

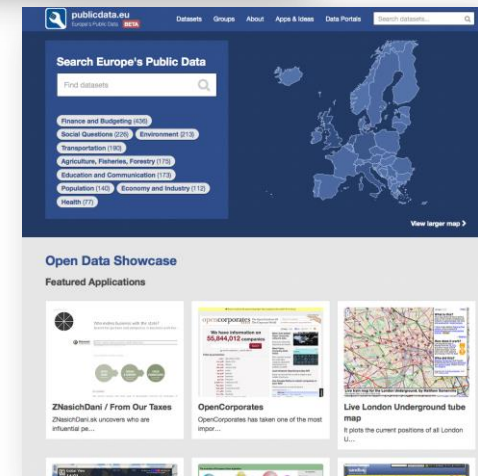
Auf dem Weg zum pan-europäischen Open Data-Portal

Fabian Kirstein M.Sc. | Fraunhofer FOKUS | 4. OGD D-A-CH-LI - Konferenz - Open X



Open Data in der Europäischen Union

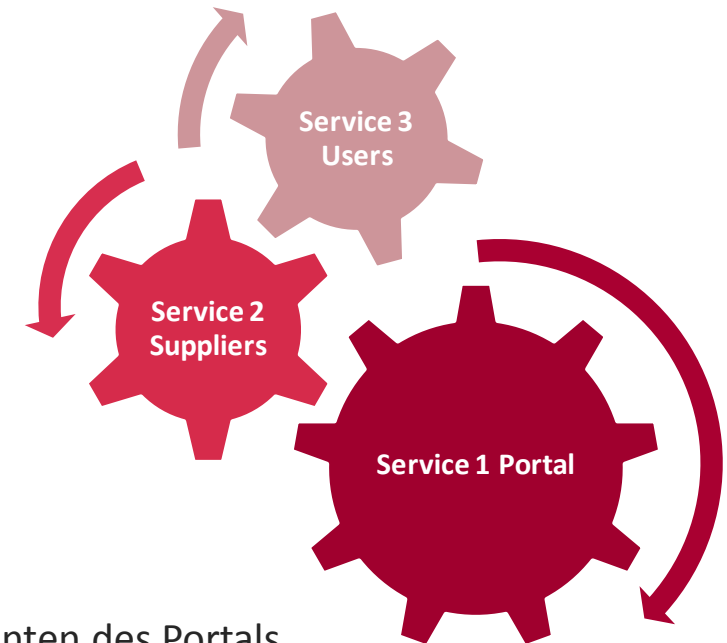
- Daten als Top-Priorität der EU:
“Enhance data driven innovation and re-use of PSI.”
(Europäischer Rat 2013)
- Open Data als Teil der Digitalen Agenda der EU:
“Action 3: Open up public data resources for re-use“
- 40 Milliarden EUR jährliches wirtschaftliches Potenzial
in der EU prognostiziert [1]
- Mehr als 200 Open-Data-Kataloge EU-weit (Open Data Monitor)
- „Vorläufer“ eines EU-weiten Datenportals
 - publicdata.eu (EU-Forschungsprojekt „Linked Open Data 2“)
 - data.opendatasupport.eu (EU-Projekt „Open Data Support“)



[1] http://europa.eu/rapid/press-release_IP-11-1524_en.htm

Ziele des Projektes „pan-europäisches Open Data-Portal“

- Übergreifendes Ziel des dreijährigen Projektes:
 - Daten, die von öffentlichen Stellen in Europa veröffentlicht werden, zentral zugänglich machen (One-Stop Shop)
- Drei Aktivitätsbereiche (Services)
 - Service 1: Konzeption, Implementation & Betrieb der pan-europäischen Datenportalinfrastruktur
 - Service 2: Unterstützung der Datenbereitstellung
 - Service 3: Unterstützung der Weiterverwendung öffentlicher Daten
- Auftraggeber: Data Value Chain Unit / DG Connect
- Acht Partner, Fraunhofer FOKUS verantwortlich für Kernkomponenten des Portals

