

Geospatial Linked (Open) Data für Österreich? Einführungen, Anwendungen und Perspektiven

Francis Harvey¹ und Johannes Scholz²

¹ University of Minnesota, Geography 468 SocSci 267, 19th Ave S, Minneapolis, MN, USA;
fharvey@umn.edu

² Research Studios Austria, Studio iSPACE, Schillerstrasse 25, 5020 Salzburg, Austria;
johannes.scholz@researchstudio.at

Abstract

The paper deals with the theoretical concepts of (spatial) Linked Open Data. In addition, selected research topics related to LOD are highlighted which are currently discussed in the scientific community. Among them are new approaches to integrate Metadata in LOD, the handling of changes over time in LOD as well as the issue of diverging “truths” regarding the Universe of Discourse. Based on several application examples, the authors elaborate on the potential and challenges of LOD. Subsequently, the possibilities for LOD to serve as basis for a new paradigm for Spatial Data Infrastructures are discussed.

Kurzfassung

Das Papier befasst sich mit den grundlegenden theoretischen Konzepten hinter (spatial) Linked Open Data, und beleuchtet zudem einige ausgewählte aktuelle Forschungsfragen die in diesem Kontext auftreten. Darunter fallen neue Ansätze zur Integration von Metadaten, das Management von Veränderungen in LOD über die Zeit und unterschiedliche Sichtweisen auf das zu modellierende Universe of Discourse. Anhand von mehreren aktuellen Beispielen aus der LOD Welt werden die Möglichkeiten und Herausforderungen von LOD erläutert und das zukünftige Potential in Richtung eines möglichen neuen Paradigmas zu Geodateninfrastrukturen beleuchtet.

Keywords: Linked Open Data, Semantik, Ontologie, Semantik Web, Web 3.0,

1. Einleitung - Was ist, was kann LOD?

Linked (Open) Data (LOD) stellen die technische Grundlage möglicher grundlegender Änderungen in der Bearbeitung und Nutzung von Geodaten dar. Hierbei ist es wichtig nicht nur technologische Aspekte - wie z.B.: die Konvertierung von Datensätzen - zu betrachten, sondern auch den Einfluss auf die gesamte Prozesskette von Produktion, Wartung, „Vertrieb“ und Nutzung sowie auf die Eigenschaften von Daten selbst. Linked Data Konzepte zielen darauf ab, explizite Verbindungen (links) zwischen maschinenlesbaren Datensätzen zu etablieren. Diese Verbindungen können nach Bedarf konstruiert und dazu eingesetzt werden, neue Sachverhalte aus bestehenden Daten zu extrahieren, und neue Datenquellen zu bestimmten Sachverhalten oder einem geographischen Ort zu finden.

Bis dato, werden Daten von den nationalen oder regionalen Behörden sowie von Firmen für deren eigene Erfordernisse erzeugt und bei Bedarf auch an die Öffentlichkeit abgegeben - oft verkauft. Aus diesem Grund gibt es, unter anderem, folgende Probleme mit der zentralen Verteilung von Daten (vgl. [18]):

- o Zugang zu den Daten ist oft abhängig von der eingesetzten Software
- o Metadaten und Daten sind unabhängige Teile
- o Daten werden als "Wahrheit" aus Sicht jedes Datenherstellers angesehen, obwohl sich die Daten manchmal widersprechen
- o Es existieren zu wenige unterschiedliche zugängliche Sichtweisen (i.e. Daten) auf das zu modellierende „Universe of Discourse“
- o Daten über ein spezielles Thema, zu einem speziellen Ort und einer speziellen Zeit sind schwierig zu finden
- o Eine eindeutige homogene globale Identifizierung von Ortsnamen ist nur schwierig zu bekommen

Durch LOD können diese Probleme nicht sofort gelöst werden, jedoch stellen Sie die technischen Grundlagen dar, um anders mit diesen genannten Problemen umzugehen und ansatzweise zu lösen. LOD können, aber müssen nicht, offen und ohne Einschränkung zur Verfügung gestellt werden, d.h. Linked Data eignen sich auch für Nutzungen in Intranetzwerken. LOD ermöglichen dennoch das höchste Maß an Folgenutzung, da sie von Einschränkungen befreit werden—for viele Anwendungsgebiete ein wichtiges Ziel mit positiven betriebs- und volkswirtschaftlich Auswirkungen [24], [13], [4]. Im besonderen Maße könnten geographische Daten als Verbindungsglied zwischen Datensätzen unterschiedlicher Domänen dienen [9]. Ein hypothetisches Beispiel bezogen auf die Stadt Salzburg, veranschaulicht die Möglichkeiten und ein Vergleich mit herkömmlichen GIS Verfahren weist auf einige wichtige Änderungen durch LOD hin. Angenommen man möchte eine Anwendung für die Stadt Salzburg erstellen, die ein Geländemodell als Grundlage benötigt. Um die Grundlagendaten zu erhalten ist der Entwickler bis dato darauf angewiesen, die möglichen Anbieter von solchen räumlichen Daten zu kennen, um von diesen die Daten käuflich zu erwerben. Da die Anbieter die Geländemodelle mit unterschiedlichen Methoden und Genauigkeitsanforderungen erstellt haben, können die Datensätze verschiedene Qualitätsstufen aufweisen. Aus diesem Grund wäre es für den Entwickler der Applikation interessant einen Überblick über die verfügbaren Daten für die Stadt Salzburg zu bekommen. Zudem wäre es von grundlegendem Interesse einen Vergleich über die Qualität der angebotenen Datensätze zu bekommen, um die „richtigen“ Daten für die Applikation auswählen zu können. Zum Beispiel wäre ein ASTER oder SRTM Geländemodell für eine detailgetreue Oberflächenmodellierung der Stadt zu grob aufgelöst, während Laserscanning Daten des Bundeslandes Salzburg in hoher Auflösung zur Verfügung stünden. Dieser Datensatz wäre für den Einsatz der detaillierten Oberflächenmodellierung sehr gut geeignet. Zudem könnten andere Datensätze, die mit der Stadt Salzburg verknüpft sind – entweder über die Geographie oder ein Stichwort – zusätzlich gefunden und integriert werden.

