

TEODOOR: Geodateninfrastruktur zur Verwaltung und Veröffentlichung von terrestrischen Beobachtungsdaten der HGF Infrastrukturmaßnahme TERENO aus verteilten Quellen

Ralf Kunkel¹, Jürgen Sorg¹, Rainer Gasche², Jens Klump³, Olaf Kolditz⁴, Mark Frenzel⁴, Frank Neidl²

¹: Forschungszentrum Jülich GmbH, Institut für Bio- und Geowissenschaften – Agrosphäre (IBG-3), 52425 Jülich, E-Mail: r.kunkel@fz-juelich.de

²: Karlsruher Institut für Technologie (KIT), Institut für Meteorologie und Klimaforschung (IMK-IFU), Kreuzeckbahnstrasse 19, 82467 Garmisch-Partenkirchen.

³: Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum GFZ, Zentrum für Geoinformationstechnologie, Telegrafenberg, 14473 Potsdam

⁴: Helmholtz-Zentrum für Umweltforschung UFZ, Departments Umweltinformatik/Department Biozönoseforschung, Permoserstraße 15, 04318 Leipzig

Abstract

Im Rahmen der HGF-Infrastrukturmaßnahme TERENO werden in Deutschland seit 2008 vier langfristig betriebene terrestrische Umweltobservatorien eingerichtet, in denen große Mengen an umweltrelevanten Daten gesammelt werden. Um diese Datenmengen adäquat zu verwalten, zu beschreiben, auszutauschen und zu veröffentlichen wurde die verteilte Geodateninfrastruktur TEODOOR (<http://www.tereno.net>) aufgebaut. Die für die Einrichtung und Betrieb der Observatorien verantwortlichen Institutionen betreiben dabei jeweils eigene Dateninfrastrukturen, die über lokale, standardisierte OGC-konforme Web-Dienste miteinander kommunizieren. Ein zentrales TEODOOR-Datenportal hat die Aufgabe, die Daten und Metadaten aus den einzelnen Observatorien zusammenzuführen und für die interne und öffentliche Verwendung bereitstellen. Eine Reihe von Werkzeugen wie z.B. die hierarchische Suche nach Stichworten, Web-GIS Funktionalitäten und Visualisierungswerkzeuge erlauben es dem Benutzer, Daten von verschiedenen Zeitreihen und Variablen zu suchen, zu selektieren, darzustellen und konform zur gemeinsamen Datenpolitik auch herunterladen werden. Derzeit sind über TEODOOR zeitaktuelle Daten von mehr als 450 automatischen Stationen frei verfügbar.

Within the TERENO initiative four long-term terrestrial observatories, which collect huge amounts of environmental relevant data, have been set up since 2008. In order to manage, describe, exchange and publish these data, the distributed Spatial Data Infrastructure TEODOOR (<http://www.tereno.net>) has been created. Each institution responsible for an individual observatory set up its own local data infrastructure,

which may communicate to each other to exchange data and metadata internally or to the public by OGC-conformal Web services. The TEODOOR data portal serves as a database node to provide scientists and decision makers with reliable and well accessible data and data products. Various tools like hierarchical search or Web-GIS functions allow a deeper insight into the different observatories, test areas and sensor networks. Sensor data can be selected according to the measured parameters, stations and/or time periods, visualized and downloaded according to the common TERENO data policy. Currently, TEODOOR provides free access to data from more than 450 monitoring stations.

1 Einleitung

Die Änderungen von Klima und Landnutzung sind wesentliche Faktoren des globalen Wandels, die nicht nur die Gesellschaft, sondern auch lebenswichtige Ökosystemfunktionen und Ressourcen stark beeinflussen. Langfristige Trends in Temperatur, Niederschlag und weiteren klimatischen Parametern beeinflussen alle Umweltkompartimente nicht nur direkt, sondern auch indirekt durch komplexe Rückkopplungsmechanismen. Eine wichtige Aufgabe der terrestrischen Ökosystemforschung ist die Entwicklung und Bereitstellung integrierter Modelle zur Quantifizierung der Auswirkungen von Umweltveränderungen auf terrestrische Systeme. Als Grundlage hierfür sind langfristige, interdisziplinäre und skalenübergreifende Umweltdaten notwendig (Bronstert et al., 2009; Reid et al., 2009; Richter and Mobley, 2009). In diesem Zusammenhang wird gerade in letzter Zeit die Entwicklung und Implementierung von Dateninfrastrukturen für die Umweltforschung sehr intensiv betrieben (Band et al., 2005; Burt et al., 2008; Keller et al., 2008; Lin, 2003; Lin, 2010; McDonnell et al., 2007; Montgomery et al., 2007; Nisbet, 2007; Parr et al., 2003; Zoback, 2001). Getrieben durch die rasante Entwicklung hochauflösender Mess- und Fernerkundungstechnik werden in diesen Infrastrukturen große Mengen heterogener Umweltdaten generiert.

Um diese Datenmengen für terrestrische Systemanalysen nutzbar zu machen, sind zwei Schritte erforderlich. Zum einen müssen neuartige Datenmanagement- und Schnittstellenkonzepte entwickelt werden, um die anfallenden Daten in den verschiedenen Zeit- und Raumskalen verarbeiten und effizient austauschen zu können. Hierzu müssen die existierenden, aktuell meist nicht kompatiblen Ansätze zur Schaffung von Interoperabilität in den einzelnen Wasser- und Klimaforschungs-Communities geeignet zusammengeführt, visualisiert und über standardisierte Webdienste publiziert werden. In diesem Zusammenhang spielen die Aktivitäten und Standards des *Open Geospatial Consortiums* (OGC) eine wichtige Rolle. Zum anderen müssen integrierte Modellansätze direkt an diese Datenströme und Informationsinfrastrukturen angebunden werden. Nur so kann es gelingen, das Systemverständnis unter Berücksichtigung der Wechselwirkungen von Atmosphäre, Landoberfläche, Gewässer, Boden

bis hin zum Grundwasser auf einer so derzeit nicht verfügbaren Umweltdatenbasis auf ein neues Niveau zu heben.

Mit dem TERENO-Vorhaben (TERrestrial ENvironmental Observatories) hat die Helmholtz-Gemeinschaft 2008 begonnen, ein Netzwerk langfristig betriebener, integrierter Umweltobservatorien in Deutschland einzurichten (Bogena et al., 2006; Zacharias et al., 2011) mit dem Ziel, die Auswirkungen des Klima- und Landnutzungswandels auf die terrestrischen Umweltsysteme langfristig, interdisziplinär und skalenübergreifend zu untersuchen. Hierzu wurden in Deutschland vier terrestrische Observatorien eingerichtet, die von jeweils einem Helmholtz-Zentrum betrieben werden (siehe Abbildung 1). Übergeordnetes Ziel von TERENO ist letztlich die Nutzung der in den einzelnen Observatorien erhobenen Umweltdaten für prognosefähige Systemanalysen, um beispielsweise die Einflüsse etwaiger Klima- und Landnutzungsänderungen auf die langfristige Verfügbarkeit und Qualität von Wasserressourcen abschätzen zu können.



