

Auf dem Weg zu Leipzig Data

Das Leipzig Data Projekt

Prof. Dr. Hans-Gert Gräbe
Leipzig Data Team

Universität Leipzig
<http://leipzig-data.de>

Version vom 14. Februar 2015

Die Herausforderung

- Web 2.0 als tief greifender technologischer Wandel digitaler Repräsentationsformen, der seit etwa 2005 deutlich Fahrt aufgenommen hat und die Kommunikationsstrukturen der Gesellschaft grundlegend umkrepeln wird.
- Basistechnologie eines neuen Technologie-Zyklus, dessen Durchdringungsphase etwa 2008 begonnen hat und als typische pervasive Technologie kaum einen gesellschaftlichen Bereich auslassen wird.
- Dieser Herausforderung muss sich Regionalentwicklung stellen.

Das Internet

Auf dem Weg
zu Leipzig
Data

Leipzig Data
Team

Der Weg zu
Leipzig Data

Die Vision

Wo stehen
wir?

Längerfristige
Teilprojekte

Weitere
studentische
Projekte

Leipzig Data
und RDF

Verfügbare
Infrastruktur

- Web der Daten? Daten werden von Akteuren zu von ihnen verfolgten Zielen zusammengetragen.
99 % der Webseiten im Internet enthalten Texte bzw. textuelle Repräsentationen.
- Web der Worte! Web der Sätze! Web der Stories!
Mit den neuen Technologien rückt dieser Aspekt gegenüber einem engen Begriff Daten in den Vordergrund.
- Aber (Faust): ... kann ich das *Wort* so hoch unmöglich schätzen ...
- Zusammen mit den Worten müssen die Akteure und ihre kooperativen unternehmerischen Taten im Blick bleiben.

Leipzig Data – Die Vision

- Mit dem Auf- und Ausbau breitbandiger leistungsfähiger digitaler Kommunikationsnetze sind die Grundlagen gegeben, unser gesamtes soziales Leben auf dieser erweiterten Basis neu zu strukturieren.
- Leipzig Data ist mehr als Leipzig Open Data, letzteres aber dessen Kern.
- Leipzig Open Data ist viel mehr als Offene Daten der Stadt, wir müssen (inhaltlich) nicht auf die Stadtverwaltung warten, um loszulegen.
- Im Mittelpunkt von Leipzig Data stehen nicht die Daten (bzw. Worte) selbst, sondern die Menschen und deren Motive, die auf dieser Grundlage planen und handeln.

Wo stehen wir?

- Vorarbeiten von API Leipzig (Orte und Adressen in Leipzig, Übersicht über Medienunternehmen, Event-Daten) sowie aus dem MINT-Netzwerk (Angebote, MINT-Orte, Träger, einige Bereiche der Jugendhilfe) sind weitgehend aufgearbeitet, in die Strukturen von Leipzig Data integriert, weiter angereichert und fortgeschrieben.
- Kurzzeitprojekt „Leipzig Open Data Initiative“ im Rahmen der städtischen Ausschreibung Open Innovation Nov. 2012 bis April 2013.
- Einbindung in die studentische Ausbildung durch Seminararbeiten und Projektpraktika. Damit Erschließung weiterer Datenbestände sowie Konsolidierung von Technologien.
- Zusammenarbeit mit dem OK Lab Leipzig – <http://codefor.de/leipzig/>.

Projekt „Gelbe Seiten“

- Aus APILeipzig wurden über 65 000 Leipziger Adressen übernommen, nach einem Standard in URIs verwandelt und mit Geodaten angereichert. Damit lassen sich Orte in Leipzig referenzieren.
- Dazu wurden (aktuell 428) Orte identifiziert, mit URI versehen und einer Adresse zugeordnet, zu denen in Leipzig Data „Geschichten“ erzählt wurden.
- Siehe <http://www.leipzig-data.de/gelbe-seiten/>.
- Zu diesen Orten wurden unsystematisch weitere standardisierte Informationen zusammengetragen.

In verschiedenen zeitlich beschränkten Projekten werden auf der Basis *Ortsinformationen* zusammengetragen, die Informationen zum jeweiligen Ort unter einem bestimmten Gesichtspunkt zu einer bestimmten Zeit darstellen, etwa das Projekt „Jugendstadtplan 2013“ im Vorfeld der Worldskills.

Projekt „Event Framework“

Auf dem Weg
zu Leipzig
Data

Leipzig Data
Team

Der Weg zu
Leipzig Data

Die Vision

Wo stehen
wir?

Längerfristige
Teilprojekte

Weitere
studentische
Projekte

Leipzig Data
und RDF

Verfügbare
Infrastruktur

- **Ziel** dieses Teilprojekts ist es, eine Infrastruktur aufzubauen, in die Event-Daten in einheitlichem RDF-Format aus verschiedenen Quellen und von verschiedenen Akteuren eingespeist werden und der Allgemeinheit zum Gebrauch zur Verfügung stehen.
- Nach diesem Schema werden Event-Informationen aus verschiedenen Quellen (APILeipzig, Netzwerk Energie & Umwelt, MINT-Netzwerk) zusammengetragen und wöchentlich aktualisiert. Events werden 90 Tage nach Ablauf aus der Infrastruktur gelöscht.
- Mit diesen Event-Informationen werden unsere Daten über Orte und Akteure in Leipzig weiter konsolidiert.
- Die prinzipiellen Möglichkeiten dieses Event Frameworks demonstriert unser *Leipzig Data Event Widget*.
- Siehe <http://www.leipzig-data.de/events/>.