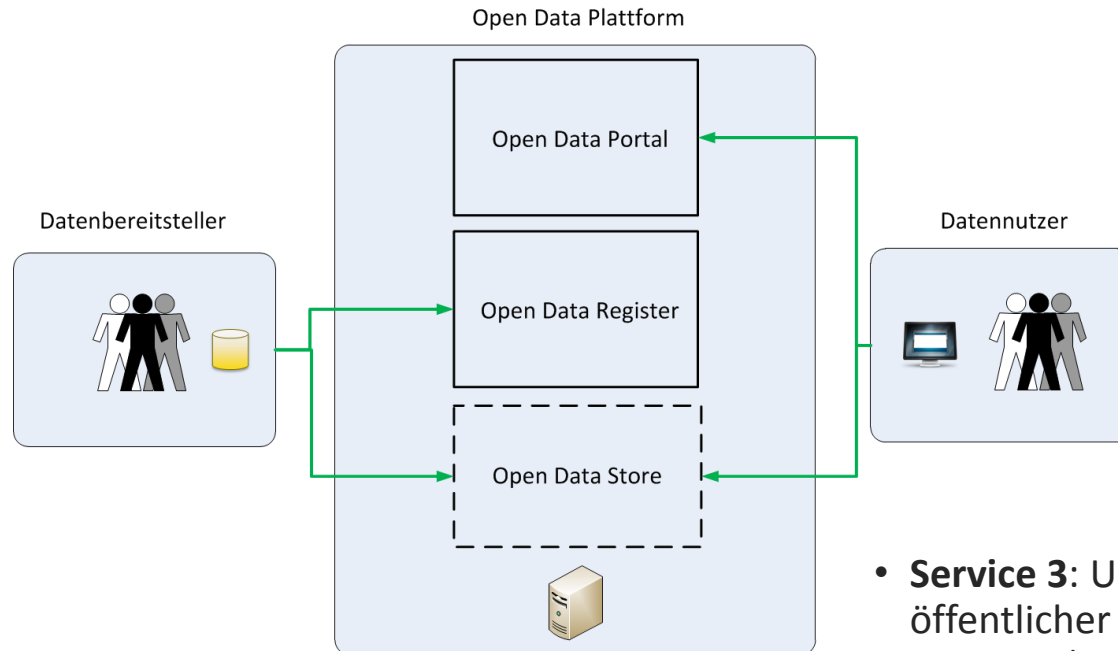


Alle drei Aktivitätsbereiche hängen voneinander ab, um den Wert von Open Data nutzbar zu machen.

Größerer Kontext des Portals

- **Service 2:** Unterstützung der Datenbereitstellung durch nationale OD Status Analysen

- OD Leaders Network
- Trainingsunterlagen
- Vor-Ort-Workshops



- **Service 3:** Unterstützung der Weiterverwendung öffentlicher Daten u.a. durch

