

Dr. Tobias Thelen
Universität Osnabrück
Zentrum für Digitale Lehre, Campus-Management und Hochschuldidaktik
Institut für Informatik
Institut für Kognitionswissenschaft

Die Informationen sind frei, wer kann sie verwerten?

Plädoyer für eine Open-Data-Initiative in Stud.IP

Meine Forderung

*Die in Stud.IP eingeschlossenen Informationen
sollen für jedermann frei zugänglich sein,
und zwar in offenen, maschinell verarbeitbaren
Datenformaten (z.B. für Big Data).*



Fotograf: Armin Kübelbeck,
[CC-BY-SA, Wikimedia Commons](#)

Informationsfreiheitsgesetz

§ 1 IFG

Jeder hat nach Maßgabe dieses Gesetzes gegenüber den Behörden des Bundes einen Anspruch auf Zugang zu amtlichen Informationen.

§ 5 IFG

(1) Zugang zu personenbezogenen Daten darf nur gewährt werden, soweit das Informationsinteresse des Antragstellers das schutzwürdige Interesse des Dritten am Ausschluss des Informationszugangs überwiegt oder der Dritte eingewilligt hat.

Art. 39 Bayerisches Datenschutzgesetz

(1) Jeder hat das Recht auf Auskunft über den Inhalt von Dateien und Akten öffentlicher Stellen, soweit ein berechtigtes, nicht auf eine entgeltliche Weiterverwendung gerichtetes Interesse glaubhaft dargelegt wird und

- 1. bei personenbezogenen Daten eine Übermittlung an nicht öffentliche Stellen zulässig ist und*
- 2. Belange der öffentlichen Sicherheit und Ordnung nicht beeinträchtigt werden.*

§3 Informationsweiterverwendungsgesetz (IWG)

(2) Informationen sind in allen angefragten Formaten und Sprachen, in denen sie bei der öffentlichen Stelle vorliegen, zur Weiterverwendung zur Verfügung zu stellen; soweit möglich und wenn damit für die öffentliche Stelle kein unverhältnismäßiger Aufwand verbunden ist, sind sie vollständig oder in Auszügen elektronisch sowie in einem offenen und maschinenlesbaren Format zusammen mit den zugehörigen Metadaten zu übermitteln. Sowohl die Formate als auch die Metadaten sollten so weit wie möglich anerkannten, offenen Standards entsprechen.

Definition Open Data

Bereitstellung von (öffentlichen) Daten

- für jedermann
- für beliebige (legale) Zwecke
- in offenen Formaten
- mit offenen Lizenzen (z.B. CC-BY oder CC-BY-SA)

Ziele der Open-Data-Bewegung

- **Open Government:** Wenn Daten frei verfügbar sind, erhöhen sich die Transparenz und damit die Voraussetzungen für demokratische Teilhabe
- **Beteiligung:** Wenn die veröffentlichten Daten kritisch genutzt werden, können Rückfragen und Feedback zur Verbesserung der Datenqualität oder der zugrundeliegenden Sachverhalte führen
- **Informationsbereitstellung:** Die veröffentlichten Daten können als authentische Daten aus erster Hand genutzt werden. Die ermöglicht kritische Überprüfungen von Behauptungen.
- **Innovation:** Die freie Bereitstellung ermöglicht die Verwirklichung innovativer Ideen.

Das Datenportal für Deutschland

Open Government: Verwaltungsdaten transparent, offen und frei nutzbar

Nach Datensätzen suchen

Suchen

Erweiterte Suche

Kartensuche

Was hat das mit
Stud.IP zu tun?

orum
in
eth

tz
Stud.IP

hline: 1 mehr ...



Brainstorming

Welche Daten aus Stud.IP
werden aus Ihrer Installation
bereits öffentlich bereitgestellt?

Open Data ≠ Webseite

Viele Daten werden digital,
aber nur für Menschen lesbar bereitgestellt.

