weiterentwickeln.

Bereits erarbeitete Ergebnisse sollen so die weitere Entwicklung ermöglichen und die Zusammenarbeit einer heterogenen Gemeinschaft von Wissenschaftlern, Entwicklern, Anwendern und anderen Interessengruppen stärken. FOSSLAB bildet dabei einen Schirm für eine Reihe von in Beziehung stehenden Geoinformatikaktivitäten, wie Softwaresystemdesign, Implementierung, Training, und Dokumentation von Best Practices für Erfahrungen und Ergebnisse bei der Arbeit mit SDIs, GIS, der räumlichen Verarbeitung auf Clustern und in der Cloud, Geomatik und vielem mehr.

Der Fokus der Veranstaltung liegt dabei auf einem aktuellen Projekt, das unter anderem ein komponenten- basiertes Softwaresystem für die Tsunami-Frühwarnung implementiert und dessen Entwicklungsteam im Rahmen von FOSSLAB Ergebnisse sichert, weiterentwickelt und Interessengruppen zur Nachnutzung und Weiterentwicklung zur Verfügung stellt.

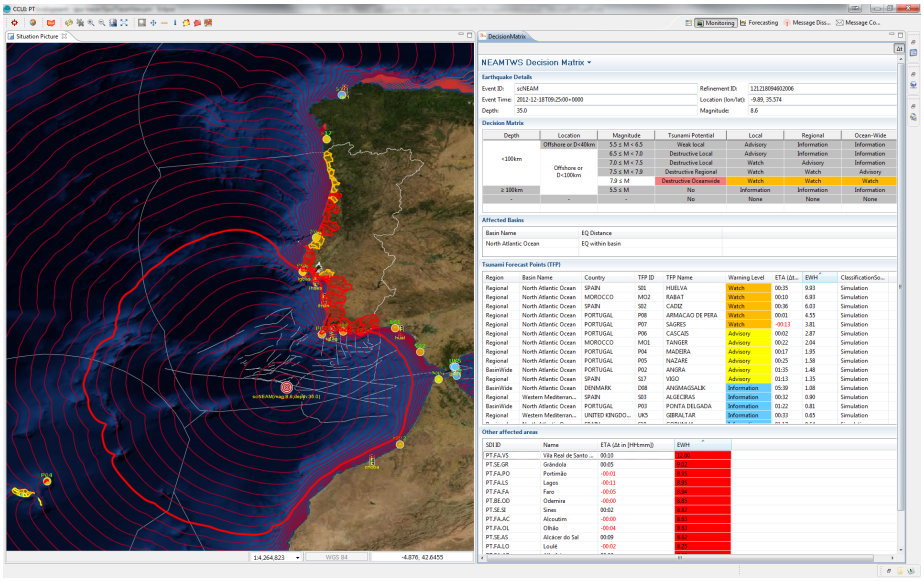


Abbildung 2: Basierend auf der ausgewählten Simulation und einer Entscheidungstabelle ermöglicht das TRIDEC System die Klassifizierung betroffener Gebiete und Referenzpunkte über die berechneten Ankunftszeiten und Wellenhöhen

2 Energy Management of Android

Bettina Schnor, Institut für Informatik

Das Open-Source Betriebssystem für mobile Geräte Android ist mit einem weltweiter Marktanteil von 70% bei Smartphones im vierten Quartal 2012 das meist genutzte Betriebssystem für Smartphones, Tablets und Uhren. Inzwischen wird es sogar in anderen Geräten unseres Alltags wie z.B. Waschmaschinen integriert und hat sich somit seit der Übernahme der Firma Android Inc. durch Google im Jahr 2005 und der ersten Veröffentlichung im Jahr 2008 rasant im Markt als defacto Standard etabliert. Der Android-Kernel (dessen grobe Architektur in Abbildung 3 dargestellt ist) basiert auf dem Linux Kernel (Version 3.0.x) und ergänzt Geräte- bzw. Chip-spezifische Treiber. Zusätzlich wurde eine reduzierte und optimierte libc Variante (die bionic libc) integriert. Applikationen laufen in der sogenannten Dalvik Virtual Machine - eine Weiterentwicklung von Apache Harmony, die für niedrigen Speicherverbrauch optimiert ist und Applikationen (deren Bytecode in einem Dex-File kompiliert wird) ausführt.