Abbildung 1: Lage der vier TERENO-Observatorien (Zacharias et al., 2011)

2 Aufbau von TEODOOR

Der gesamte Erfolg von Langzeitprojekten wie TERENO hängt ganz entscheidend von einer gut organisierten Datenverwaltung, dem Datenaustausch zwischen den beteiligten Teilprojekten und der Verfügbarmachung der erhobenen Daten ab. Um dieses zu gewährleisten, wurde die räumliche Dateninfrastruktur (SDI) TEODOOR (TEReno Online Data RepOsitORY) aufgebaut, welche das Bindeglied bildet, über das Wissenschaftlern und Entscheidungsträgern zuverlässige und zugängliche Umweltdaten und weiterführende Produkte zur Verfügung gestellt werden.

Aufgrund der Heterogenität der Daten, der großen Menge anfallender Daten und der bereits bestehenden Dateninfrastrukturen in den beteiligten Institutionen wurde in TERENO ein dezentrales Datenmanagementkonzept umgesetzt (siehe Abbildung 2). Die Grundidee hierbei ist die Verknüpfung lokaler unabhängiger Infrastrukturen (Daten- bzw. Metadatenbanken) in den einzelnen Observatorien mit einer zentralen TERENO-Portalanwendung (TEODOOR). Das TEODOOR Geodatenportal bündelt die über die verschiedenen Webdienste distributiv zur Verfügung gestellten Daten

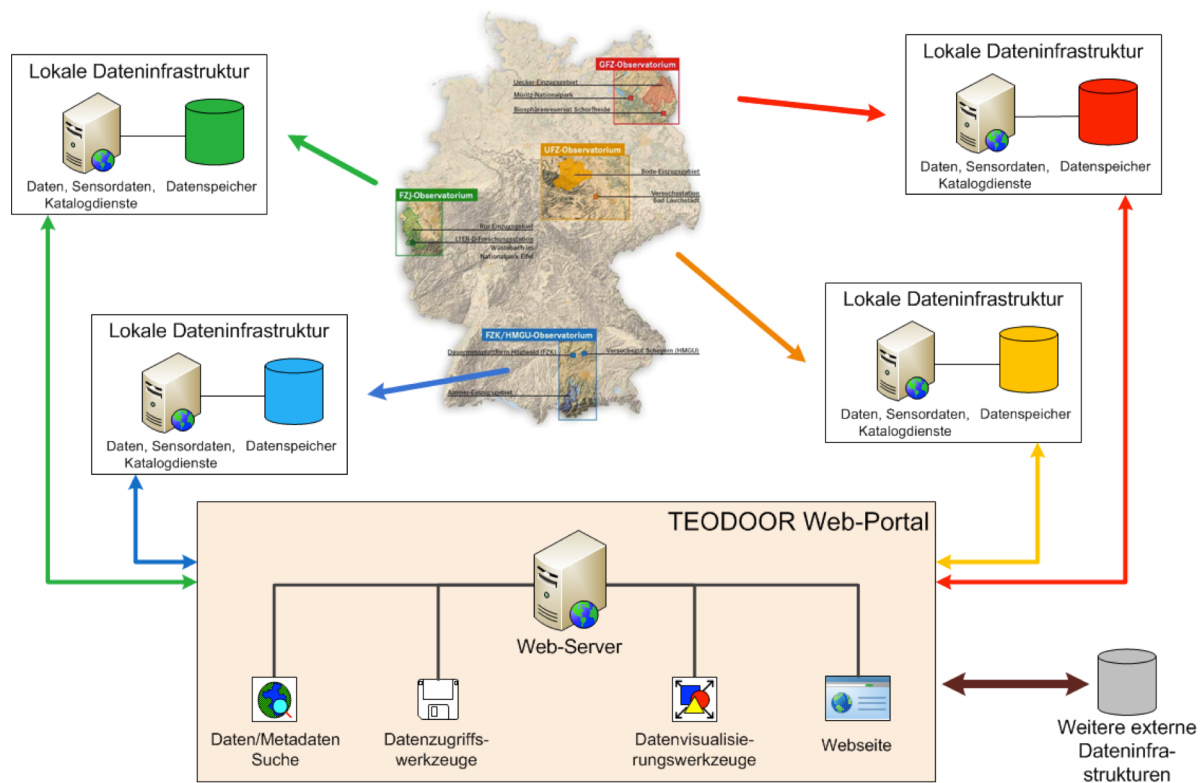


Abbildung 2: Aufbau der Dateninfrastruktur TEODOOR

der einzelnen Observatorien über Such-, Visualisierungs- und Downloadwerkzeuge und stellt sie der wissenschaftlichen Gemeinschaft zur Verfügung. Hierzu werden Webdienste verwendet, die durch das Open Geospatial Consortium (OGC) standardisiert wurden (OGC, 2012), wie z.B. Web Coverage Services (WCS), Web Feature Services (WFS), Web Map Services (WMS), Web Processing Services (WPS), Web Catalogue Services (CSW) und insbesondere dem Sensor Web Enablement (SWE). Diese Webdienste stellen Standards und Schnittstellen zum Suchen in und zum Zugriff auf Geodatenbestände zur Verfügung. Durch die Verwendung standardisierter Schnittstellen wird zum einen eine größtmögliche Flexibilität des Datenmanagements in den einzelnen Observatorien sichergestellt. Zum anderen wird auch die Übertragbarkeit der Daten aus den einzelnen Observatorien gewährleistet, so dass auch Daten von weiteren, neu eingerichteten Observatorien oder anderen behördliche Datenhaltern ohne größeren Aufwand eingebunden werden können.

Ein wesentlicher Aspekt von TERENO ist die Verfügbarmachung der gemessenen Daten für die Öffentlichkeit. Als Grundlage hierfür wurde in einem partizipatorischen Prozess eine gemeinsame TERENO-Datenpolitik erarbeitet, in der festgelegt wurde, welche Daten unter welchen Bedingungen zugänglich sind (TERENO, 2010). Innerhalb des TERENO-Verbundes sind alle Daten frei verfügbar. Der Öffentlichkeit werden grundsätzlich alle Daten umgehend zugänglich gemacht, wenn die Daten mindestens eine erste Qualitätskontrolle erfolgreich durchlaufen haben und der Veröffentlichung keine anderen Nutzungsvorbehalte, z.B. für laufende Promotionsvorha-

ben entgegenstehen. Neben der Datenqualität spielt insbesondere die adäquate Beschreibung der Datenbestände über Metadaten eine entscheidende Rolle. Aus diesem Grunde werden Daten immer als eine Kombination der Daten und der beschreibenden Metadaten behandelt. Durch ein gemeinsames TERENO-Metadatenprofil wurde eine einheitliche Datenbeschreibung innerhalb des TERENO-Verbundes sichergestellt.

Während der Einrichtung und Instrumentierung der TERENO-Observatorien wurden von den beteiligten Zentren Dateninfrastrukturen eingerichtet, in welche die TERENO-Daten sukzessive integriert werden. Auch wenn die vier Observatorien Unterschiede hinsichtlich des Schwerpunktes der untersuchten Fragestellungen aufweisen, sind die anfallenden Daten bei den verschiedenen Observatorien sehr ähnlich und sollen am Beispiel der Zeitreihendaten des vom FZJ betriebenen Eifel-Rur Observatoriums exemplarisch verdeutlicht werden.