Wichtig ist LOD als Entwicklung und Versuch von neuen technischen und organisatorischen Möglichkeiten zu sehen. Es gilt LOD als Weiterentwicklung von neueren vernetzten Fähigkeiten, bekannt mit dem Begriff Web 2.0 oder Web 3.0, zu verstehen. Einige Punkte in der Entwicklung

von LOD konnten im Zuge des „*Linked Open Data Workshops @ GI_Forum 2014*“ in Salzburg herausgearbeitet und diskutiert werden
(<http://www.johannesscholz.net/LinkedOpenData2014.htm>):

- Datenmengen und Angemessenheit
- Änderungen der kartographischen Ansätze und digitalen Repräsentation: „Das Ende der Layer-Metapher in räumlichen Informationssystemen“
- Ontologie, Vocabularies und/oder Patterns: Beiträge der Philosophie und/oder Ingenieure zu Wissenschaft
- Komplexität der Schnittstellen und analytische Schärfe
- Integration von vernetzten Informationen und Suchfähigkeiten um neue Anwendergruppen zu erreichen

Das Papier gliedert sich wie folgt. In dem nächsten Abschnitt werden wir die theoretischen Konzepte von LOD vorstellen sowie wissenschaftliche Fragen diskutieren, die für die praktische Umsetzung von Interesse sind. Darauf aufbauend werden in Kapitel 3 Beispiele aus der Anwendungswelt erläutert, die die Konzepte von LOD in die Realität umsetzen. Eine Zusammenfassung des Papiers sowie eine Sammlung der zukünftigen Entwicklungen finden sich in Kapitel 4.

2. Überblick von Konzepten

Der Begriff LOD umfasst eine Vielzahl von Technologien und Prinzipien, die die Realisierung eines vollkommen neuen Paradigmas zur Publikation, Abfrage, Wiederverwertbarkeit und Integration von Daten ermöglichen. Anders als im „klassischen“ Internet – welches dokumentenzentriert ist – werden im LOD – Web Links zwischen den Daten erstellt. Dabei geht es primär nicht um Links zwischen menschenlesbaren Webseiten, die für Maschinen schwierig zu interpretieren sind, sondern um explizite Links zwischen rein maschinenlesbaren Datensätzen. Dies ist deshalb notwendig, da Webseiten und andere, für den Menschen interpretierbare, Daten im Internet, für Computer nur eingeschränkt automatisiert interpretierbar sind. Dies ist jedoch eine Voraussetzung um das Konzept „Semantic Web“ und Web 3.0 realisieren zu können. Das Semantic Web beschreibt eine Erweiterung des herkömmlichen WWW, um die Kontextualisierung von Informationen – i.e. werden Informationen mit eindeutigen Bedeutungen versehen.

Damit kann eine Zusammenarbeit von Mensch und Maschine ermöglicht werden, wie von Berners-Lee et al. [3] definiert: „*The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.*“ Um diese Vision zu realisieren, können Konzepte der Künstlichen Intelligenz eingesetzt werden, die bereits seit einigen Jahren im Fokus der Forschung liegen. Dies wird auch zusammenfassend von Heath und Bizer [11] so beschrieben: „*For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. Artificial Intelligence researchers have studied such systems since long before the Web was developed. (...) The challenge of the*

Semantic Web, therefore, is to provide a language that expresses both data and rules for reasoning about the data and that allows rules from any existing knowledge-representation system to be exported onto the Web.”.

Als weiteres Beispiel kann die Stadt Salzburg, die innerhalb des Bundeslandes Salzburg liegt, dienen. Der Link zwischen diesen beiden Daten würde im LOD Paradigma explizit, gerichtet und annotiert veröffentlicht werden. Damit kann direkt ausgesagt werden, dass die Stadt Salzburg im Bundesland Salzburg liegt, was mit dem Prädikat „*LiegtIn*“ unterstrichen wird. Zusätzlich ist es von Vorteil, die Datensätze mit eindeutigen Datentypen zu versehen, um die Aussagekraft und Verknüpfbarkeit zu erhöhen. Für das eben genannte Beispiel, wären das eine Stadt und das Bundesland – da das Wort „Salzburg“ potentiell eine Instanz beider Typen sein kann.

Zusammengefasst kann man jedes Statement in LOD als Tripel auffassen, welches aus Subjekt, Prädikat und Objekt besteht – i.e. die Stadt Salzburg (*Subjekt*) LiegtIn (*Prädikat*) dem Bundesland Salzburg (*Objekt*). Um solche sprachlichen Elemente entsprechend digital zu modellieren und repräsentieren, gibt es das *Resource Description Framework* (RDF). Berners-Lee [2] definiert vier Prinzipien von LOD:

- Uniform Resource Identifiers (URIs) bezeichnen Dinge direkt
- HTTP URIs werden dafür verwendet um Ressource eindeutig zu identifizieren und dereferenzierbar – i.e. auffindbar – zu machen (von Computern und Menschen)
- RDF und OWL – als aktuelle W3C Standards – sollten verwendet werden, wenn Informationen hinter URIs bereitgestellt werden.
- Daten, die reale Objekte beschreiben, sollten zu Daten von anderen Objekten weiterverweisen (unter der Prämisse, dass URIs verwendet werden)

Basierend auf den genannten Prinzipien stellen URIs einen einheitlichen Bezeichner für alle Dinge dar, die mit LOD beschrieben werden sollen. Darunter fallen Webseiten, Dokumente und/oder sonstige Dateien, sowie Dinge außerhalb des World Wide Webs, wie real existierende Objekte und (abstrakte) Konzepte. Da reale Objekte keine Informationsquelle im eigentlichen Sinne sind, ist es nicht möglich Informationen zu den Objekten mit Browsern oder anderen Werkzeugen abzufragen. Dennoch kann eine digitale Informationsquelle zu dem Objekt – als RDF – angeboten werden. Für die Stadt Salzburg könnte dies – unter Verwendung des offenen SPARQL Endpunktes <http://live.dbpedia.org>, und des darauf aufbauenden Namespaces – die folgende URI sein: dbpedia:Salzburg. Diese URI ist eine Abkürzung für die folgende URI: <http://live.dbpedia.org/page/Salzburg>. Das Subjekt „Salzburg“ wird nun mit weiteren Objekten (i.e. URIs) über Prädikate verbunden, wie bereits oben erwähnt. Die Prädikate können, wie im Falle „Salzburg“, z.B. folgendes umfassen: z.B. geo:lat, geo:lon, prop:bundesland, prop:mayor. Die Objekte würden in den oben genannten Fällen zu den Geographischen Koordinaten führen, das Bundesland auf dbpedia:Salzburg_(state) sowie auf den Bürgermeister dbpedia:Heinz_Schaden verweisen. Durch das hinzufügen von weiteren Prädikaten und Objekten kann eine dicht vernetzte Linked Data Wolke aufgebaut werden, die eine reichhaltige Informationsquelle bereitstellt.