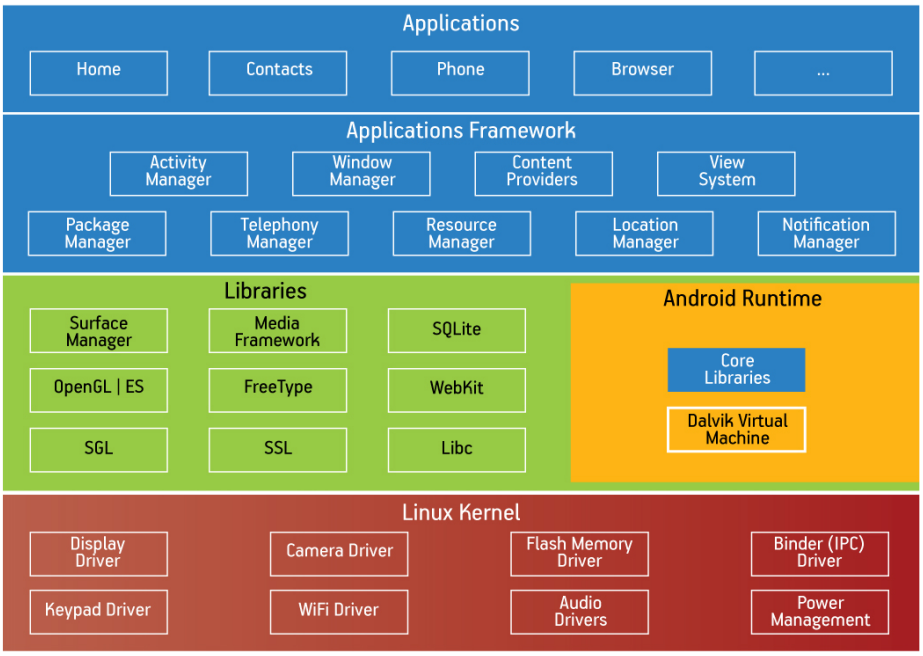


Abbildung 3: Android-Architektur (Quelle: `http://www.techdesignforums.com`)

Dieser Vortrag behandelt eine weitere Spezialität des Android Betriebssystems: Das Android-Energiemanagement (der Android Version 4.3) - eine essentielle Komponente für den Einsatz mobiler Geräte. Wie der Name nahe legt, ist die Kernaufgabe des Energiemanagements den kritischen Energieverbrauch der Software und der genutzten Komponenten zu kontrollieren, um akzeptable Gerätelaufzeiten von mindestens 16 Stunden zu gewährleisten. Dies wird hauptsachlich durch die frühe Deaktivierung von Komponenten erreicht. Dies unterscheidet sich signifikant von klassischen Desktop-Betriebssystemen, bei denen der default Betriebszustand „an“ ist und im Wesentlichen nur das komplette System mittels der Zustände „G1 = sleeping, S3 = suspend to RAM“

des sogenannten Advanced Configuration & Power Management (ACPI) deaktiviert werden kann.

Das Android-Energiemanagement geht hier folgenden Weg: Das System deaktiviert nicht benötigte HW-Komponenten oder reduziert ihren Energieverbrauch (z.B. durch herunter takten), solange diese nicht benötigt werden. Das System kann bei Bedarf durch Ereignisse (wie z.B. Nutzereingabe, Ablauf von Timern, Sensor- oder Netzwerkaktivität) in Kürze wieder reaktiviert werden.

Mittels sogenannter "Wake-Locks" belegen Applikationen bestimmte Ressourcen wie z. B. Bildschirm oder CPU und verhindern das Deaktivieren (bzw. Runterfahren oder Dimmen) dieser Ressourcen.

Wake-Locks sind dabei als Kernel-Mechanismus implementiert, der benötigte Komponenten aktiv hält, solange sie von mindestens einer Applikation reserviert sind. Dabei können Anwendungen verschiedene Wake Lock Typen (siehe Tabelle 1) anfordern. Da die CPU erst in den Standby-Zustand geschaltet wird, wenn kein Wake-Lock mehr aktiv ist, ist eine sorgfältige Programmierung essentiell, um Energie sparen zu können.

Type	CPU	Bildsch.	Tast.bel.
PARTIAL_WAKE_LOCK	an	aus	aus
SCREEN_DIM_WAKE_LOCK	an	gedimmt	aus
SCREEN_BRIGHT_WAKE_LOCK	an	hell	aus
FULL_WAKE_LOCK	an	hell	hell

Tabelle 1: Wake-Lock-Typen

Neben dem Wakelock Mechanismus beinhaltet das Energiemanagement auch eine Batterieüberwachung mit der Applikationen sich über Veränderungen des Batterie-Ladestandes und der Temperatur informieren lassen können bzw. diese Informationen abfragen können.

Somit bietet das Android-Energiemanagement eine umfangreiche Schnittstelle zur Programmierung energiebewusster Anwendungen und hat wesentlichen Anteil am Erfolg des Android-Betriebssystems.