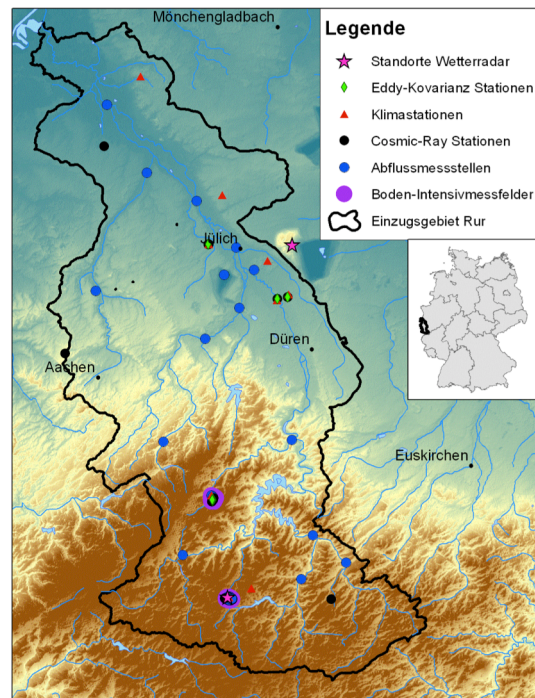


Abbildung 3: Lage der Intensivtestgebiete im Eifel-Rur Observatorium.

3 Lokale Dateninfrastruktur „Eifel-Rur Observatorium“

Das Observatorium Eifel-Rur umfasst das Einzugsgebiet der Rur mit einer Größe von 2354 km². In drei repräsentativen Regionen des Einzugsgebietes wurden Intensivmessgebiete eingerichtet, in denen eine Reihe von Parametern, wie z.B. Bodenfeuchte und Bodentemperaturen, Grundwasserständen sowie Energie und Gasflüssen erfasst werden. Verteilt über das Einzugsgebiet werden zusätzlich an Cosmic-Ray Stationen Bodenfeuchtedaten (Desilets et al., 2010) sowie an Eddy-Kovarianz Stationen Klima-, Energie- und Gasflussdaten erhoben (siehe auch Abbildung 3):

- 53 automatische Stationen liefern mit einer Frequenz von 1-5 h⁻¹ Abfluss-, Klima-, Oberflächengewässerqualitäts- und Bodendaten. Pro Tag werden von diesen Stationen insgesamt 26400 einzelne Datenwerte erfasst.
- Zwei SoilNet Beobachtungsnetzwerke (Bogena et al., 2010) erfassen mit einer Frequenz von 6 h⁻¹ Bodentemperaturen, Bodenfeuchten und Klimaparameter an insgesamt 404 Messstellen mit einem Umfang von insgesamt 686000 Messwerten pro Tag.
- Durch sieben Eddy-Kovarianz Stationen (Vanderborcht et al., 2010) werden eine Reihe von Klimaparametern und Gaskonzentrationen mit einer Frequenz

von 20 s^{-1} (Rohdaten) bzw. 6 h^{-1} (prozessierte Daten) gemessen, die sich pro Tag auf 133 Millionen einzelner Messwerte summieren.

- Ein Wetterradar auf der Sophienhöhe nahe Jülich und ein kleinerer Rainscanner im Süden des Einzugsgebietes erzeugen mit einer Frequenz von 12 h^{-1} bzw. 60 h^{-1} jeweils ein Niederschlagsraster mit 800×800 Pixeln. Pro Tag fallen 1728 Rasterdateien an, die zusammen mit den Rohdaten insgesamt 4.3 GB Speicherplatz beanspruchen.

Zusätzlich werden Abfluss- und Klimadaten aus externen Beobachtungsmessnetzen von Wasserverbänden und dem Bundesland NRW integriert. Neben den Stationsdaten fallen auch Dokumente, Bilder und GIS-Daten in größerem Umfang an. Diese Daten werden über ein Content-Management System (Plone) und einen GIS-Server verarbeitet, was allerdings hier nicht weiter betrachtet werden soll.

3.1 Stationsdaten

Die Verarbeitung der Stationsdaten erfolgt automatisiert durch ein für diesen Zweck im FZJ entwickeltes integriertes Zeitreihenmanagementsystem (TSM 2.0). Die Sensordaten, welche die Infrastruktur über unterschiedliche Kanäle, wie z.B. DSL, E-Mail, GSM, Richtfunk oder manuellen Upload erreichen, werden in einem ersten Schritt in ASCII-kodierte Dateien überführt. Der Inhalt der Dateien wird in einem nächsten Schritt in eine relationale Datenbank überführt, wobei der Überführungsprozess für die einzelnen Stationen bzw. Stationstypen individuell konfigurierbar ist. Gegebenenfalls erfolgt an dieser Stelle auch eine Umrechnung der Rohdaten, z.B. von Signalstärken in Temperaturen, die Anwendung von Umrechnungsvorschriften für mehrere Eingangsgrößen, wie z.B. die Bildung von Mittelwerten oder zeitlichen Aggregaten. Konfigurierbare Filter ermöglichen es, eine erste Plausibilitätsprüfung und Kennzeichnung der Datenqualität vorzunehmen. Nach erfolgreichem Datenimport werden die Originaldateien archiviert. Das System benachrichtigt die für eine Messstelle verantwortlichen Techniker und/oder Wissenschaftler über den erfolgten Importprozess, über aufgetretene Fehler, Warnungen und Datenlücken, aber auch über Probleme bei der Übertragung der Rohdaten von der Messstelle zur Infrastruktur.

Gemessene Daten sind primär durch Ort und Zeitpunkt der Messung, die physikalische Größe und den Messwert charakterisiert. Für eine spätere Nutzung der Daten sind diese Angaben allein allerdings nicht ausreichend. Hierzu sind zusätzliche beschreibende Angaben, die sogenannten Metadaten, erforderlich (Bose, 2002; Gray et al., 2005; Michener et al., 1997; Tomasic and Simon, 1997). Metadaten enthalten alle notwendigen Angaben zur Charakterisierung der gemessenen Größe, wie z.B. Namen, Attribute, Einheiten, Variablenbeschreibungen, Datengenauigkeit, Datenlayout, beschreiben aber auch den Prozess der Messung, Aufbereitung und Berechnung der Daten (Gray et al., 2005). Um dem Rechnung tragen zu können, wurde ein Datenmodell für Zeitreihendaten entwickelt, in dem neben den eigentlichen Daten auch alle relevanten Metadaten verwaltet werden können. Das Datenmodell basiert in weiten

Teilen auf dem Observation Data Model (ODM), das im Rahmen der CUAHSI-Initiative entwickelt und eingesetzt wird (Horsburgh et al., 2008; Tarboton et al., 2009). Es wurde hier allerdings erweitert, um die Prozesse von Datenerfassung, Datenimport und Datenveröffentlichung adäquat abbilden zu können.