Das *Resource Description Framework Schema* (RDFS) ermöglicht die Definition von Klassen, die dazu eingesetzt werden können, um Ressourcen eindeutigen Typen zuzuordnen. Zudem

können Festlegungen von properties getroffen werden – die Definition von validen Verbindungen zwischen Ressourcen. Mittels RDFS kann eine Semantik von Klassen in Form einer Taxonomie – als Sub-Klasse – aufgebaut und einfache Restriktionen, wie Kardinalität, Range und Domain können definiert werden.

SPARQL ist die Abfragesprache für LOD, im Speziellen für RDF. Eine geographische Erweiterung – geoSPARQL – erlaubt es, topologische Relationen in Abfragen einzubauen [23]. Damit können z.B.: Objekte in einer Bounding Box abgefragt werden. Auch komplexere Abfragen sind möglich wie z.B.: alle Wasserflächen innerhalb 1km um die Route 3

```
SELECT ?water ?wWKT
WHERE { ?water rdf:type :WaterBody .
?water :exactGeometry ?wGeo .
?wGeo ogc:asWKT ?wWKT .
:Route_3 :exactGeometry ?r3Geo .
:r3Geo ogc:asWKT ?r3WKT .
FILTER(ogcf:distance(?r3WKT, ?wWKT, ogc:km) <= 1)
```

Es existieren nur wenige Implementierungen von SPARQL Endpunkten, die geoSPARQL unterstützen. Projekte die geoSPARQL unterstützen sind: Parliament, Sesame via uSeekM Add-on und zum Teil Strabon.

Die abstrakten technischen LOD Konzepte beschreiben jedoch eine Veränderung des traditionellen Weges um Karten zu erstellen und räumliche Information zu generieren. Durch die Vielzahl von LOD gibt es eine Menge von Daten, die miteinander in Verbindung gebracht werden können. Deshalb ist es erforderlich, aus der Vielzahl an unterschiedlichen Daten die für den Anwendungsfall „richtigen“ herauszufiltern. Dies muss einerseits im Hinblick auf die thematische Ebene sowie auf qualitativer Ebene erfolgen. Besonders die qualitative Ebene ist bei weiterführenden Analysen von besonderer Bedeutung, da diese maßgeblichen Einfluss auf die Qualität der Aussagen/Resultate hat. Wichtig in diesem Zusammenhang ist auch zu sehen, dass LOD eine Modellierung der Welt widerspiegeln, und damit unterschiedliche Datensätze verschiedene Sichtweisen ergeben [8]. Dies kann als direkte Konsequenz des Ausspruchs von Allemang und Hendler [1] „*Anyone can say Anything about Any topic*“ gelten, was von Janowicz (2010) noch verschärft wurde: „*Anyone can say Anything about Any topic at Any time at and Anywhere*“. Diese Aussagen verdeutlichen, dass es zeitgleich unendliche Verknüpfungs-Links durch LOD geben kann und wird. Die Wahl des passenden Datensatzes aus der Vielzahl von verfügbaren Daten (sowohl thematisch als auch zeitlich) für einen gegebenen Use Case ist daher von größter Bedeutung.

Diese Sichtweisen können sich auch über die Zeit verändern, da die Welt in keinem statischen Zustand verharrt. Während für „einfache“ Phänomene, wie Niederschlag oder Temperatur, die Veränderungen in Form von Snapshots abgebildet werden können, ist dies bei manchen komplexeren Prozessen nicht so einfach möglich. Dies deshalb, da diese Veränderungen auch nachfolgende Schlussfolgerungen grundlegend verändern. Dies kann aufgrund von verschiedenen Änderungsprozessen passieren, wie z.B.: Grenzverschiebungen (z.B. administrative Regionen),

Verbindungen zwischen Objekten werden verändert (z.B. ein neuer Bürgermeister wird in der Stadt Salzburg eingesetzt), geänderte Grundkonzepte (vgl. [18]). Eine Auflistung von Veränderungen in zeitlich-räumlichen Systemen kann z.B. in Medak [20] sowie Kauppinen und Hyvonen [16] gefunden werden. Für LOD erscheint es daher wichtig, die Gültigkeitsdauer eines LOD Statements zu vermerken. Damit könnte die zeitliche Abfolge von gültigen Ortsnamen für einen Ort nachvollziehbar dargestellt werden, oder administrative Grenzen mit Gültigkeitsintervallen versehen werden und mit Objekten gelinkt werden. Damit wird es möglich diese Veränderungen eindeutig zu monitoren und historische Schlussfolgerungen und Analyseergebnisse nachvollziehbar zu gestalten.

Nach [9] sind LOD im Grunde nichts anderes als Behauptungen die von Personen/Organisationen gemacht werden - vgl. dazu [8] „*Anyone can say Anything about Any topic*“. LOD sind daher einfache Statements in RDF-Tripel Struktur (Subjekt - Prädikat - Objekt). LOD eignen sich daher nicht gut um komplexe relationale Datenmodelle - wie sie in der Datenbank-Welt vorherrschen - abzubilden. Da jedoch räumliche Daten zumeist in komplexen Datenmodellen gespeichert werden, müssen diese Datenmodelle in geeigneter Form abstrahiert werden, um die Grundidee von LOD zu erfüllen. Wenn wir den komplexen Datenmodellen unterstellen als komplexe (isolierte?) Ontologie das gegebene “Universe of Discourse” abzubilden, dann kann es schwierig werden Datensätze miteinander in Verbindung zu bringen (i.e. zu linken). Die Verlinkung von Datensätzen beruht auf der Ähnlichkeit der Semantik eines Konzeptes/Begriffs oder Datensatzes, z.B.: Daten über den Bürgermeister der Stadt Salzburg, können mit der Stadt Salzburg in Verbindung gebracht werden, wohingegen der Landeshauptmann des Bundeslandes Salzburg primär mit dem Bundesland Salzburg verbunden ist. Die automatisierte Detektion von solchen Verbindungen ist - bei komplexen Datenmodellen der Ausgangsdaten - zunehmend schwierig zu bewerkstelligen, da eine Vielzahl von unterschiedlichen semantischen Konzepten “aufeinanderprallt”. Aus diesem Grund wird zunehmend versucht auf standardisierte Vocabularies zurückzugreifen, um die Daten semantisch auszuzeichnen.