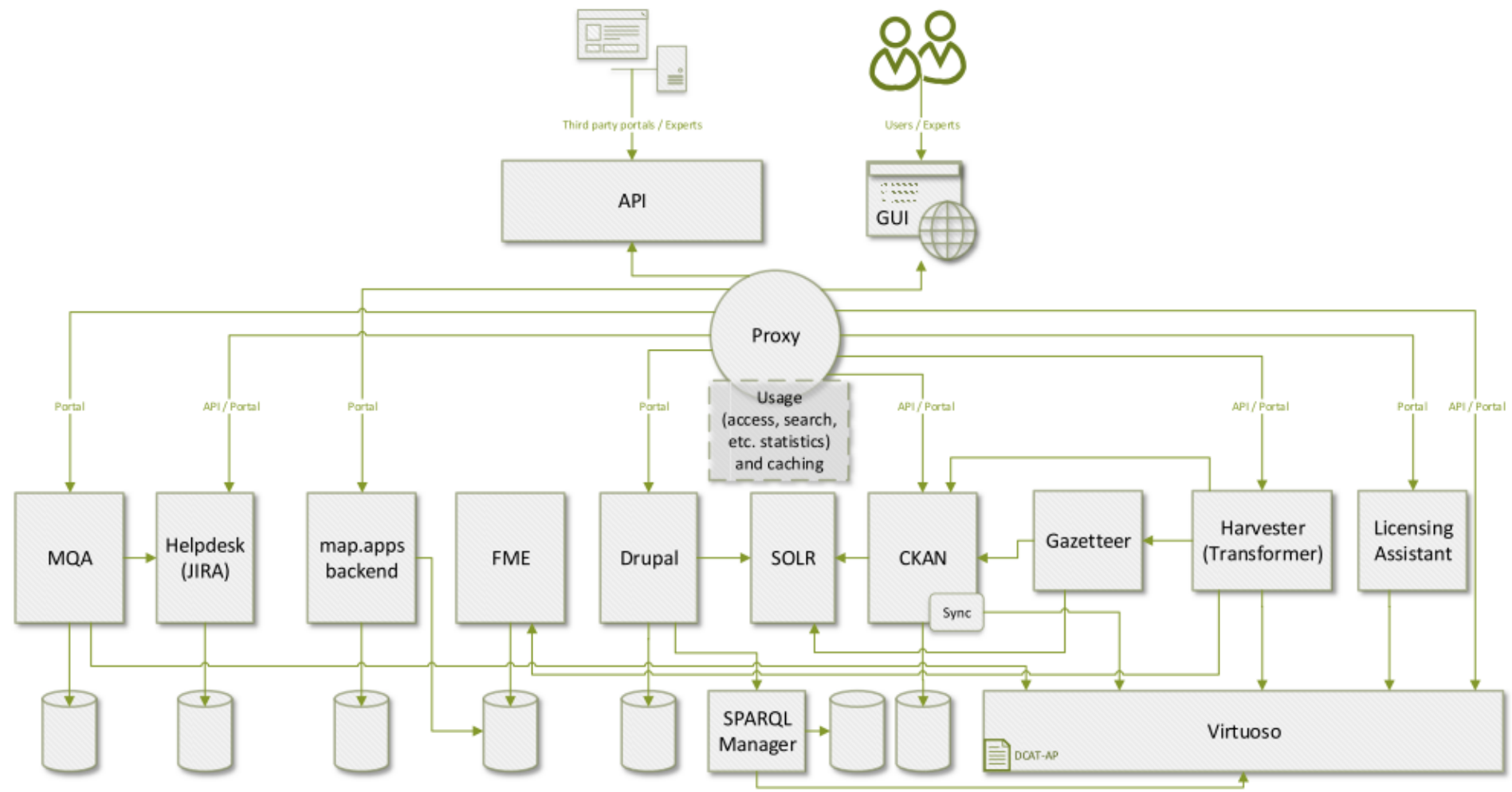
- Marketing der Portalinfrastruktur und ihrer Inhalte
- Untersuchung ökonomischer Effekte der Weiterverwendung
- Förderung von Fachcommunities

Herausforderung Heterogenität

- Technische Basis der Datenquellsysteme
 - Einbeziehung von OD-Portalen und Fachportalen (z.B. Geodateninfrastrukturen)
 - Verschiedene technologische Lösungen & variierende Performanz der Quellsysteme
- Format / Qualität der Metadaten
 - Unterschiedliche nationale Metadatenspezifikationen
 - Keine einheitlichen Vorgaben zur Datenqualität
- Multilingualität
 - Mehrsprachigkeit betrifft die Benutzeroberfläche, redaktionelle Inhalte, Datensatzbeschreibungen, Suchergebnisse sowie Schulungsunterlagen
 - Übersetzung in 24 offizielle EU-Sprachen
- Lizenzierung und Weiterverwendung der Datensätze
 - Viele verschiedene offene Lizenzen, wie z.B. nationale Open-Data-Lizenzen



Architektur



Ausgewählte Module des Portals I

■ Datenkatalog (CKAN)

- Speicherung von Metadaten, keine Speicherung von Ressourcen (Inhaltsdaten)
- Je Metadatensatz Angabe mehrerer Ressourcen in unterschiedlichen Formaten
- Open Source Lösung CKAN (Open Knowledge Foundation) als Basis für die Katalogsoftware
- Verwendung eines einheitlichen Metadatenschemas innerhalb des Portals
- Konform mit DCAT-AP Standard

■ Linked Data Datenbank (Virtuoso)

- Neben der Speicherung in CKAN, werden die Metadaten im RDF-Format (DCAT-AP) abgespeichert
- Nutzung des Datenbanksystems Virtuoso (u.a. Triplestore)
- Direkter Zugriff über eine SPARQL-Schnittstelle