Eine schematische Darstellung des Datenmodells gibt Abbildung 4. Es beinhaltet mehrere Komponenten, über die eine detaillierte Charakterisierung der Messwerte (data values), der Messstellen (sites) der Messgrößen (variables) erfolgt. Zusätzlich werden die bei der Messung verwendeten Sensoren, die Konfiguration der einzelnen Logger und die beim Import anzuwendenden Berechnungsverfahren beschrieben (logger, sensors, agents and functions). Das Modell enthält darüber hinaus Tabellen zur Spezifizierung der jeweils verantwortlichen Personen (sources), zur Verwaltung von Proben (samples) sowie zur Veröffentlichung der Daten (publishing and maintenance).

Neben der Langzeitspeicherung und Dokumentation der Sensordaten ist insbesondere die Schaffung von Zugängen zu den Daten ein wesentlicher Aspekt von TERENO. Innerhalb des distributiven Aufbaus von TERENO stellt dies nicht zuletzt aufgrund der großen Vielfalt an Herstellern von Sensorsystemen und den unter-

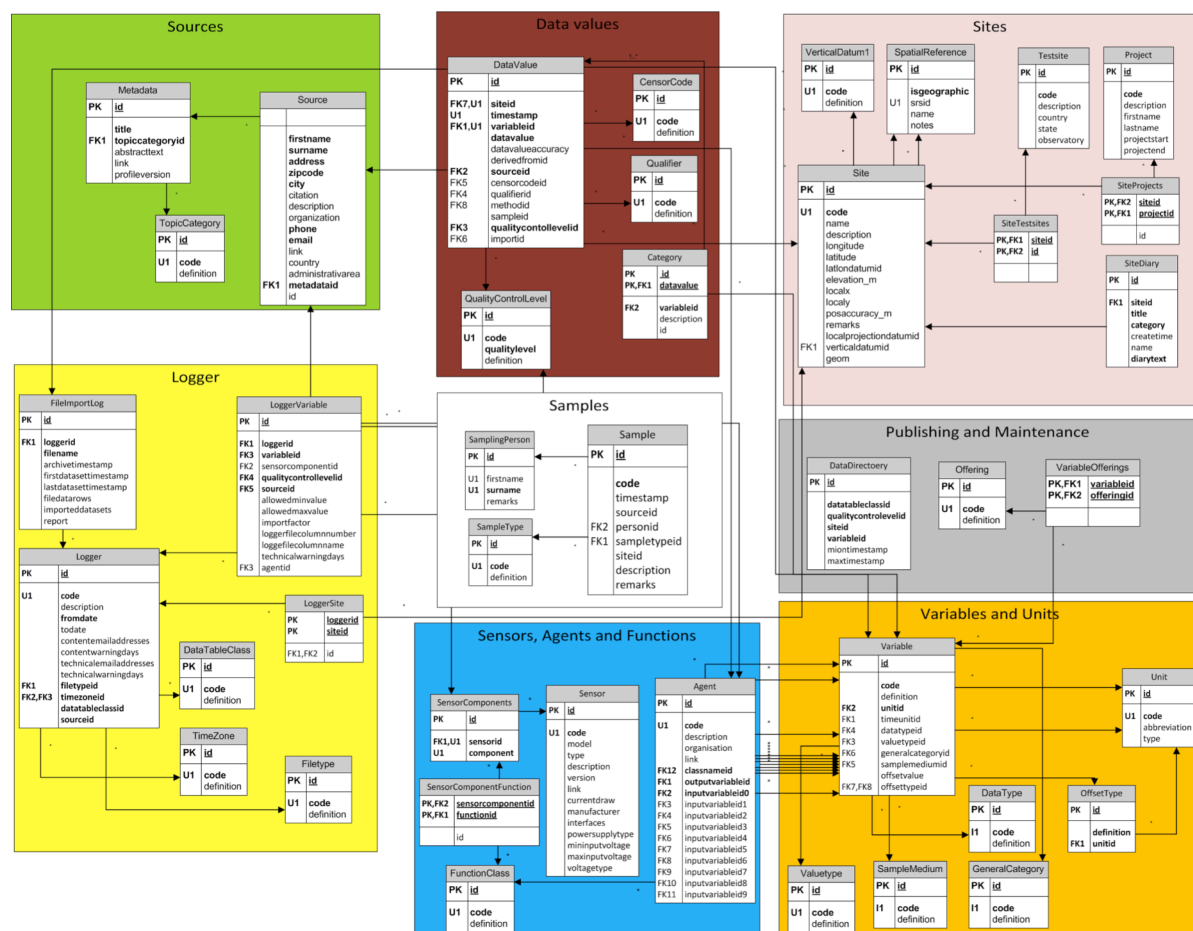


Abbildung 4: Datenmodell zur Verarbeitung der Sensordaten im TERENO-Observatorium Eifel-Rur.

schiedlichen Übertragungsprotokollen eine große Herausforderung dar. Der Aufbau von Beobachtungsnetzwerken erfordert eine interoperable, hardware- und plattform-unabhängige Behandlung der einzelnen Sensoren (Bröring et al., 2011a; Bröring et al., 2011b; Jirka et al., 2012), so dass die Daten kohärent beschrieben, gesucht, ggf. kontrolliert sowie über einheitliche Schnittstellen auf sie zugegriffen werden kann.

Eine solche einheitliche Behandlung von Sensoren kann durch den vom OGC bereitgestellten Sensor Web Enablement (SWE) realisiert werden, welcher eine Reihe aufeinander abgestimmter Standards und Schnittstellendefinitionen zur Verfügung stellt (Botts et al., 2008): der Observation & Measurement Standard zur Beschreibung und zum Austausch von Beobachtungsdaten (Cox, 2010; Cox, 2011), die Sensor Model Language (SensorML) als Informationsmodell zur Beschreibung von Sensoren und der von ihnen gelieferten Informationen (Botts, 2007) und die TransducerML (TML) als ein Protokoll für den Austausch von Informationen zwischen den Sensoren und einer SensorWeb-Infrastruktur. Als Dienste definiert der SWE den Sensor Observation Service (SOS) für die Suche, das Filtern und den Abruf von Sensordaten (Bröring et al., 2012; Na and Priest, 2007), den Sensor Planning Service (SPS) zum Abrufen und Web-basierten Steuern von Sensoren (Simonis and Echterhoff, 2011), den Sensor Alert Service (SAS), der es Nutzern ermöglicht, über eine Standardschnittstelle Alarmnachrichten von einzelnen Sensoren zu erhalten und schließlich den Web Notification Service (WNS) (Simonis and Wytzisk, 2003) für den asynchronen Austausch von Informationen zwischen den einzelnen Webdiensten.

Die Open-Source Initiative 52North hat die SWE-Dienste implementiert und als freie Software veröffentlicht (52°North, 2012). Im Rahmen von TERENO werden die Sensordaten über die 52North Implementierung zugänglich gemacht, wobei zum einen der interne Zugriff auf die Daten an das verwendete Datenmodell angepasst und zum anderen in das O&M-Modell zum Datenaustausch TERENO-relevanten Angaben, wie z.B. die Qualitätskennzeichnungen, aufgenommen wurden. Für das Observatorium Eifel-Rur wurde für jedes Intensivmessgebiet und das gesamte Einzugsgebiet jeweils ein Dienst für interne und externe Nutzung aufgesetzt, so dass die Sensordaten durch insgesamt 14 Sensor Observation Services veröffentlicht werden.