Zudem sind die Open-Daten und deren Links sind nicht auf die im eigenen Haus beschränkt. Um nicht nur Daten zu verlinken, sondern auch Personen, deren Aktivitäten und Verbindungen zu anderen Personen und Objekten, existiert das Projekt “*Friend of a Friend*” (FOAF) (www.foaf-project.org). FOAF verwendet RDF und XML um das Hinzufügen von persönlichen Metadaten zu FOAF Profilen zu ermöglichen, womit eine maschinenlesbare Ontologie gefüllt wird. FOAF schafft daher eine dezentrale Architektur die Social Media Webseiten und die Personen darin miteinander verknüpft, und damit personenbezogene Applikationen unterstützen kann.

Ein anderer Ansatz verfolgt die Entwicklung von “(geo-)Ontology Design Patterns (ODPs)“ (z.B.: [12], [6]), die eine einheitliche Konzeptualisierung eines eingeschränkten Ausschnitts der Realwelt darstellen. ODPs werden als konsensuales Ergebnis von Domain Experten und Ingenieuren aus dem Bereich Wissensmodellierung gesehen, und können als Grundbausteine für den Aufbau von weiteren, möglicherweise komplexeren Ontologien gesehen werden. ODPs sind grundsätzlich in sich abgeschlossen und sind für die Wissensmodellierung von Domänen und Applikationen geeignet - besonders in den Fällen wo man abstrakte und/oder komplexe

Ontologien nicht oder nur schwer zur Modellierung des “Universe of Discourse” einsetzen kann. Die ODPs stellen somit “Mini-Ontologien” dar, die gut dokumentiert sind, leicht generalisiert werden können und wiederverwendet und mit anderen ODPs kombiniert werden sollen. Solche ODPs wurden und werden im Zuge von Geo-Vocabulary Camps (http://vocamp.org/wiki/Main_Page) erstellt und getestet. Wir unterscheiden zwei Arten von ODPs [6]:

- Logische - ODPs: befassen sich mit der formalen Beschreibung der Semantik, und sind daher unabhängig von der Domäne;
- Inhalt - ODPs: dienen der Beschreibung des Domänenwissens;

Bisher wurden Daten und Metadaten immer als getrennte Entitäten gesehen worden. Da Daten und Metadaten zumeist in separaten Datenformaten und Speicherorten erfasst und gespeichert werden, ist die Erfassung der Metadaten häufig als nachrangig erachtet worden - sobald die Daten in genügend hoher Qualität erfasst worden sind. Beim Suchen von Daten bedient man sich jedoch häufig der (oftmals “lückig” erfassten) Metadaten, wodurch die Suchresultate die Benutzer nicht zufrieden stellen. LOD stellen einen anderen Ansatz zur Integration von Metadaten dar. Da Metadaten auch als RDF-Datensätze gespeichert werden können, ist es möglich Beziehungen zwischen den Daten und Metadaten herzustellen und diese gemeinsam via SPARQL Endpunkten abzufragen. Hu et al. [12] meinen außerdem, dass Metadaten auch zeitlich-räumlicher Natur sein können - unabhängig davon ob die Daten zeitlich-räumlicher Natur sind. Durch die Integration der (räumlichen) Metadaten in die Daten, können diese gemeinsam mit den eigentlichen Daten mit Methoden der Geoinformatik analysiert werden - was einen generellen Strukturwandel bei Geoportalen nach sich ziehen kann [14]. Der Wandel wird durch die integrale Sicht auf die Daten und Metadaten initiiert sowie durch die Ausführung von zeitlich-räumlichen Abfragemöglichkeiten - nicht nur auf einzelne Datensätze - sondern auf ganze Geodateninfrastrukturen eingeleitet.

3. Beispiele aus der Anwendungswelt

In diesem Kapitel versuchen die Autoren die theoretischen Konzepte aus Kapitel 2, anhand entsprechender Praxisbeispiele aus der Anwendungswelt, zu erläutern und deren praktischen Nutzen zu verdeutlichen. Das Kapitel beleuchtet vier Anwendungsbereiche, die hier exemplarisch diskutiert werden. Zum ersten wird das Open Governmental Data Portal data.gv.at eingehend diskutiert, gefolgt von Anwendungen von Tomi Kauppinen. Linked Data Projekte von Krzysztof Janowicz bilden den nächsten Schwerpunkt und sollen zeigen, wie fortgeschritten LOD auch im Bereich der Anwendung bereits sind, gefolgt von dem reichhaltigen LOD Angebot der Verwaltung von Großbritannien (UK).

3.1. Data.gv.at - die österreichische Open Governmental Data Platform

Die österreichische Open Governmental Plattform ist eine Initiative von mehreren Städten und dem Bundeskanzleramt, die als zentraler „Österreich“-Katalog die Metadaten der dezentralen Datenkataloge in Österreich in sich aufnehmen und die Datensätze manuell wie automatisiert abrufbar halten soll. Darunter fallen die Daten der Städte und Gemeinden, der Bundesländer, dem

Bund und von ausgelagerten Organisationen (wie z.B. Umweltbundesamt oder Bundesrechenzentrum).

Die Plattform stellt im Moment (Stand: Juli 2014) eine Sammlung von 1398 Datensätzen von unterschiedlichen Dienststellen in z.T. mehreren Formaten (z.B.: pdf, xls, doc, html, etc.). Zu diesem Zeitpunkt existieren nur wenige RDF Datensätze in dem Datenbestand, da die Einführung von LOD aktuell in einem Pilotprojekt bearbeitet wird [15]. Diese Datensätze werden in dem offenen Datenportal CKAN [21] gespeichert und publiziert.

Zu jedem Datensatz existiert ein separater Metadatensatz der in einem entwickelten Metadatenstandard [7]. Der Metadatenstandard OGD 2.2. umfasst 11 verpflichtende sowie 22 optionale Metadatenelemente. Die gesammelten Metadaten zu einem Datensatz und den verfügbaren Datenformaten können über die CKAN Schnittstelle abgerufen werden. Für die gesamten Metadaten zum Datensatz "Bombenverdachtsflächen in der Stadt Salzburg" ist die folgende URI verfügbar: http://www.data.gv.at/katalog/api/3/action/package_show?id=bomben-verdachtsfl-chen. Die Metadaten können auch bei der Suche nach Datensätzen durchsucht werden. Dennoch sind diese nicht vollständig integriert in den Datensatz.

Basierend auf diesen offenen Daten wurden 244 Applikationen (Stand Juli 2014) von der User Community entwickelt und auf der Plattform data.gv.at der Öffentlichkeit zur Verfügung gestellt. Die Applikationen sind für Mobiltelefone und/oder Desktop PCs entworfen und nutzen die offenen Daten von data.gv.at. Die Apps drehen sich um z.B. Touristik, wie auch profane Dinge des Lebens (Radwege, Spielplätze, Hunde-Stationen), räumliche Daten (e.g. Geländemodelle).