3 Indoor-Lokalisierung

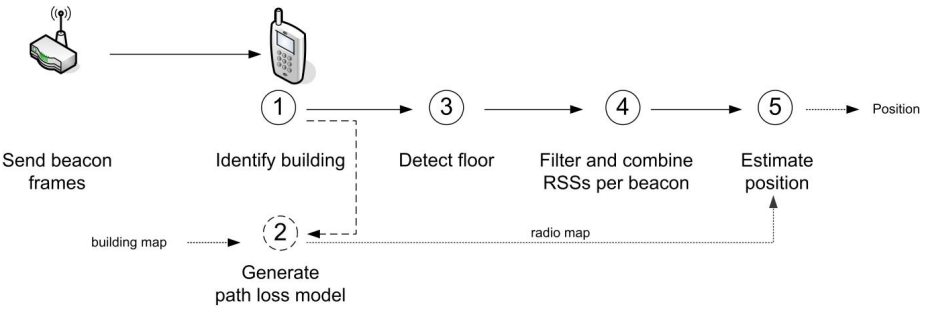
Sebastian Fudickar, Institut für Informatik

Smart phone applications (such as navigation, information or social-applications) can be enhanced with location awareness. While outdoor localization-technologies such as GPS or cell-tower communication are not applicable for indoor environments due to radio-signal shielding, the radio frequencies of 2.4 GHz (as used by WiFi) or Sub GHz (such as 868 MHz) are well suited for indoor localization. For indoor localization, beacons that regularly transmit messages are placed within buildings. Mobile devices estimate their current position from the received messages typically via the received signal strength indicator (RSSI).

While most RSSI based systems apply fingerprinting approaches, the proposed system instead applies a model based approach to overcome the laborious manual collection of RSSI fingerprints for each supported building. The proposed localization system therefore, generates a path loss model for the identified building. Received beacon-frames of surrounding beacons are used to identify the building and the current floor-level. Furthermore, after being filtered and merged they are used for position estimation on the identified floor by the generated path-loss model.

For the proposed system the optimal path loss-models, algorithms and parameter settings that localize most accurately were identified based on recordings via a simulator. The recordings covered all accessible areas of two (office style) buildings with long floors and rooms attached to them.

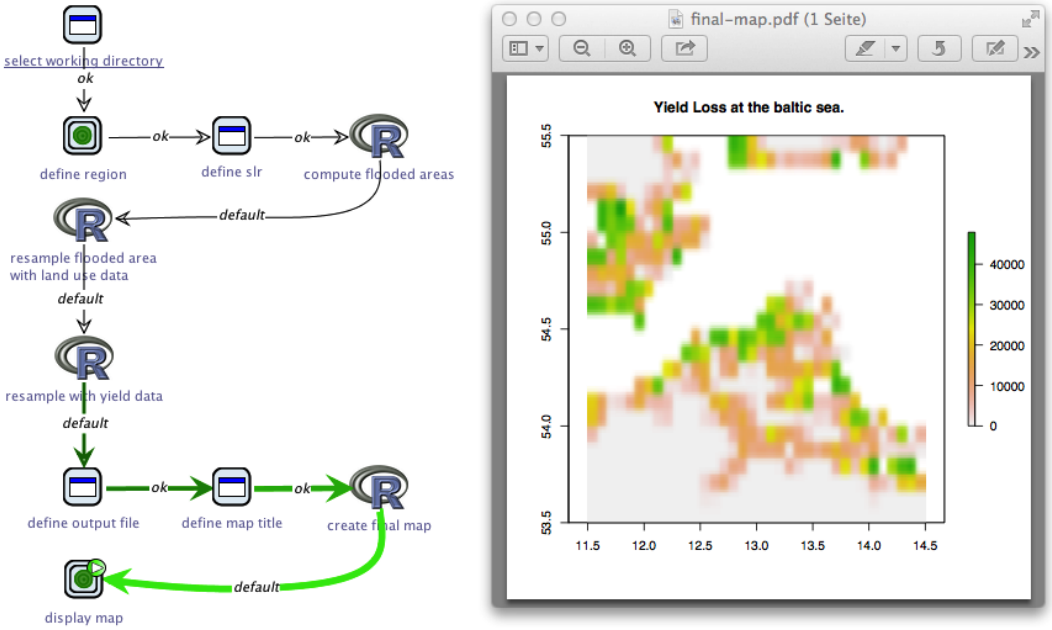
The resulting localization system can localizes peasants with a accurate median error distance below 5 meters for Sub GHz and below 6 meters for WiFi and achieves acceptable localization-error and floor-error rates in setups where beacons are deployed with a density of 0.008 beacons per m^2. Aside of the high accuracy, the system is well suited for continuous tracking due to its energy efficiency and its ability to adapt to previously unknown buildings (via a map-exchange mechanism).



References:
On the Comparability of Indoor Localization Systems Accuracy: Sebastian Fudickar, Sebastian Amend, Bettina Schnor. Proceedings of ISA '13, Orlando, USA, November 2013

4 Scientific Workflows: Dealing with a World full of Services and their Varying Compositions

Tiziana Margaria-Steffen, Anna-Lena Lamprecht, Institut für Informatik



Scientific application domains are characterized by the long-term availability of the basic computational components, but software systems for managing the actual workflows of the scientificday-to-day life have to deal with changing service interfaces and varying service compositions. Accordingly, the past decade has seen a lot of research on scientific workflow management,and service-oriented and model-based approaches have become the methods of choice for user-centric development of variant-rich workflows in many application domains. In particular, typical daily adaptations and modifications should be in the hands of the application expert. Our reserach addresses concepts and technologies to make this happen, and apply them to create tools and frameworks that offer automated support to model-driven service orientation.