3.2 Rasterzeitreihen

Die von einem Sensor gemessenen Werte werden einem Geoobjekt zugeordnet, dem sogenannten *Feature Of Interest* (FOI). Dieses Geoobjekt, für das die Messwerte gelten, kann der Punkt der Messung sein, aber auch aus einer Fläche oder einem Volumen bestehen. Die bisher betrachteten Messdaten hatten alle zu Eigen, dass die mit einem Sensor zu einem bestimmten Zeitpunkt gemessenen Größen ein Integral für das gesamte Feature-Of-Interest repräsentieren. Sensoren können aber auch für einen Zeitpunkt flächendifferenzierte Daten, meist Rasterdaten liefern. Hierzu zählen alle Arten von Fernerkundungsdaten, bei denen bei einem Überflug eines Flugzeugs oder Satelliten die flächenhafte Variation einer Zustandsgröße (Landnutzung, Radar-

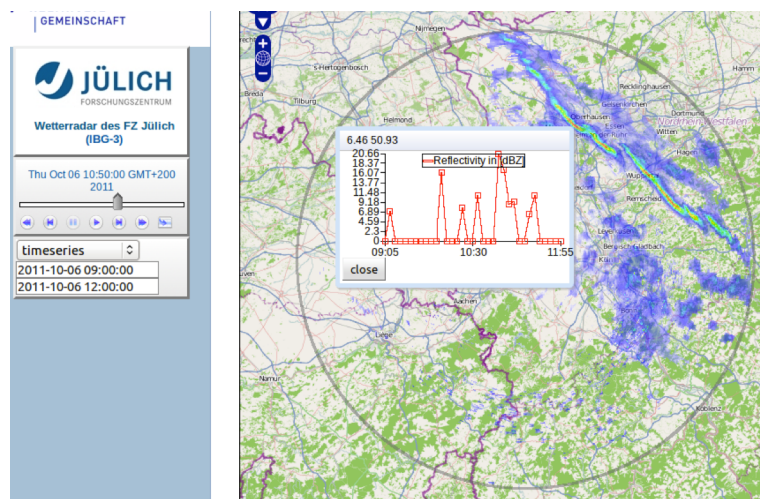


Abbildung 5: Klientanwendung zur animierten Darstellung von Rasterzeitreihen am Beispiel des Wetterraders Sophienhöhe.

echo etc.) für ein bestimmtes Gebiet gemessen wird. Hierzu zählen aber auch die Daten, die von den im TERENO-Vorhaben eingesetzten Radargeräten erzeugt werden. Diese liefern in regelmäßigen Zeitabständen (1-5 Minuten) jeweils einen Rasterdatensatz mit Reflektivitäten, aus denen über Berechnungsvorschriften u.a. Niederschlagshöhen ermittelt werden können. Auch diese Rasterzeitreihendaten sind in der OGC-Spezifikation

von Sensor Observation Services vorgesehen; es existierte bislang aber keine Implementierung des Standards. Aus diesem Grunde wurde auf Grundlage des SOS des 52North ein OGC konformer Sensor Observation Service entwickelt, der auch für Rasterzeitreihendaten anwendbar ist (Sorg, 2012). Hierzu wurde ein Datenmodell entwickelt, welches dem hohen Speicherbedarf der Rasterdaten (hier etwa 65000 mal größer als Punktdaten) Rechnung trägt, es aber auch ermöglicht, die Daten in Bilder mit variabler Farbskala umzuwandeln. Darüber hinaus wurden effiziente Algorithmen zur räumlichen Suche und zur Filterung von Rasterdaten entwickelt. Damit wurde im Ergebnis ein OGC-konformer Dienst zu Verfügung gestellt, der einen Zugang zu einzelnen Reflektivitätsrastern und frei wählbaren Ausschnitten bis hin zu Zeitreihendaten an frei wählbaren Einzelpunkten ermöglicht. Durch die Kopplung mit einem Sensor Alarm Service (SAS) wurden die Möglichkeiten eines SOS zur Filterung von Daten genutzt, in dem Nutzern die Möglichkeit geschaffen wurde, sich sektorbezogen vor Starkregenereignissen warnen zu lassen (Sorg, 2012).

Zur Visualisierung von Rasterzeitreihendaten wurde eine Klientanwendung entwickelt, die ausschließlich über die OGC-Standardschnittstellen mit einem Raster-SOS kommunizieren kann. Die Anwendung beinhaltet einen Player zur Darstellung animierter Rasterzeitreihen aus einem wählbaren Zeitraum mit wählbarer räumlicher Ausdehnung und die Möglichkeit, sich eine Zeitreihe der Rasterdaten an einem wählbaren Ort ausgeben zu lassen. Dies ist in Abbildung 5 am Beispiel einer Zeitreihe von Radarreflektivitäten des Wetterraders Sophienhöhe dargestellt. Der Klient ist in das TEODOOR-Datenportal eingebunden und wird dort derzeit zur Visualisierung von Wetterrardaten eingesetzt¹, kann aber für alle Arten von Rasterzeitreihendaten, die über einen Raster-SOS zur Verfügung gestellt werden, verwendet werden.

¹ http://teodoor.icg.kfa-juelich.de/@@ibg.coverage_player

3.3 *Datensuche und Katalogdienst*

Eine Reihe von Untersuchungen haben gezeigt, dass die Nachfrage an distributiv verfügbaren Daten zwar kontinuierlich ansteigt, ihre tatsächliche Nutzbarkeit aber durch mangelnde Kenntnis der vorhandenen Daten, durch schlechte Dokumentation der Daten und insbesondere durch Dateninkonsistenzen eingeschränkt wird (Nogueras-Iso et al., 2005; Tomasic and Simon, 1997). Eine weitergehende Nutzung der in einer Dateninfrastruktur verwalteten Daten und der zur Verfügung gestellten Datendienste erfordert zwingend deren umfassende Beschreibung durch Metadaten. In Abhängigkeit vom Datentyp existieren gängige Standards, die festlegen, durch welche Angaben die Daten und die Datendienste beschrieben werden (FGDC, 2000; ISO, 2003; ISO, 2009). Für einzelne Standards werden oftmals ergänzend Profile definiert, mit denen die zwingend oder optional geforderten Elemente näher spezifiziert, festgelegt oder eingeschränkt werden. Auch im Rahmen des TERENO-Vorhabens wurde ein Metadatenprofil definiert, welches auf der EU-INSPIRE-Richtlinie Metadaten (ECJRC, 2009) basiert.

Katalogdienste ermöglichen es Nutzern oder anderen Geodiensten, herauszufinden, welche Datendienste verfügbar sind und welche für die Nutzer relevanten Daten sie enthalten. In der Katalogdienst-Spezifikation des OGC (Nebert et al., 2007) ist festgelegt, wie diese Katalogdienste aufgebaut sind und auf welche Art und Weise Metadaten gesucht und übertragen werden. Katalogdienste werden für die meisten gängigen Typen von Geodaten, aber auch für nicht raumbezogenen Daten standardmäßig eingesetzt.