3.2. Anwendungen von Tomi Kauppinen, Krzysztof Janowicz und Karsten Kessler

Weitere Anwendungen von Linked Data wurden im Rahmen von wissenschaftlichen Projekten erarbeitet, von denen wir zwei hier vorstellen wollen, da sie auch für die Anwender - Community von Nutzen erscheinen. Die erste Anwendung beschreibt den Einsatz von Linked Open Data bei Naturkatastrophen, während das zweite Beispiel das Verlinken von Sensor Daten beschreibt - was ebenfalls mit LD Strategien erfolgen kann.

LOD für das Management von Naturkatastrophen im weiteren Sinne wurde von z.B.: Ortmann et al. [22] beschrieben, und die Applikation ist unter <http://observedchange.com/demos/linked-haiti/> online abrufbar. Das Paper beschreibt den Paradigmenwechsel bei der Datenakquise nach Naturkatastrophen, der sich darin manifestiert, dass lokale Personen, also auch weltweit verstreute Personen versuchen aktuelle Daten und Informationen zu zeitnah sammeln und bereit zu stellen - und die Datensammlung nicht nur den staatlichen Organen/Stellen überlassen wird. Problematisch ist in die Datensammlung durch Freiwillige deshalb, da die Daten unstrukturiert erhoben und bereitgestellt werden, was eine weitere Analyse und Datenintegration schwierig macht. Deshalb schlägt das Papier - anhand der entwickelten Ushahidi Haiti Plattform - die Anforderungen und die Werkzeuge für eine LOD Plattform vor, sodass Freiwillige zur Datensammlung bei Katastrophen beitragen können.

Um die Komplexität des Krisenmanagements nach Katastrophen zu modellieren und analysieren - anhand des Beispiels Haiti - werden die Linked Data der Beobachtungen der ansässigen

Bevölkerung anhand eines standardisierten Vokabulars kategorisiert. Das verwendete Vokabular verwendet das *Management of a Crisis vocabulary* (MOAC) [19].

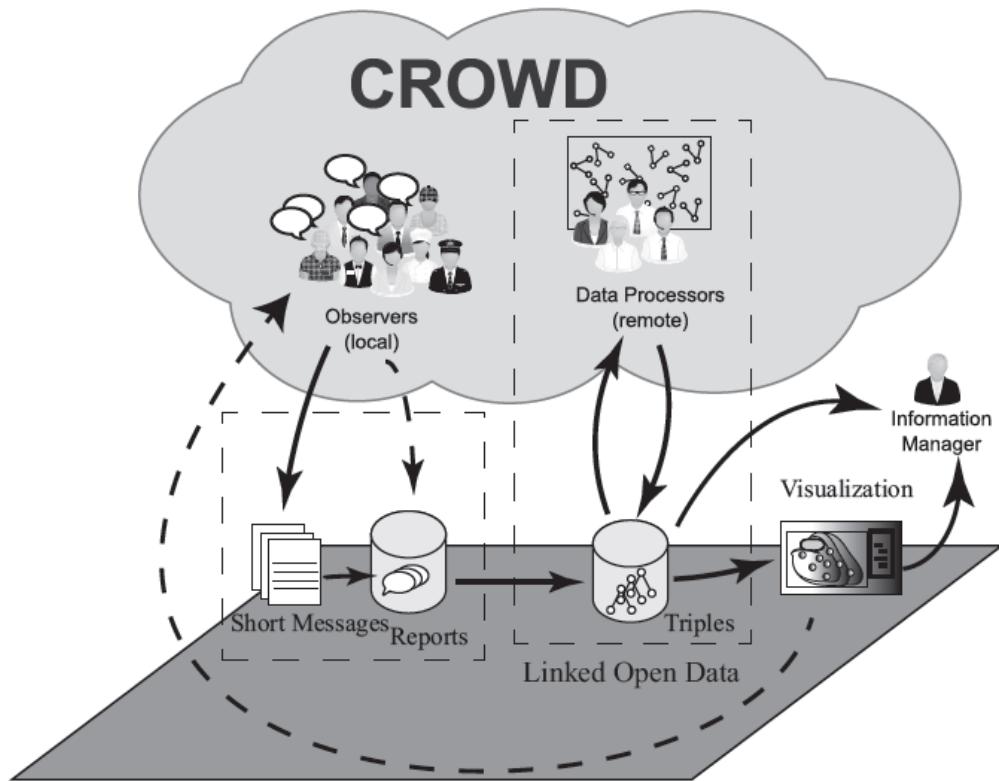


Abb. 1.:

Workflow zur Integration von crowd-sourced LOD. Die gesammelten Daten (Nachrichten und Reports) werden angereichert und als RDF Tripel zur Verfügung gestellt. Eine nachfolgende Visualisierung dient zur Information des Managements und der lokalen Bevölkerung (Quelle: [22]).

Linked Haiti Data: Reports

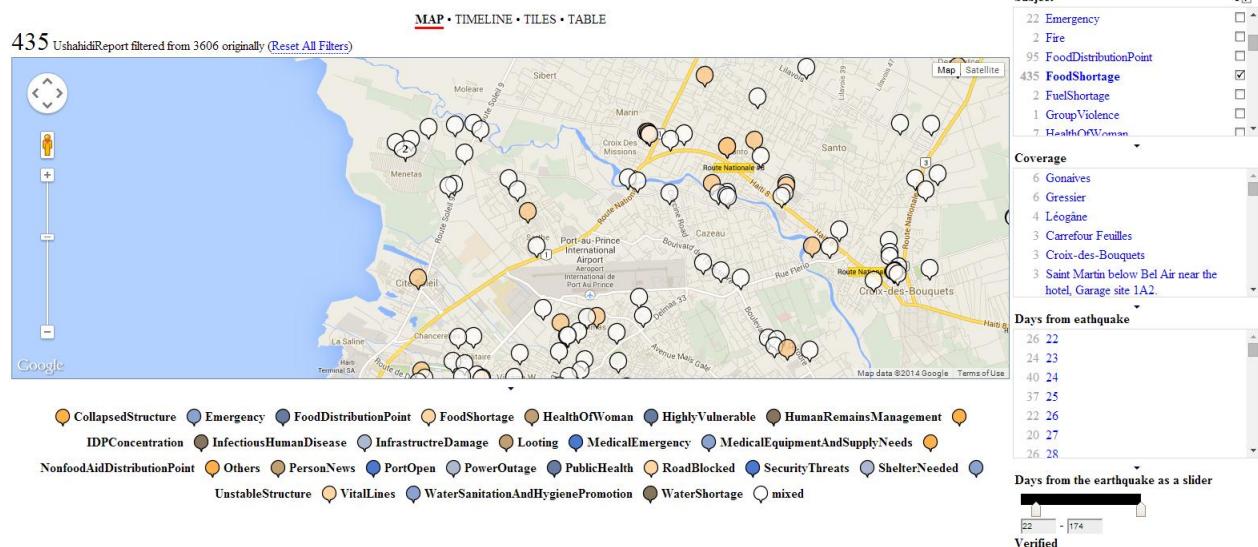


Abb. 2.: Ein Screenshot der visuellen Darstellung Haiti Applikation (<http://observedchange.com/demos/linked-haiti/>), unter Verwendung einer Kartenmetapher. Ein Klick auf einen Marker öffnet das RDF Tripel für den Report.