■ Open Data Lizenzassistent (Licensing Assistant)

- Open-Data-Lizenzen Inhaltsservice
- Open-Data-Lizenzen Empfehlungsservice
- Open-Data-Lizenzen Kombinationsservice



Datenlizenz Deutschland – Zero – Version 2.0



Open Government Licence
for public sector information

Open Data Commons Open
Database License (ODbL)

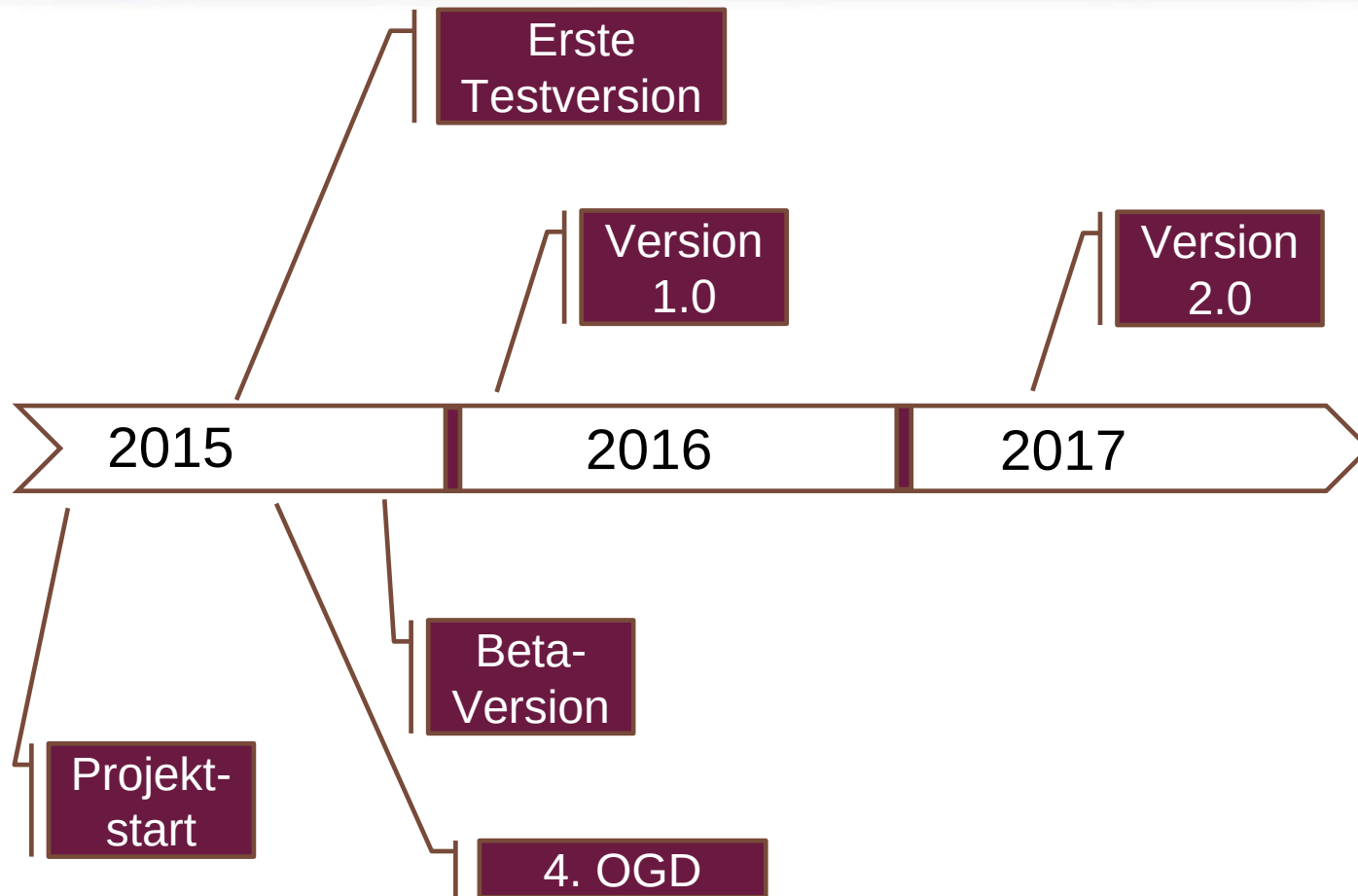
Ausgewählte Module des Portals II

- Content Management (Drupal)
 - Pflege und Bereitstellung von redaktionellen Inhalten
 - Zentraler Einstiegspunkt des Portals
- Metadatenqualität (MQA)