The extreme model-driven approach that we implement in the jABC framework supports the agile management of processes and workflows in really huge and trulyheterogeneous application domains. The key concept is the use of different levels of abstraction, which leverages workflow design to a more conceptual level, enabling scientists (our end-users) to work with a world-wide distributed collection of tools and data using their own domain language, that is, without requiring technical knowledge about IT and programming in generaland about the used services in particular.

We have applied this approach in different scientific diciplines, including bioinformatics, computerlinguistics, geoinformatics and medicine. A comprehensive book on this topic, titled "Process Design for Natural Scientists: An Agile Model-DrivenApproach", will soon appear in Springer's CCIS series. It focuses on our experiences in the geo- and bioinformatics application domains and incorporates a number of student project reports from our course "Informatics for scientists"

References:

- Process Design for Natural Scientists: An Agile Model-DrivenApproach. A. Lamprecht, T. Margaria (editors). To appear in Springer's CCIS series.
- User-level workflow design. A bioinformatics perspective. A. Lamprecht. To appear in Springer's LNCS series.
- Scientific Workflows: eternal components, changing interfaces, varying compositions. A. Lamprecht, T. Margaria. 5th International Symposium on Leveraging Applications (ISoLA 2012), Springer LNCS, Volume 7609, 2012.
- Service Orientation: Conquering Complexity with XMDD. T. Margaria, B. Steffen. Conquering Complexity, Springer 2012.

5 Sensor-Integrationsplattform - Connector-Plugins und Mapping zu O&M / SWE-Interfaces

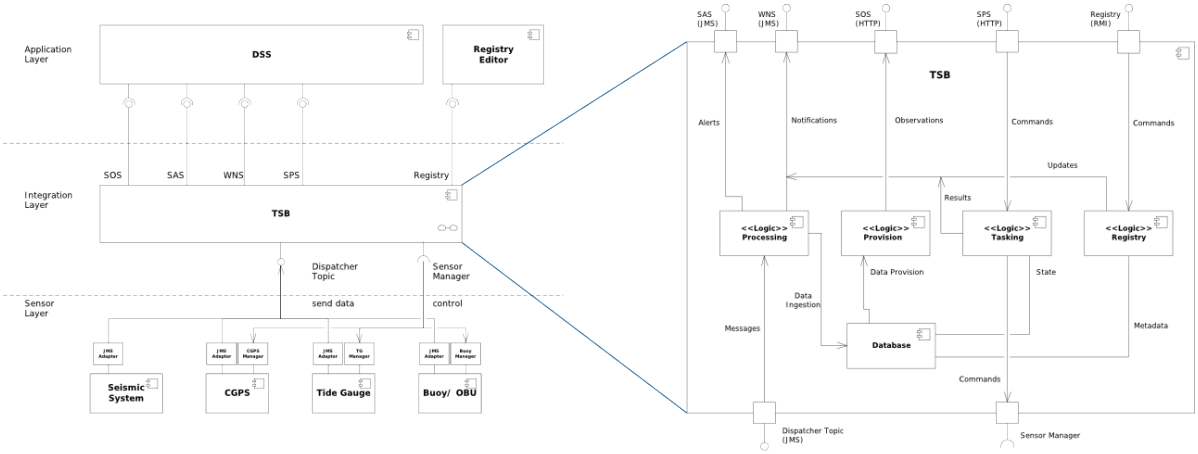
Stephan Gensch, GeoForschungsZentrum

One of the major challenges for the German Indonesian Tsunami Early Warning System (GITEWS) was to facilitate integration of a complex and heterogeneous sensor data infrastructure, being prepared for future modifications of both single sensors and sensor systems. Following the paradigm of a Service Oriented Architecture (SOA), the so-called Tsunami Service Bus (TSB) is an integration platform implementing OGC Sensor Web Enablement (SWE) standards and services. It is deployed as a Java EE application hosted within a JBoss Application Server.

The following SWE specifications have been implemented:

- Observations & Measurements (O&M): model for observations and measurements.
- Sensor Model Language (SensorML): model for describing sensor systems.
- Sensor Observation Service (SOS): service for obtaining sensor observations.
- Sensor Planning Service (SPS): service for tasking sensors.

This talk will present the concept of plug-ins for sensor data ingestion and the respective mappings to an internal Objects & Measurements data representation for seismic and cGPS sensor systems and tide gauge sensors. Emphasis is also laid on the standardized client interface for accessing sensor data and sensor Meta data management.