Ein Projekt von Krzysztof Janowicz und Karsten Kessler beschäftigt sich mit dem verlinken von Sensor Daten aus Sensorwolken - wie der OGC Sensor Web Enablement Initiative [17]. Die Sensor Web Enablement Initiative ermöglicht es Clients - die dem Standard entsprechend aufgebaut sind - auf die Daten der Sensoren zuzugreifen, jedoch kann auf anderes Wissen aus den vorhandenen Infrastrukturen nicht oder nur eingeschränkt zugegriffen werden - z.B. für einen Vergleich der Sensordaten. Keßler und Janowicz [17] haben daher einen Weg vorgeschlagen, um Sensordaten für alle Clients verfügbar zu machen - unter der Verwendung von LOD Prinzipien und RESTful services. Wichtig für die Integration von Sensordaten in die Linked Data Cloud ist jedoch die Einordnung anhand der Semantik der Daten und die Erstellung von geeigneten Links (siehe Abb. 3). Um das zu erreichen, wird von Janowicz und Kessler (2010) ein Ansatz vorgeschlagen um die Qualität der Links zu verbessern. Nach der Konvertierung der Sensordaten nach RDF, werden Schlüsselwörter extrahiert und mit den vorherigen LOD Quellen verglichen. Daraus werden Vorschläge für Übereinstimmungen automatisch erarbeitet und dem User vorgeschlagen, der diese dann verifiziert und verbessert. Diese verbesserten Links werden auch als Input für die Ähnlichkeitsbestimmung verwendet, die ähnliche Konzepte und Elemente in der LOD Cloud findet - die wieder Grundlage von Vorschlägen sind.

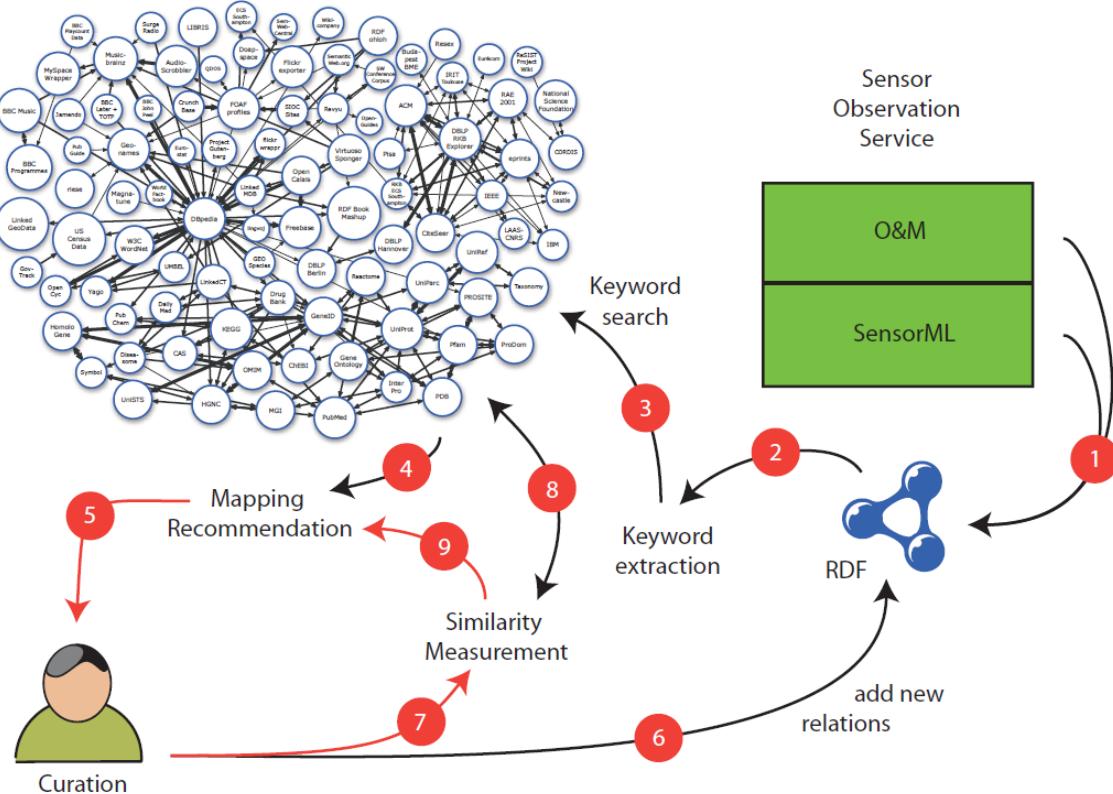


Abb. 3.: Vorgangsweise bei der Integration von Sensordaten in die Linked Data Cloud (Quelle: [17]).

3.3 LOD in Großbritannien

Verwaltungsdaten werden in Großbritannien auf dem Portal [data.gov.uk](#) veröffentlicht, was ähnlich dem österreichischen Pendant [data.gv.at](#) auf einer CKAN Installation beruht. Im Vergleich ist das Portal des Vereinigten Königreichs jedoch mit einem reichhaltigen Datenbestand bestückt (ca. 19000 Datensätze per Juli 2014). Auf dem Portal der britischen Verwaltung werden momentan ca. 340 Applikationen gelistet, die direkt auf dem Datenbestand beruhen. Um das Portal weiterzuentwickeln und die Nutzbarkeit der Daten für die Öffentlichkeit zu erhöhen, werden seit mehreren Jahren Anstrengungen unternommen, um die offenen Verwaltungsdaten in die LOD-Welt zu integrieren. Shadbolt et al. [25] und Sheridan et al. [26] beschreiben die Herangehensweise, die Herausforderungen und die Lehren des Projektes.