Projekt „Geolokale Informationen“

- Eine der großen Killerapplikationen Offener Daten ist die Zuordnung von Geokordinaten zu konkreten Aktionen und Beschreibungen und die kartenmäßige Darstellung dieser geolokalen Informationen.
- Mit dem Teilprojekt *Gelbe Seiten* sind hierfür die erforderlichen Informationen für Anwendungsprojekte verfügbar.
- Es wurden verschiedene kürzere Projekte (Jugendstadtplan, Karte der Standorte von EEG-Anlagen, Gentrifizierungsprojekt) umgesetzt, um die technische Basis für derartige Kartendarstellungen zu konsolidieren – siehe <https://github.com/LeipzigData/Karten>.

Weitere studentische Projekte

Auf dem Weg
zu Leipzig
Data

Leipzig Data
Team

Der Weg zu
Leipzig Data

Die Vision

Wo stehen
wir?

Längerfristige
Teilprojekte

Weitere
studentische
Projekte

Leipzig Data
und RDF

Verfügbare
Infrastruktur

- Gentri-14 – Informationen zu Gentrifizierungsprozessen in Leipzig. Siehe <http://www.leipzig-data.de/Gentri-14/>
- Energie-13 – Dezentralen Energieerzeugungsanlagen in Leipzig. Siehe <http://www.leipzig-data.de/Energie-13/>
- JSP-13 – Jugendstadtplan von Leipzig. Siehe <http://www.leipzig-data.de/Jugendstadtplan/>

RDF – Grundlegende Konzepte

Auf dem Weg
zu Leipzig
Data

Leipzig Data
Team

Der Weg zu
Leipzig Data

Die Vision

Wo stehen
wir?

Längerfristige
Teilprojekte

Weitere
studentische
Projekte

Leipzig Data
und RDF

Verfügbare
Infrastruktur

- RDF = Resource Description Framework
 - Zentrale Idee: Portioniere Informationen in elementaren Drei-Wort-Sätzen, speichere diese als Tripel und verwende standardisierte Werkzeuge zu deren Management.
- *Resources*: URI (Unique Resource Identifier), HTTP-Zugriff
 - Zugriff auf weltweit verteilte Daten nach einheitlichem Prinzip.
- *Resource Descriptions*: Liefere auf eine HTTP-Anfrage eine angemessene Portion Information in strukturiertem RDF-Format aus, die mit Informationen aus anderen Quellen zu neuen RDF-Sätzen kombiniert werden kann (Linked Data Prinzip).
- (Federated) Query Language SPARQL als RDF-Anfragesprache.
- *RDF Triple Stores* und *SPARQL Endpunkte* als Teil einer weltweiten verteilten Infrastruktur von Datenspeichern.

RDF und die Leipzig Data Infrastruktur

- Tools und Daten sind in mehreren Repos über den github-Account <http://github.com/LeipzigData> öffentlich zugänglich.
- Grundlegende Konzepte und Ansätze werden im Wordpress-basierten Portal <http://leipzig-data.de> genauer dargestellt.
- Ein zentraler Virtuoso basierter RDF Triple Store <http://leipzig-data.de/Data> mit eigenem SPARQL Endpunkt <http://leipzig-data.de:8890/sparql> arbeitet nach Linked Data Prinzipien.
- Regelmäßige Dumps der RDF-Daten im Turtle-Format
- Anleitung zur Installation der Tools und Daten in einem Virtuoso RDF Store innerhalb eines lokalen Apache Webserver.

Leipzig Data Datentruckturen

- AdministrativeGliederung.ttl – Ortsteile und Stadtbezirke in Leipzig
- Adressen.ttl – Leipziger Adress-Daten
- API-Referenzen.ttl – Zuordnung von URIs zwischen Leipzig Data und APILeipzig
- GeoDaten.ttl – GeoDaten zu den Adress-Daten
- Events.ttl – Dump der aktuell verfügbaren Event-Informationen
- Orte.ttl – Orte in der Stadt Leipzig, die in Leipzig Data Geschichten schon einmal eine Rolle gespielt haben
- Personen.ttl – Personen, die in Leipzig Data Geschichten schon einmal eine Rolle gespielt haben
- Strassenverzeichnis.ttl – Leipziger Straßenverzeichnis einschließlich Straßennummern der Stadtverwaltung.

Verfügbare Infrastruktur (Zusammenfassung)

- <http://www.leipzig-data.de> – Das Leipzig Data Portal.
- <http://www.leipzig-data.de/Data> – RDF Data Store
- <http://leipzig-data.de:8890/sparql> – SPARQL Endpunkt für Anfragen auf den Daten.
- <https://github.com/LeipzigData> – github Organisationsaccount von Leipzig Data mit den Projekten
 - RDFData – RDF-Datenbasis
 - Tools – eine Reihe von RDF-Werkzeugen (für Entwickler)
 - Karten – Beispiele für geolokale Darstellungen von Daten
 - Publications – Veröffentlichungen und Präsentationen zum Projekt
- <http://www.leipzig-data.de/Backups> – wöchentliche Dumps der Daten.