Bei der Kopplung von Katalogdiensten mit Sensor(meta)daten gibt es derzeit allerdings noch eine Reihe von Problemen, die aus der besonderen Charakteristik von Sensornetzwerken resultieren. An erster Stelle spielt hierbei die Tatsache eine Rolle, dass bei Sensornetzwerken Zeitreihen von meist mehreren Parametern anfallen. Aufgrund der Dynamik der Datenerfassung ist dies durch die gängigen Metadatenstandards kaum abbildbar. Weiterhin werden die Sensormetadaten, d.h. Informationen zu den einzelnen Sensoren, den Messprozessen, den gemessenen Eingabe- und Ausgabegrößen sowie andere wichtige Informationen entsprechend dem SWE durch die SensorML spezifiziert. Diese ist allerdings nur eingeschränkt kompatibel zu den für allgemeine Geodaten verwendeten Standards (ISO, FGDC, DublinCore). Schließlich bestehen bei Sensornetzwerken besondere Anforderungen an die Suche nach Informationen, z.B. nach bestimmten Messgrößen, Messorten, Zeitpunkten der Messung, oder Datengruppen (Jirka et al., 2009). Diese Informationen sind ebenfalls in der SensorML spezifiziert und über Anfragen an einen Sensor Observation Services zugänglich, aufgrund der unterschiedlichen Strukturen nicht jedoch über die klassischen Protokolle der Web-Katalogdienste.

Eine Lösung für dieses Problem bietet das Electronic Business Registry Information Model, kurz ebRIM, welches ein generelles XML-basiertes Modell zur Verwaltung

elektronischer Daten darstellt (OASIS, 2002). ebRIM stellt dabei nur eine Grundstruktur zur Verfügung, die mittels einer Erweiterung für SensorML-Metadatendokumente (Chen et al., 2009) anwendbar gemacht wurde. Um die sehr umfangreichen Möglichkeiten, Sensoren mittels SensorML zu beschreiben, auf ein Minimum an obligatorischen Werten zu beschränken, wird ein Profil der SensorML definiert, dass aus Elementen zur Identifikation und Klassifikation sowie aus räumlichen und zeitlichen Elementen besteht. Dadurch wird sichergestellt, dass diese obligatorischen Elemente in jedem auf ebRIM basierenden Katalogdienst (CSW) verfügbar sind, der das SensorML-Profil unterstützt. Realisiert wurde der auf ebRIM basierende Katalogdienst durch das quelloffene Buddata ebXML Registry/Repository (Buddata-ebRR) Projekt, dass den OASIS ebXML Registry und den OGC CSW Standard implementiert (kZen Labs, 2012). Aus den vorhandenen SensorML-Dokumenten wurden mittels xslt-Transformationen ebRIM-XML-Dokumente erzeugt und über die vom Buddata-ebRR-Dienst bereitgestellten Schnittstellen in das ebRIM überführt. Im TEODOOR Portal wurde eine benutzerfreundliche, grafische Schnittstelle implementiert, die eine Suche nach ebRIM-Metadaten, z.B. nach Sensornamen, Parametern, Messzeiträumen oder Datengruppen im Buddata-ebRR-Dienst durch Nutzung der dort implementierten standardisierten CSW-Operationen (Lesage, 2007) ermöglicht.

4 TEODOOR Datenportal

Zusätzlich zu seiner Funktion als gemeinsame Informations- und Austauschplattform für den TERENO-Verbund hat das TEODOOR-Portal die Funktion, die in den einzelnen Observatorien gemessenen und über OGC-Dienste zur Verfügung gestellten Daten zusammenzuführen und der Öffentlichkeit zugänglich zu machen. In Abbildung 6 ist der Aufbau des TERENO Datenportals TEODOOR dargestellt. Im Kern

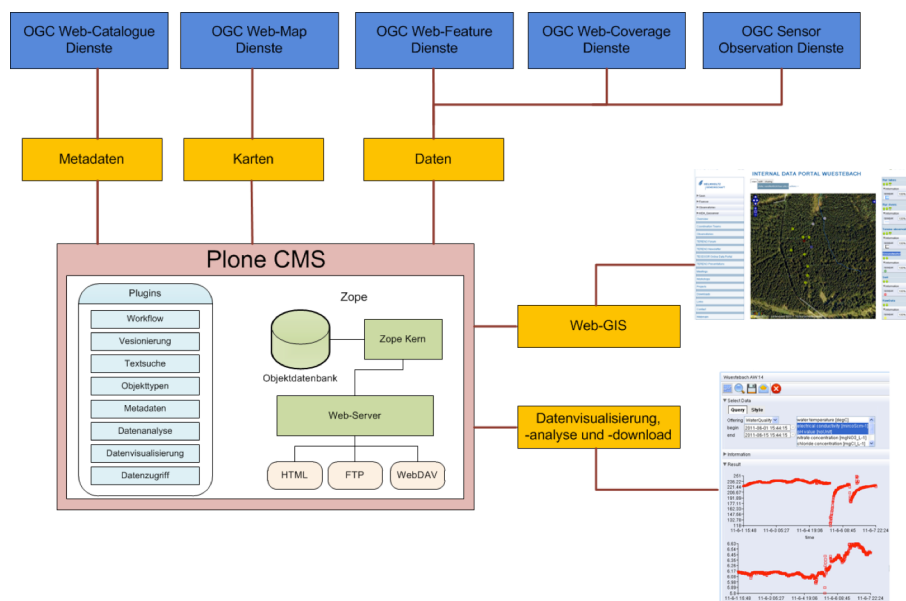


Abbildung 6: Aufbau des TEODOOR Datenportals.

besteht das System aus dem Content Management System Plone (Plone Foundation, 2004), welches die TERENO Webseite zur Verfügung stellt und es gleichzeitig TERENO Nutzern erlaubt, eine große Vielfalt von Objekttypen (Dokumente, Daten, Termine, Ressourcen, ...) in das System zu laden, zu verwalten und zu teilen. Plone verfügt über umfangreiche Such- und Datenzugriffsfunktionalitäten und unterstützt nativ den DublinCore Metadatenstandard (ISO, 2009).

Der überwiegende Teil der zur Verfügung stehenden Daten, insbesondere die Geo und Sensordaten, werden jedoch nicht über TEODOOR verwaltet, sondern sind über OGC-Schnittstellen an entfernte Datenquellen angekoppelt. Hierzu wurde das System erweitert, so dass eine Datensuche in verteilten OGC-Katalogdiensten möglich ist. Diese umfasst die hierarchische Suche nach Schlüsselworten, die Suche in Textdokumenten, aber insbesondere auch die Suche nach Sensordaten in entfernten ebRIM-Metadatenbeständen. Der Datenzugriff erfolgt direkt über die in den Metadaten angegebenen Speicherorte (Online Resources).