Diese Lehren aus dem Projekt, können gleichermaßen für das österreichische Portal [data.gv.at](#) gelten. Problematisch ist bis dato, dass viele Datensätze nicht direkt zugänglich sind. Sehr oft müssen die Datensätze - obwohl in CKAN angelegt - in HTML Seiten gesucht und gefunden werden. Das ist problematisch, wenn Datensätze (semi-)automatisiert verlinkt werden sollten. Interessant ist auch die Beobachtung von Shadbolt et al. [25], dass Metadaten - und deren Standards - stark von den Anforderungen der Datenerzeuger geprägt sind und weniger die Interessen der Nutzer von LOD abdecken. Dies erscheint unabdingbar notwendig, um neue Nutzerkreise von LOD zu erschließen. Eine direktere Integration der Metadaten in die Daten - wie von Kuhn et al. [18] und im Kapitel 2 angesprochen, würde dieses Anliegen durchaus unterstützen, und für deine Vereinfachung vor allem aus Sicht der Nutzer führen.

Problematisch bei LOD erscheinen aus Sicht von [data.gov.uk](#) die temporalen und geographischen Lücken in den Daten, die die Nützlichkeit von Links zwischen Datensätzen erheblich einschränken. Aus dem Grund erscheint es notwendig die zeitliche Komponente (i.e. zeitliche Gültigkeit, Versionierung) in LOD explizit zu erfassen und offen zu legen, um die Benutzer entsprechend zu unterstützen.

Eine Herausforderung bei der Generierung von LOD aus den Open Government Daten in UK war der semantische Abgleich - i.e. das finden von Begriffen ähnlicher Semantik. Da viele Datensätze der Verwaltung grundsätzlich nicht für den Zweck LOD entwickelt worden sind, sind auch transparente Semantiken (und Ontologien) nicht oder nur eingeschränkt verfügbar. Das erschwert das Verbinden von Datensätzen erheblich - vor allem wenn die Verlinkung (semi-) automatisiert ablaufen soll/muss.

Grundsätzlich erscheint die Verknüpfung von Open Government Daten und LOD als eine Initiative die beiden Welten Vorteile bringt, da einerseits die LOD Welt mit hochqualitativen Verwaltungsdaten angereichert wird, und andererseits müssen die Herausforderungen zur

Integration von heterogenen Daten mit wenig Struktur und geringer inhärenter Semantik gelöst werden, was in einem breiteren abstrakteren Kontext von Interesse ist.

3.5. Publicdata.eu

Eine Antwort auf die Pan-Europäischen Bemühungen zur Vereinheitlichung des Zugangs von offenen Verwaltungsdaten ist die Initiative Publicdata.eu. Diese Website versucht Verwaltungsdaten der Europäischen Staaten in einem One-Stop Shop zur Verfügung zu stellen. Das Portal wird von der Open Knowledge Foundation entwickelt und beherbergt momentan (Stand: Juli 2014) ca. 48000 Datensätze, davon ca. 2000 RDF Datensätze.

Problematisch bei diesem Portal sind, wie bereits erwähnt, die intransparente Semantik der angebotenen Datensätze. Vor allem durch die EU-weite Sammlung von Verwaltungsdaten erscheint es schwierig die Daten in geeigneter Form zu verlinken, da eine “gemeinsame” Bedeutung von den verwendeten Begriffen noch nicht Teil des Projektes ist. Hier könnten ODP oder standardisierte Vocabularies einen Beitrag leisten, um Daten anhand einer geeignete Semantik des Universe of Discourse zu klassifizieren und wenn möglich zu verlinken.

4. Zusammenfassung und Zukünftige Entwicklungen

Das Papier gibt einen Überblick über die wichtigsten Konzepte und Theorien hinter LOD. Die beschriebenen Konzepte dienen dazu, um den aktuellen Stand in der Wissenschaft zu dokumentieren und um neue Konzepte, die aktuell im wissenschaftlichen Diskurs stehen, zu beleuchten. Im Zuge dieses Papiers zeigt sich, dass im Bereich LOD viel Grundlagen- wie angewandte Forschung betrieben worden ist, jedoch noch viele Fragen offen sind. Ein Teil dieser Fragen wurde auf dem *Linked Open Data Workshop im Zuge des GI-Forums 2014* in Salzburg diskutiert, und ein großer Teil findet sich in dem Papier von Kuhn et al. [18].

Einer der zentralen Punkte bei LOD nimmt die Semantik ein - was quer über alle wissenschaftliche Literatur als Grundtenor erscheint. Um heterogene Daten zu verlinken sind (immer noch) große Aufwände zu betreiben, und es existiert keine standardisierte Vorgehensweise dafür. Zudem bietet die Wissenschaft noch unterschiedliche methodische Ansätze um die Semantik von Datensätzen in den “Griff” zu bekommen. Im Zuge dieses Papiers wurden Ontologien, Ontology Design Patterns und standardisierte Vocabularies angesprochen. Aus der Literatur heraus ist es schwierig eine Methode herauszufiltern die besser geeignet erscheint die offenen Fragen von LOD zu lösen - jede Methodik hat Vor- und Nachteile aufzuweisen. Generell kann die Referenzierung mittels Vocabularies als momentan sehr häufig anzutreffende Methode im Bereich LOD angesehen werden, was im “*Linked Open Data Workshop @ GI_Forum 2014*” zu Tage getreten ist. ODP stellen einen neuen Ansatzpunkt dar, die Welt mittels kleiner abgeschlossener wiederverwendbarer Ontologien zu modellieren, und in weiterer Folge diese Mini-Ontologien miteinander zu verbinden.

Da in einer LOD Cloud eine große Menge an Daten zu erwarten sind, muss man sich über eingesetzte Algorithmen und Methoden Gedanken machen, ob diese mit den großen Datenmengen - und deren Semantik - in entsprechender Form umgehen können. Zudem ist es notwendig die analytische Schärfe bei Abfragen und Analysen zu LOD kritisch zu betrachten, da

in LOD oftmals mehrere unterschiedliche Sichtweisen auftreten können - die sich zudem über die Zeitachse verändern können. Deshalb erscheint die explizite Auszeichnung der temporalen Gültigkeit eines Datensatzes von dringender Notwendigkeit, um die Aussagekraft von LOD-basierten Analysen zu erhöhen.