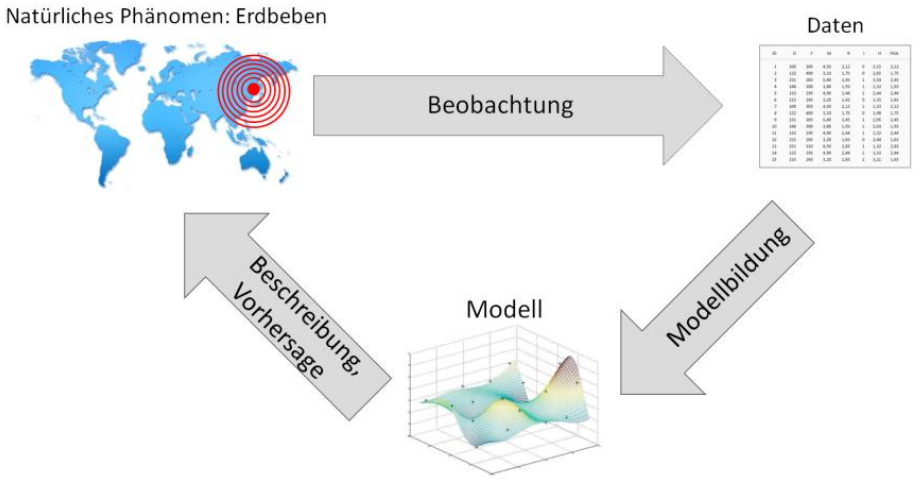


The Tsunami Service Bus is architecturally divided into sensor, integration and application layer.
Sensor Layer: *Sensors and sensor systems are located at the Sensor Layer, providing data for the warning system. Data providers from the sensor layer publish data onto a Java Messaging System (JMS) Dispatcher Topic, where it is then processed and stored.*
Integration Layer: *The integration layer consists of producer interfaces for the application layer (Sensor Web Enablement (SWE) services SOS, SAS, SPS, and WNS) and consumer interfaces for the sensor layer (Dispatcher Topic and Sensor Manager). The TSB internally implements Processing and Provision of sensor data, Sensor Tasking and the Sensor Registry, which is responsible for all sensor metadata management in GITEWS.*
Application Layer: *Client applications reside at the topmost layer, one of these being the Decision Support System (DSS), which aggregates the sensor data further to highly aggregated information products to assist the Chief Officer on Duty in his/her decision whether a tsunami warning should be disseminated. Other applications or GUIs can easily be developed to use the standardized interfaces.*

GFZ Helmholtz Centre Potsdam
CC BY
HELMHOLTZ ASSOCIATION

6 Maschinelles Lernen zur Modellbildung in den Naturwissenschaften

Niels Landwehr, Institut für Informatik



Die Analyse experimentell gewonnener Daten bildet eine Säule des Erkenntnisgewinns in den Naturwissenschaften. Sie ermöglicht die Bildung von Modellen, die natürliche Phänomene beschreiben; gewonnene Modelle werden häufig auf der Grundlage experimenteller Daten evaluiert. Datenanalyseverfahren aus dem Bereich der Informatik (maschinelles Lernen, Data Mining) werden in verschiedenen naturwissenschaftlichen Disziplinen eingesetzt, um Modelle natürlicher Systeme aus Beobachtungen abzuleiten. Ziel ist die Vorhersage des Systemverhaltens, Einsicht in Systemparameter sowie die vergleichende Evaluierung alternativer Modelle, die unterschiedliche wissenschaftliche Erklärungsansätze reflektieren. Allerdings basieren verbreitete methodische Werkzeuge des maschinellen Lernens auf Unabhängigkeits- und Verteilungsannahmen, die experimentell gewonnene Daten in verschiedener Weise verletzen können. So werden die Verteilungseigenschaften von Experimentaldaten durch die Wahl von Beobachtungsprotokollen und experimentelle Parameter bestimmt; die gewonnenen Daten liefern ein Abbild der Wirklichkeit, das durch den Prozess der Messung geprägt ist.

Die im Februar 2013 am Institut für Informatik eingerichtete Emmy Noether - Nachwuchsgruppe Machine Learning and Scientific Data Analysis beschäftigt sich mit der Entwicklung von Verfahren des maschinellen Lernens zur Modellbildung und Modellevaluierung, welche den experimentellen Beobachtungsprozess und die daraus resultierende Verschiedenartigkeit von Natur und Experiment explizit abbilden. Ergebnisse des Projektes sollen die experimentellen Kosten der Modellbildung reduzieren und auf Experimentaldaten gestützte Schlussfolgerungen belastbarer machen.

Am Institut für Informatik kooperieren wir eng mit der Arbeitgruppe Maschinelles Lernen (Prof. Tobias Scheffer). In Zusammenarbeit mit Arbeitsgruppen aus der Geophysik (Prof. Frank Scherbaum) und der kognitiven Psychologie (Prof. Reinhold Kliegl) arbeiten wir darüber hinaus an exemplarischen naturwissenschaftlichen Modellbildungsproblemen. Im Bereich der Geophysik beschäftigen wir uns mit der Bildung seismischer Risikomodelle aus Bodenbeschleunigungsdaten, die in unterschiedlichen Erdregionen aufgenommen wurden und daher Verteilungsverschiebungen unterliegen. Ansätze des Transferlernens korrigieren regionale Verteilungsunterschiede durch Inferenz entsprechender Korrekturmodelle aus beobachteten Daten. In Vorstudien konnten wir zeigen, dass sich damit genauere Bodenbeschleunigungsvorhersagen treffen lassen. Im Bereich der kognitiven Psychologie liegt ein Fokus auf der Modellierung menschlicher Blickbewegungen insbesondere beim Lesen von Texten. Solche Blicksteuerungsmodelle sind ein aktives Forschungsgebiet in der Experimentalpsychologie, da sie kognitive (Textverstehen), visuelle (Worterkennung) und motorische (Augenbewegungen) Prozesse miteinander in Beziehung setzen. Wir beschäftigen uns insbesondere mit Modellen interindividueller Unterschiede in Blickbewegungsmustern. Solche Modelle ermöglichen einerseits die Untersuchung individueller Komponenten der Blicksteuerung. Andererseits lassen sich möglicherweise Rückschlüsse auf die Identität eines Lesers aus Blickbewegungsdaten ziehen; wir werden daher auch den Nutzen von Blickbewegungsmodellen in biometrischen Anwendungen untersuchen.