Zur Darstellung und Visualisierung der räumlichen Daten wurden in TEODOOR WebGIS Funktionalitäten integriert. Diese ermöglichen es, die über entfernte Kartendienste (WMS) zur Verfügung gestellten Daten (Raster- oder Vektordaten) grafisch darzustellen. Darüber hinaus wurden Anwendungen entwickelt, um die Daten der Sensornetzwerke (alle oder Suchergebnisse) von entfernten SOS abzurufen, darzustellen und zu visualisieren. Auch hierbei wurden ausschließlich die OGC-Standards verwendet. Der Nutzer kann hierzu in einer WEBGIS Anwendung die für ihn relevanten Messstellen selektieren, und über eine Klientanwendung (siehe Abbildung 7) Parameter und Zeitreihe auswählen. Die Daten werden dann als Grafiken dargestellt und können im O&M-Format (Cox, 2011) heruntergeladen werden. Entsprechend der TERENO Datenpolitik sind die Daten grundsätzlich frei verfügbar; sie werden den Nutzern nach Anerkennung der Nutzungsbedingungen per E-Mail zugestellt, wobei die für eine bestimmte Messstelle verantwortlichen Wissenschaftler parallel über den Datenabruf benachrichtigt werden.



Abbildung 7: Visualisierung von Sensordatenzeitreihen in TEODOOR am Beispiel der Klima- und Cosmic-Ray Station Gevenich im Rureinzugsgebiet.

5 Zusammenfassung und Ausblick

Im Rahmen der HGF-Infrastrukturmaßnahme TERENO werden in vier terrestrischen Umweltobservatorien große Mengen an umweltrelevanten Daten gesammelt. Um diese Datenmengen adäquat zu verwalten, zu beschreiben, auszutauschen und zu veröffentlichen, wurde die verteilte Dateninfrastruktur TEODOOR eingerichtet. Die für die Einrichtung und Betrieb der Observatorien verantwortlichen Institutionen betreiben dabei jeweils eigene lokale Dateninfrastrukturen, in denen effektive Werkzeuge für eine automatisierte Datenspeicherung und Datenarchivierung sowie für regelmäßige Plausibilitätsprüfungen der Daten zum Einsatz kommen.

Zusätzlich zu seiner Funktion als gemeinsame Informations- und Austauschplattform für den TERENO-Verbund hat das TEODOOR-Portal (<http://www.tereno.net>) die Funktion, die Daten aus den einzelnen Observatorien zusammenzuführen und der wissenschaftlichen bzw. nichtwissenschaftlichen Öffentlichkeit zugänglich zu machen. Hierzu wurden eine Reihe von Werkzeugen wie z.B. der hierarchischen Suche nach Stichworten oder durch Web-GIS- Funktionalitäten entwickelt, die es dem Benutzer erlauben, einen detaillierten Einblick in die verschiedenen Observatorien, Testgebiete und Messnetze zu erhalten. Datenvisualisierungswerkzeuge erlauben es, die Daten von verschiedenen Zeitreihen und Variablen zu selektieren, darzustellen, zu überprüfen und schließlich konform der gemeinsamen Datenpolitik auch herunter zu laden. Momentan sind zeitaktuelle Daten von mehr als 450 automatischen Stationen frei verfügbar.

Die Definition und Umsetzung geeigneter Schnittstellen und Standards für den Datenaustausch zwischen den lokalen Dateninfrastrukturen und TEODOOR ist die größte Herausforderung. In TERENO erfolgt dieser Austausch über standardisierte OGC-konforme Web- Dienste, die für jedes Observatorium aufgebaut und betrieben werden und welche die Daten und Metadaten für die interne und öffentliche Verwendung bereitstellen. Ein neu entwickelter OGC-konformer Sensor Observation Service (SOS) ist in der Lage, Rasterzeitreihendaten liefern zu können, und eröffnet damit eine Reihe erweiterter Funktionalitäten für die Veröffentlichung von Wetterradar- und Fernerkundungsdaten.

6 Referenzen

- 52°North (2012): Sensor Web Community. 52°North - Initiative for Geospatial Open Source Software GmbH, <http://52north.org/communities/sensorweb/>, accessed 2012-08-30.
- Band, L., Ogden, F., Goodrich, D., Hooper, R., Kane, D., Lyons, B., McKnight, D., Miller, N., Williams, M., Potter, K., Scanlon, B., Pielke, R.A., Sr., Reckhow, K. (2005): Designing a network of hydrologic observatories as a community service. Technical Report Number 7, CUAHSI, pp. 44.
- Bogena, H.R., Herbst, M., Huisman, J.A., Rosenbaum, U., Weuthen, A., Vereecken, H. (2010): Potential of Wireless Sensor Networks for Measuring Soil Water

- Content Variability. *Vadose Zone Journal* (4): 1002-1013, doi: 10.2136/vzj2009.0173.
- Bogena, H.R., Schulz, K., Vereecken, H. (2006): TERENO: Towards a network of observatories in terrestrial environmental research. *Advances in Geosciences*: 109-114
- Bose, R., 2002. A conceptual framework for composing and managing scientific data lineage, 14th International Conference on Scientific and Statistical Database Management. IEEE Press, Piscataway, N. J., pp. 15-19.
- Botts, M. (Ed.), (2007): OpenGIS Sensor Model Language (SensorML) - Implementation Specification. DocNr. OGC-07-000, Ver 1.0.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 180.
- Botts, M., Percivall, G., Reed, C., Davidson, J. (2008): OGC Sensor Web enablement: Overview and high level architecture. In: Nittel, S., Labrinidis, A., Stefanidis, A. (Eds.), *Geosensor Networks. Lecture Notes in Computer Science*. Springer-Verlag Berlin, Berlin, pp. 175-190.
- Bronstert, A., Kneis, D., Bogena, H. (2009): Interactions and feedbacks in hydrological change: Relevance and possibilities of modelling. *Hydrologie und Wasserbewirtschaftung* (5): 289-304
- Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., Lemmens, R. (2011a): New Generation Sensor Web Enablement. *Sensors* (3): 2652-2699, doi: doi:10.3390/s110302652.
- Bröring, A., Maué, P., Janowicz, K., Nüst, D., Malewski, C. (2011b): Semantically-Enabled Sensor Plug & Play for the Sensor Web. *Sensors* (8): 7568-7605
- Bröring, A., Stasch, C., Echterhoff, J. (Eds.) (2012): Sensor Observation Service Interface Standard. DocNr. OGC 12-006, Ver 2.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 163.
- Burt, T.P., Howden, N.J.K., Worrall, F., Whelan, M.J. (2008): Importance of long-term monitoring for detecting environmental change: lessons from a lowland river in South East England. *Biogeosciences*: 1529-1535
- Chen, N.C., Di, L.P., Yu, G.N., Gong, J.Y., Wei, Y.X. (2009): Use of ebRIM-based CSW with sensor observation services for registry and discovery of remote-sensing observations. *Computers & Geosciences* (2): 360-372, doi: 10.1016/j.cageo.2008.08.003.
- Cox, S. (Ed.), (2010): Geographic Information: Observations and Measurements - OGC Abstract Specification Topic 20. DocNr. OGC 10-004r3, ISO 19156, Ver 2.0.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 57.
- Cox, S. (Ed.), (2011): Observations and Measurements - XML Implementation. DocNr. OGC 10-025r1, Ver 2.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 76.
- Desilets, D., Zreda, M., Ferre, T.P.A. (2010): Nature's neutron probe: Land surface hydrology at an elusive scale with cosmic rays. *Water Resources Research*, doi: W11505 10.1029/2009wr008726.
- ECJRC (Ed.), (2009): INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119. European Commission Joint Research Centre, pp. 74.
- FGDC (Ed.), (2000): Content Standard for Digital Geospatial Metadata Workbook. Federal Geographic Data Committee, Reston, Virginia, USA, pp. 126.
- Gray, J., Liu, D., Nieto-Santisteban, T.M., Szalay, A., DeWitt, D.J., Heber, G. (2005): Scientific data management in the coming decade. *SIGMOD Rec.* (4): 34-41, doi: doi:10.1145/1107499.1107503.

- Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I. (2008): A relational model for environmental and water resources data. *Water Resources Research* (5), doi: W05406 10.1029/2007wr006392.
- ISO (Ed.), (2003): ISO/DIS 19115: Geographic information - Metadata. DocNr. ICS 35.240.70, Ver ISO/DIS 19115. International Organization for Standardization, pp. 163.
- ISO (Ed.), (2009): Information and documentation - The Dublin Core metadata element set. DocNr. ISO 15836:2009. ISO, pp. 6.
- Jirka, S., Bröring, A., Kjeld, P., Maidens, J., Wytzisk, A. (2012): A Lightweight Approach for the Sensor Observation Service to Share Environmental Data across Europe. *Transactions in GIS* (3): 293-312
- Jirka, S., Broring, A., Stasch, C. (2009): Discovery Mechanisms for the Sensor Web. *Sensors* (4): 2661-2681, doi: 10.3390/s90402661.
- Keller, M., Schimel, D.S., Hangrove, W.W., Hoffman, F.M. (2008): A continental strategy for the National Ecological Observatory Network. *Frontiers in Ecology and the Environment* (282-284)
- kZen Labs (2012): Buddata ebXML Registry/Repository, <http://www.buddata-open.org/>, accessed 2012-08-30.
- Lesage, N. (Ed.), (2007): OGC Cataloguing of ISO Metadata (CIM) - Using the ebRIM profile of CS-W. DocNr. OGC 07-038, Ver 0.1.7. Open Geospatial Consortium, pp. 95.
- Lin, H. (2003): Hydropedology: bridging disciplines, scales, and data. *Vadoze Zone Journal*: 1-11
- Lin, H. (2010): Earth's Critical Zone and hydropedology: concepts, characteristics, and advances. *Hydrol. Earth Syst. Sci.*: 25-45
- McDonnell, J.J., Sivapalan, M., Vachè, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M.L., Selker, J., Weiler, M. (2007): Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resour. Res.* (7): W07301, doi: 10.1029/2006wr005467.
- Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G. (1997): Nongeospatial metadata for the ecological sciences. *Ecological Applications* (1): 330-342, doi: 10.2307/2269427.
- Montgomery, J.L., Harmon, T., Haas, C.N., Hooper, R., Clesceri, N.L., Graham, W., Kaiser, W., Sanderson, A., Minsker, B., Schnoor, J., Brezonik, Patrick (2007): The WATERS Network: An Integrated Environmental Observatory Network for Water Research. *Environmental Science & Technology* (19): 6642-6647, doi: 10.1021/es072618f.
- Na, A., Priest, M. (Eds.) (2007): Sensor Observation Service. DocNr. OGC 06-009r6, Ver 1.0. Open Geospatial Consortium Inc., Wayland, MA, USA, pp. 104.
- Nebert, D., Whiteside, A., Vretanos, P. (Eds.) (2007): OpenGIS Catalogue Services Specification. DocNr. OGC 07-006r1, Ver 2.0.2, Corrigendum 2 Release. Open Geospatial Consortium, Wayland, MA, USA, pp. 218.
- Nisbet, E. (2007): Earth monitoring: Cinderella science. *Nature*: 789-790
- Nogueras-Iso, J., Zarazaga-Soria, F.J., Béjar, R., Álvarez, P.J., Muro-Medrano, P.R. (2005): OGC Catalog Services: a key element for the development of Spatial Data Infrastructures. *Computers & Geosciences* (2): 199-209, doi: 10.1016/j.cageo.2004.05.015.

- OASIS (Ed.), (2002): OASIS/ebXML Registry Information Model. Ver 2.0. The Organization for the Advancement of Structured Information 1640 Standards [OASIS], pp. 60.
- OGC (2012): OGC Standards and Supporting Documents. Open Geospatial Consortium, <http://www.opengeospatial.org/standards>, accessed 2012-08-20.
- Parr, T.W., Sier, A.R.J., Battarbee, R.W., Mackay, A., Burgess, J. (2003): Detecting environmental change: science and society, Äperspectives on long-term research and monitoring in the 21st century. Science of the Total Environment: 1-8, doi: 10.1016/s0048-9697(03)00257-2.
- Plone Foundation (2004): Plone, <http://www.plone.org>, accessed 2012-08-30.
- Reid, W.V., Bréchnignac, C., Tseh Lee, Y. (2009): Earth System Research Priorities. Science (5938): 245-245
- Richter, D.d., Mobley, M.L. (2009): Monitoring Earth's Critical Zone. Science (5956): 1067-1068
- Simonis, I., Echterhoff, J. (Eds.) (2011): OGC Sensor Planning Service Implementation Standard. DocNr. OGC 09-000, Ver 2.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 248.
- Simonis, I., Wytzisk, A. (Eds.) (2003): Web Notification Service. DocNr. OGC 03-008r2, Ver 0.1.0. Open Geospatial Consortium, Wayland, MA, USA, pp. 46.
- Sorg, J. (2012): Entwurf, Implementierung und Anwendung eines OGC-konformen Sensor Observation Service für flächenbezogene Rasterzeitreihendaten. Master Thesis, Fernuniversität Hagen, Hagen, Germany, 96 pp.
- Tarboton, D.G., Horsburgh, J.S., Maidment, D.R., Whiteaker, T., Zaslavsky, I., Piasecki, M., Goodall, J., Valentine, D., Whitenack, T. (2009): Development of a Community Hydrologic Information System. 18th World Imacs Congress and Modsim09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, 988-994 pp.
- TERENO (2010): TERENO Data policy, [http://teodoor.icg.kfa-juelich.de/downloads/TERENO Data policy.pdf](http://teodoor.icg.kfa-juelich.de/downloads/TERENO_Data_policy.pdf), accessed 2012-08-20.
- Tomasic, A., Simon, E. (1997): Improving access to environmental data using context information. SIGMOD Rec. (1): 11-15, doi: 10.1145/248603.248606.
- Vanderborght, J., Graf, A., Steenpass, C., Scharnagl, B., Prolingheuer, N., Herbst, M., Franssen, H.-J.H., Vereecken, H. (2010): Within-Field Variability of Bare Soil Evaporation Derived from Eddy Covariance Measurements. Vadose Zone Journal (4): 943-954, doi: 10.2136/vzj2009.0159.
- Zacharias, S., Bogen, H., Samaniego, L., Mauder, M., Fuss, R., Puetz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., Munch, J.C., Papen, H., Priesack, E., Schmid, H.P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., Vereecken, H. (2011): A Network of Terrestrial Environmental Observatories in Germany. Vadose Zone Journal (3): 955-973, doi: 10.2136/vzj2010.0139.
- Zoback, M.L. (2001): Grand challenges in earth and environmental sciences: sciences, stewardship, and service for the twenty-first century. GSA Today: 41-47