Da LOD eine Abkehr von der traditionellen layer-basierten Sichtweise auf die Welt darstellen können, könnten auch Geodateninfrastrukturen einem Wandel unterzogen sein. [4], [24], [13] und [10] beschreiben die potentielle Anwendung und “Verwertung” im Sinne von “neuen” Geodateninfrastrukturen. Diese “neuen” Geodateninfrastrukturen - SDI 3.0 [10] - ermöglichen es, dass alle Daten als LOD publiziert werden und somit für jedermann/frau jederzeit konsumierbar und in Applikationen verwertbar sind. Mit dieser Herangehensweise könnten eine Reihe von neuen Anwendergruppen für räumliche Daten erschlossen werden, die sonst keinen - oder nur schwierigen - Zugang zu diesen Daten hätten, womit die wirtschaftliche Bedeutung von räumlichen Daten in einer Informationsgesellschaft steigt.

Referenzen

- [1] Allemang, D., and Hendler, J. 2008. Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL. Morgan Kaufmann Publishers Inc., SanFrancisco, CA, USA.
- [2] Berners-Lee, T. 2009. Linked Data - Design Issues. Online:
<http://www.w3.org/DesignIssues/LinkedData.html>; (zuletzt besucht 22-07-2014).
- [3] Berners-Lee, T., Hendler, J., Lassila, O. 2001. The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In: Scientific American Magazine (May). Web: <http://www.scientificamerican.com/article.cfm?id=thesemantic-web> (zuletzt besucht am 29-07-2014)
- [4] Coetzee, S., Harvey, F., Iwaniak, A., und Cooper, A. 2013. Sharing and Coordinating SDIs in the age of Crowd Sourcing and Volunteered Geographic Information. Proceedings of the International Cartographic Congress.
- [5] Egenhofer, M., and Wilmsen, D. 2006. Changes in topological relations when splitting and merging regions. In: 12th International Symposium on Spatial Data Handling, pp. 339-352, Vienna, Austria, Springer-Verlag.
- [6] Gangemi, A. 2005. Ontology design patterns for semantic web content. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) The Semantic Web - ISWC 2005, 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings. Lecture Notes in Computer Science, vol. 3729, pp. 262-276. Springer.
- [7] Habernig, C., Harm, R., Höchtl, J., Jörg, W., Kaltenböck, M., Lutz, B., Mittlböck, M. und Pawel, S. 2013. Cooperation OGD Österreich - OGD Metadaten - 2.2. E-Government Bundesländer-Gemeinden. Web: https://www.ref.gv.at/uploads/media/OGD-Metadaten_2_2_2013_12_01.pdf (zuletzt besucht am 28-07-2014)

- [8] Halpin, H., Hayes, P. J., McCusker, J. P., McGuinness, D. L., and Thompson, H.S. 2010. When owl:sameas isn't the same: An analysis of identity in linked data. In: The Semantic Web-ISWC 2010, pp 305-320, Springer.
- [9] Hart, G. and Dolbear, C. 2013. Linked Data: A Geographic Perspective. Taylor and Francis.
- [10] Harvey, F., Jones, J., Scheider, S., Iwaniak, A., Kaczmarek, I., Lukowicz, J., und Strzelecki, M. 2014. Little Steps Towards Big Goals. Using Linked Data to Develop Next Generation Spatial Data Infrastructures (aka SDI 3.0). In Proceedings of the 17th AGILE Conference on Geographic Information Science, Castellón, Spain.
- [11] Heath, T. and Bizer, C. 2011. Linked data: Evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology, 1(1), 1-136.
- [12] Hu, Y., Janowicz, K., Carral, D., Scheider, S., Kuhn, W., Berg-Cross, G., Hitzler, P., Dean, M., and Kolas, D. 2013. A geo-ontology design pattern for semantic trajectories. In: *Spatial Information Theory*, pp. 438-456. Springer International Publishing.
- [13] Janowicz, K., Schade, S., Bröring, A., Kessler, C., Maué, P., und Stasch, C. 2010. Semantic Enablement for Spatial Data Infrastructures, *Transactions in GIS* 14(2): 111-129.
- [14] Janowicz, K. 2010. The role of space and time for knowledge organization on the semantic web. *Semantic Web*, 1(1):25-32.
- [15] Kaltenböck, M. 2014. Project Description of the Linked Open Data (LOD) PILOT Austria. Presentation at the PiLOD event at VU Amsterdam (Netherlands) on 29.01. 2014. Web: <http://de.slideshare.net/MartinKaltenboeck/linked-open-data-lod-pilot-austria> (zuletzt besucht am 28-7-2014).
- [16] Kauppinen, T. und Hyvonen, E. 2007. Modeling and Reasoning about Changes in Ontology Time Series, pp. 319-338. *Integrated Series in Information Systems*. Springer-Verlag, New York.
- [17] Keßler, C. und Janowicz, K. 2010. Linking Sensor Data - Why, to What, and How? In Kerry Taylor, Arun Ayyagari, David De Roure (Eds.): The 3rd International workshop on Semantic Sensor Networks 2010 (SSN10) in conjunction with the 9th International Semantic Web Conference (ISWC 2010), 7-11 November 2010, Shanghai, China
- [18] Kuhn, W., Kauppinen, T. and Janowicz, K. 2014 (in press). Linked Data -- A Paradigm Shift for Geographic Information Science. In proceedings of The Eighth International Conference on Geographic Information Science (GIScience2014), Vienna University of Technology, Austria, September 2014.
- [19] Limbu, M. 2012. Management of a Crisis (MOAC) Vocabulary Specification. Web: <http://observedchange.com/moac/ns/> (zuletzt besucht am: 29-07-2014)

- [20] Medak, D. 1999. Lifestyles—an algebraic approach to change in identity. In: Spatio-Temporal Database Management, pp. 19-39. Springer Berlin Heidelberg.
- [21] Open Knowledge Foundation. 2014. CKAN - The open source data portal software. Web: www.ckan.org (zuletzt besucht am 28-7-2014)
- [22] Ortmann, J., Limbu, M., Wang, D. and Kauppinen, T. 2011. Crowdsourcing Linked Open Data for Disaster Management. In proceedings of Terra Cognita, The 10th International Semantic Web Conference, October, 2011.
- [23] Perry, M., and Herring, J. 2011. OGC GeoSPARQL-A geographic query language for RDF data. OGC Implementation Standard.
- [24] Schade, S., und Smits, P. 2012. Why linked data should not lead to next generation SDI. IGARSS 2012: 2894-2897.
- [25] Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. 2012. Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent Systems*, 27(3), 16-24.
- [26] Sheridan, J., und Tennison, J. 2010. Linking UK Government Data. In: Proceedings of LDOW 2010. Web: http://ceur-ws.org/Vol-628/l dow2010_paper14.pdf (zuletzt besucht am 28-07-2014).