7 Geomatik & Human-Computer Interfaces

Peter Löwe, Léonie Schäfer, GeoForschungsZentrum

This presentation provides an introduction to the state-of-the-art in Human-Computer Interfaces (HCI) for Geospatial Technologies:

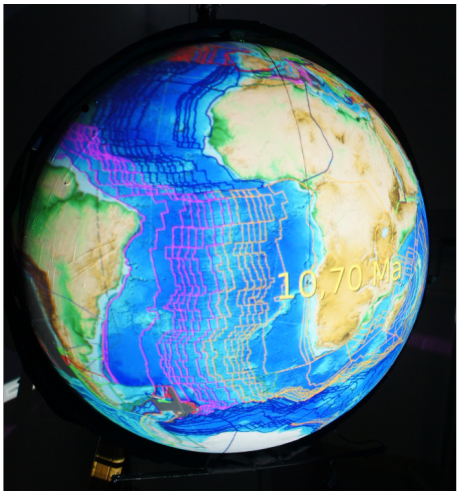
The daily use of Geographical Information Systems and other applications of computerised mapping have become mainstream in recent years: Computer-based maps are common practice on the World Wide Web, for smartphone use, navigation systems and for traditional desktop computing.

Sophisticated software applications allow the manipulation and analysis of geographical information and are used in location decisions of new businesses, for public service delivery and for planning decisions by local and central government.

However, many applications of GIS are hard to learn and to master: Until quite recently, the main focus of software vendors in the area of GIS was on the delivery of basic functionality and development of methods to present and manipulate geographical information using the available computing resources.

As a result, little attention was paid to usability aspects of GIS. This is evident in many public and private systems where the terminology, conceptual design and structure are all centred around the engineering of GIS and not on the needs and concepts that are familiar to the user.

In this presentation, the linkage between the fields of HCI and Geomatics is outlined and explored - mostly from a geographers perspective. In particular, the linkage between the areas of geographical information science and cartography on the one hand and HCI and usability engineering on the other shows that there is a disciplinary gap that makes more integrated interactions challenging.



8 Big Data: Anwendungsfelder in der Wissenschaft und Technologien

André Luckow, Institut für Informatik, BMW Group

Die Menge an Daten, die weltweit anfallen und verarbeitet werden, steigt ständig an. Ein Grund ist die allgegenwertige Verfügbarkeit von Sensoren, z.B. in mobilen Geräten, Maschinen, usw., die das umfangreiche Sammeln von Daten ermöglichen. Nach einer aktuellen Studie des IDC wurden im Jahr 2012 rund 2.2 Zetabyte digitale Daten erstellt, abgerufen oder repliziert; dazu zählen Texte, Bilder, Videos, Aufnahmen von Überwachungskameras, Lokationsdaten, usw. Der gröte Teil dieser Daten wird dabei von Maschinen, z.B. mobilen Geräten, Flugzeugen, Fahrzeugen, Industrie-Robotern, erzeugt. Insbesondere in verschiedenen Wissenschaftsdisziplinen werden immense Datenmengen erzeugt: so generieren die Detektoren des Large Hydron Colliders am CERN rund 1 PB Daten am Tag; eine moderne Genome Sequenzierungsmaschine kann $\gg$1 TB/ Daten pro Tag erzeugen. Das Datenwachstum wird in naher Zukunft weiter zunehmen. Die Bewältigung dieser Datenflut ist mit verschiedenen infrastrukturellen Herausforderungen verbunden: so müssen neben dem Storage, Datentransfers, Datenreplikationen effizient durchgeführt werden; ein gutes Daten/Compute Co-Placement ist die Voraussetzung für die optimale Verarbeitung der Daten.