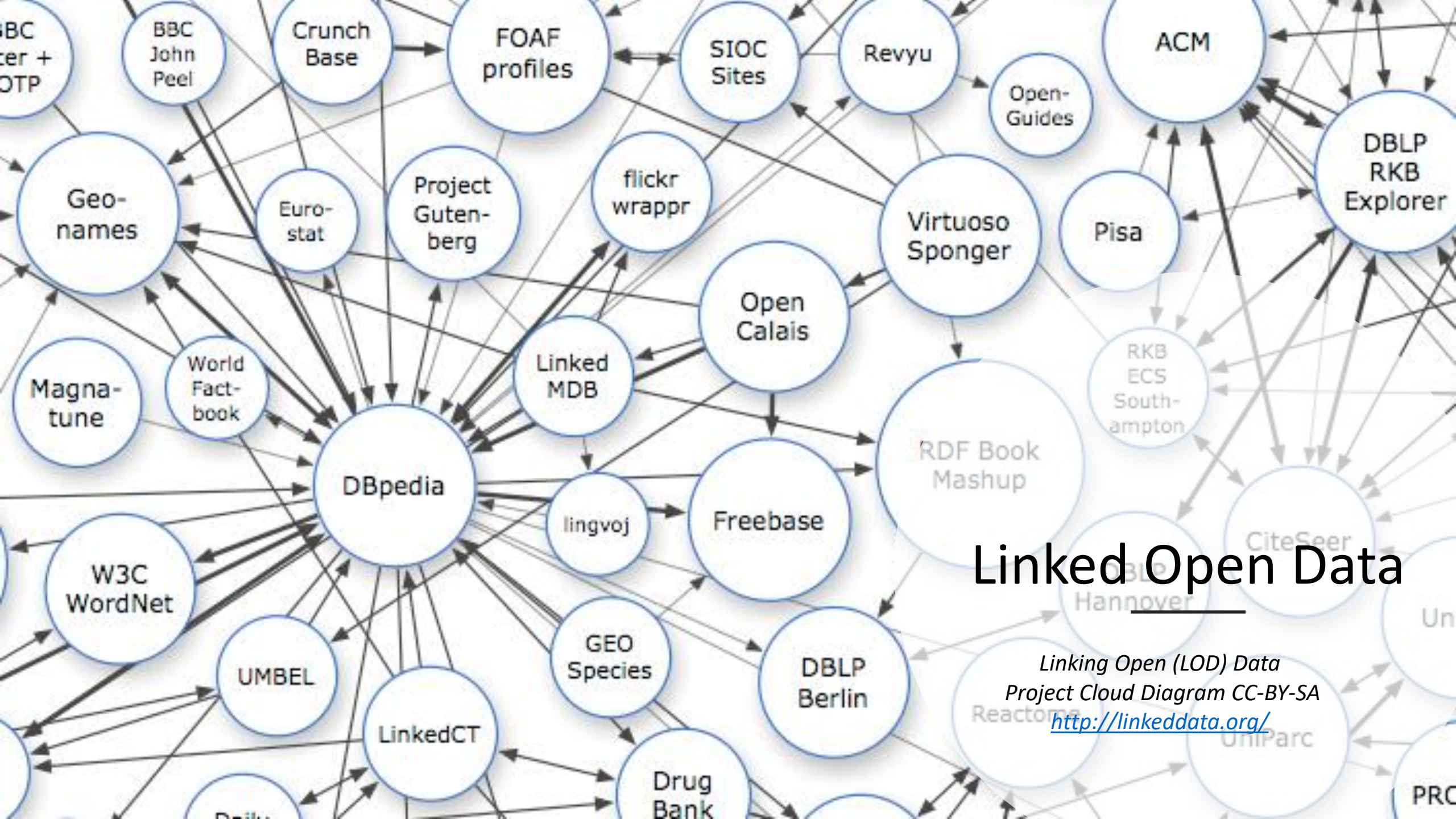
Forderung:
Maschinenlesbare und
maschinenverarbeitbare Bereitstellung.

Grade von Verarbeitbarkeit

- Daten in Text (PDF oder Webseite)
- Daten in proprietärem Format
 - jeweils aktuell abrufbare CSV-Tabellen
 - Webservice (z.B. ReST) mit JSON oder XML
 - RDF
 - GraphQL
 - *Semantik fehlt!*
- Daten in bekannten Formaten
 - ... gibt es i.d.R. nicht ...
- Daten mit semantischer Verlinkung
 - Linked Open Data
 - verwendet bekannte allgemeine Ontologien (DBPedia, Wikidata, YAGO, GND, ...)