Cloud-Anbieter, wie Google und Amazon besitzen mittlerweile groe, verteilte Infrastrukturen für die Verarbeitung von den durch ihre Diensten generierten Daten. Einen Teil der Infrastruktur wird auch externen Nutzern als Cloud-Dienst zur Verfügung gestellt. Infrastrukturen, wie EGI/Prace in Europa oder XSEDE in den USA, bieten immense Compute und Daten-Ressourcen für wissenschaftliche Anwendungen. Klassisch stand der Aspekt Compute im Mittelpunkt von solchen High Performance Infrastrukturen. Technologien aus dem Cluster und Grid Computing beschäftigten sich primär mit der Ausführung von Anwendungen in Rechnerverbänden, so genannten Cluster, sowie in hoch-verteilten Rechen-Grids. Daten-Handling Aspekte (Dateitransfers, File I/O) wurden in der Regel dem Compute Aspekt untergeordnet. Die Auswertung von Daten spielt aber zunehmend eine wichtigere Rolle - der Anteil an datenintensiven Anwendungen steigt unaufhaltsam.

Big Data beschreibt den aktuellen Trend und Hype um Technologien die den Umgang mit groen Datenmengen ermöglichen. Ein Beispiel für eine solche Technologie ist Hadoop, ein Framework welches auf dem MapReduce Programmiermodell aufbaut und die effiziente Ausführung von datenintensiven Anwendungen in einem Cluster erlaubt. Im Rahmen dieses Vortrages werden verschiedene Anwendungsfelder sowie Technologien für den Umgang mit Big Data vorgestellt.

9 Interaktive Routenvisualisierung auf einem taktilen Flächendisplay

Ulrike Lucke, Mihail Ivanchev, Institut für Informatik

Abbildung 4: Interaktive Karten- und Routen-Darstellungen werden für die Pin-Matrix-Repräsentation aufbereitet und taktil ausgegeben

Der eingeschränkte Informationszugang für Blinde und Sehbehinderte kann z.T. durch auditive oder taktile Alternativen kompensiert werden. Fr ein selbstständiges Leben ist dabei v.a. das Thema Mobilität von Bedeutung. Hier spielen sowohl die vorbereitende Erkundung einer Wegstrecke als auch die Orientierung in einer fremden Umgebung eine Rolle. Der erste dieser beiden Aufgaben ist die vorgestellte Arbeit gewidmet.

Unter Zugriff auf vorhandene Geo-Informationen wird die Präsentation einer Karte inkl. Route auf ein taktiles Flächendisplay abgebildet. Die Herausforderung liegt dabei sowohl in der auerordentlich geringen Auflösung derartiger Geräte als auch im Fehlen von Farbinformationen. Zusätzlich ist die interaktive Erkundung der Karte erforderlich. In dem Vortrag werden die prototypische Realisierung eines solchen Systems sowie derzeit noch offene Fragen diskutiert.

10 Information Modelling in the Geological Sciences

Jens Klump, CeGIT, German Research Centre for Geosciences GFZ

Geoscience data and derived information comes in many different shapes and formats. They cover the entire physical planet, its interactions with other spheres, and other extra-terrestrial bodies. Given this broad scope it is particularly challenging to develop an information model for the geosciences. Even though single investigations only study small subsets of the system, the interconnected nature of earth processes sometimes requires the integration of data from several different fields.

ISO TC 211 and its working groups, together with the Open Geospatial Consortium (OGC), have developed a whole family of standards that contribute to an information model for spatial information, notably the Geography Mark-up Language (GML). To enable the exchange of interpreted geological data the government geological surveys and the International Union of Geological Sciences (IUGS) initiated the development of a geosciences application schema of GML, called Geoscience Mark-up Language (GeoSciML). This mark-up language was designed to facilitate the exchange of geoscience information between government geological surveys through OGC web services.

With the rise of web services and XML came the idea that it should be possible to assemble components of a service oriented architecture in a modular fashion, similar to plug and play hardware. Web services were thought to integrate into web based, service oriented architectures by automatic inclusion in a registry of available services with standardised, machine interpretable descriptions of their interfaces. Information exchange was thought to be between machines only with no need for human intervention. Mark-up languages were seen as well suited media for machine-machine communication since they can be validated by machines and leave no room for interpretation or ambiguity. But does this assumption hold true? Is it possible to pre-define an unambiguous structure for this communication a perfect language? Or is the solution yet another mark-up language? Are there principal limitations to formal information models underlying mark-up languages and ontologies?

This lecture will introduce the fundamental concepts of information models in the geosciences and look into their implementation as mark-up languages and web services. The implementation will be illustrated by use cases. The lecture will also discuss the limitations of formal information models and resulting implications for the development of information models.

Curriculum Vitae: Dr. Jens Klump is a senior research scientist at the German Research Centre for GeoSciences (GFZ) in Potsdam, Germany. His field of research is data driven science. Research topics in this field are numerical simulation of natural gas reservoirs, virtual research environments, remotely operated instruments, high performance and cloud computing, long-term preservation of research data, and the development of system solutions for geoscience projects. Jens earned degrees in geology and in oceanography from the University of Cape Town (UCT) and received his PhD in marine geology from the University of Bremen, Germany.

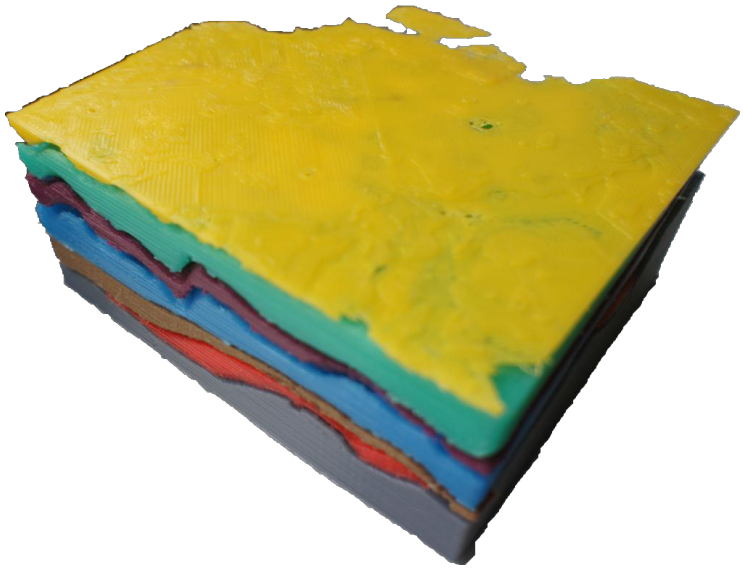


Abbildung 5: Darstellung eines Kunststoffmodells der Geologie Brandenburgs

11 Projektportale für raumbezogene Daten - Architektur und Entwicklung

Vivien Stender, Matthias Schroeder, GeoForschungsZentrum

Geowissenschaftliche Forschungen erzeugen stets auch raumbezogene Daten (Geodaten). Der bisherige Umfang an erhobenen Geodaten wird durch die fortschreitende Entwicklung der Informations- und Computer- Technologie (ICT) zukünftig um ein Vielfaches steigen. Insbesondere die Weiterentwicklung und der ansteigende Einsatz von Sensoren sorgen für einen enormen Anstieg an gemessenen Zeitreihen. Die so erhobenen Geodaten werden häufig, um sie allgemein zugänglich und wieder auffindbar zu machen, in Web-Portalen der geowissenschaftlichen Projekte visualisiert. Dabei werden die Geodaten oft mit projekteigenen Content- Management Systemen verknüpft.

Um eine interdisziplinäre, einrichtungsübergreifende Nutzung und den Austausch von Daten zu gewährleisten, ist die Anlehnung an anerkannte Web-Architekturen, die Verwendung von Standards für die Daten und Systemschnittstellen z.B. nach W3C, ISO und OGC sowie europäischen Infrastrukturmaßnahmen wie INSPIRE erforderlich. Die interdisziplinären Projekte SuMaRiO und TERENO können als Beispiele für derartige Projektportale herangezogen werden. In beiden Projekten werden standardisierte Geowebdienste in die Projektportale eingebunden. Auf diese Weise entsteht eine umfassende Informationsinfrastruktur als Grundlage und Voraussetzung für vernetztes, wissenschaftliches Arbeiten und ein Ergebnis- und Wissenstransfer. Diese Infrastruktur dient sowohl der Außendarstellung als auch als interne Informations- und Kommunikationsplattform. In späteren Projektphasen sollten diese Dienste für die Bereitstellung und Visualisierung von Ergebnissen für externe Partner und Interessenten nachgenutzt werden.

Der Beitrag stellt die grundlegenden Architekturkonzepte anhand der genannten Projekte vor. Es werden beispielhaft einige Softwarekomponenten und deren Einsatz für räumliche Daten vorgestellt.

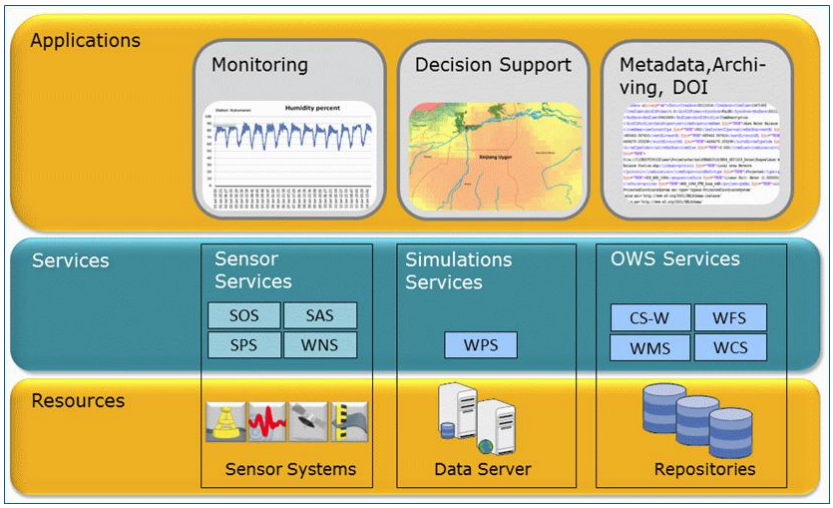


Abbildung 6: Quelle: @Center for GeoInformation Technology (CeGIT)