Triviales (?) Beispiel: Semesterzeiten

- Annahme: Die Semester- und Ferienzeiten sollen aus Stud.IP als Open Data bereitgestellt werden.
- Varianten:
 - Externe Seite oder andere dynamische Übernahme auf Uni-Webseiten
 - Abruf von proprietärer Repräsentation über spezifische URL:
<https://studip.uni-osnabrueck.de/./rest/site/semester/2019/winter/...>
 - Abruf von Linked Open Data (RDF, abstrahiert):
<http://uni-osnabrueck.de
[http://dbpedia.org/Academic term](http://dbpedia.org/Academic_term)
(25.10.2019-04.02.2020) >



Linked Open Data

*Linking Open (LOD) Data
Project Cloud Diagram CC-BY-SA*

<http://linkeddata.org/>

Braucht man nicht!

Google ferien in niedersachsen

Alle Bilder Maps News Shopping Mehr Einstellungen Tools

Ungefähr 62.200.000 Ergebnisse (0,62 Sekunden)

Schulferien Niedersachsen 2019/2020 aktuell

Herbstferien 2019	Fr 04.10. - Fr 18.10.
Winterferien 2020	Mo 03.02. - Di 04.02.
Osterferien 2020	Mo 30.03. - Di 14.04.
Pfingstferien 2020	Fr 22.05./Di 02.06.
Sommerferien 2020	Do 16.07. - Mi 26.08.

[2 weitere Zeilen](#)

Schulferien Niedersachsen 2019, 2020
<https://www.schulferien.org> › Niedersachsen › niedersachsen

Informationen zu diesem Ergebnis Feedback geben

Suchmaschinen, v.a. Google weiten semantische Suchfunktionen aus, d.h. Informationen werden nicht einfach nur aufgrund von Textvorkommen verlinkt, sondern Anfragen "verstanden" und passende Informationen direkt aufbereitet.

Nachteile gegenüber Open Data:

- Nur Google selbst kann die Erkenntnisse nutzen
- Keine Kontrolle über die Extraktion und Verarbeitung

Anwendungsfälle

Kurzes Brainstorming:

Wozu könnte die Open-Data-Veröffentlichung von Semester- und Feriendaten gut sein?

Datenqualität

Häufig das Hauptproblem bei der Veröffentlichung von Daten bei:

- Vollständigkeit der Daten
- Korrektheit der Daten
- Schlechte Dokumentation der Daten (Zweckentfremdung von Feldern, schlecht eingehaltene Konventionen)



Datenqualität im Check

Wie gut (vollständig, korrekt, konsistent) wären bei Ihnen folgende Daten aus Stud.IP?

- Semester- und Ferienzeiten
- Aufbauorganisation (Einrichtungen, Fakultäten, Institute, ...)
- Lehrangebot
- Lehrveranstaltungszeiten und -orte

Einstellungen

Themen öffentlich einsehbar

Weiteres triviales (?) Beispiel

Wer hat diesen Schalter gesehen?

Veröffentlichung von Veranstaltungs-Einzelterminen

Diskussion:

Was spricht für oder gegen die Veröffentlichung von
Veranstaltungs-Einzelterminen (und unter welchen Bedingungen
und Voraussetzungen)?

Diskussion: Welche Daten
haben Hochschulen in
Stud.IP, die als Open Data
freigegeben werden
könnten?

Nächste Schritte

- Aktuell:
 - Explorative Bachelorarbeit (Cognitive Science / Artificial Intelligence)
- Planung Uni Osnabrück / SIDDATA-Projekt:
 - Veröffentlichung einiger interessanter Daten
 - Wissenschaftliche Evaluation und Publikation der Ergebnisse
- Gesucht:
 - Weitere Betreiber mit Interesse
- Perspektivisch:
 - Gemeinsame Entwicklung/Modifikation eines Plugins/Kernfunktionalität
 - Gemeinsame Policies für Datenqualität
 - Bereitstellung über govdata.de
 - <http://data.studip.de> ?

Vielen Dank!



... Diskussion und Fragen