 - Automatisierte Checks zur Qualität der Metadaten
 - Bspw. Übereinstimmung mit vorgegebenem Metadatenschema oder Aktualität der Ressourcenverweise
 - Generierung von Qualitätsberichten für Datenbereitsteller
- Harvester / Transformer
 - Zentrale Komponente zur automatisierten „Befüllung“ des Datenkataloges
 - Abrufen der Metadaten von den Datenbereitstellern (Nationale Open Data Portale, Fachsysteme, regionale OD Portale)
 - Transformation in das Metadaten-Schema (CKAN / DCAT-AP) des Portals
 - Erstellung der Datensätze

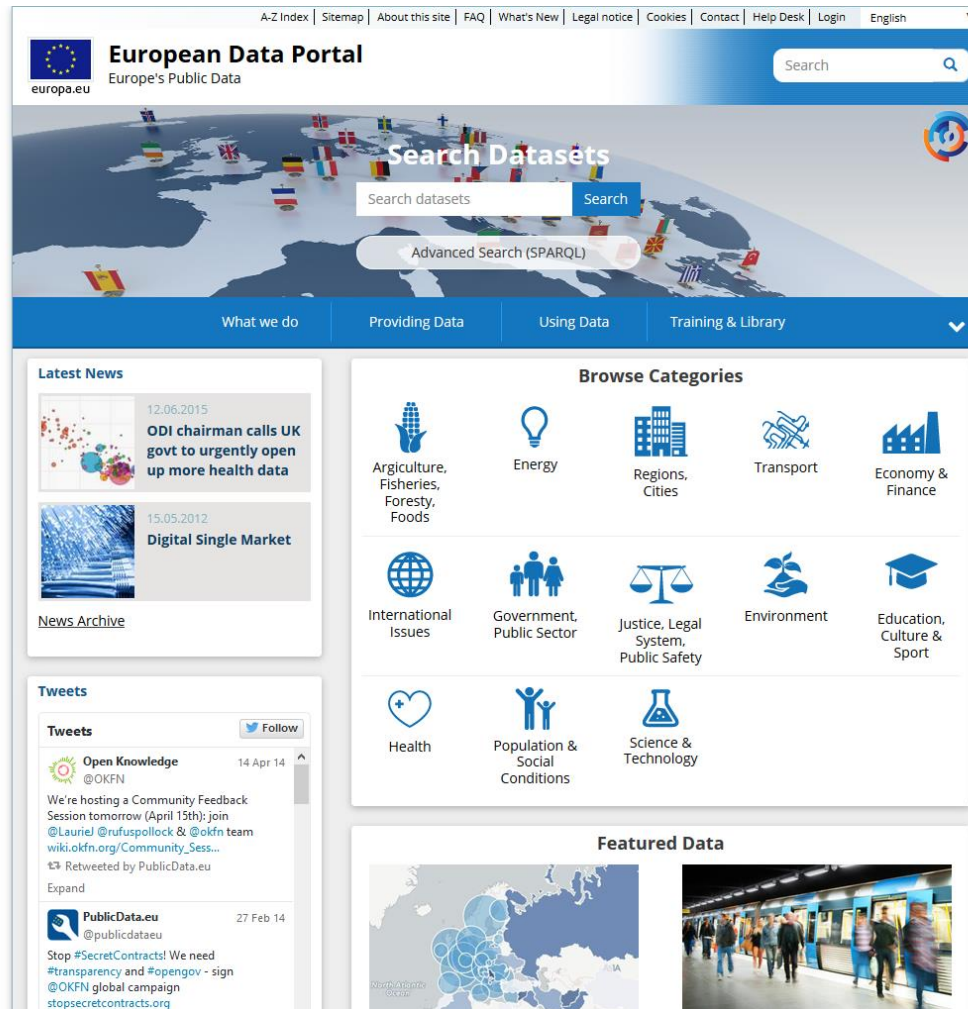
Umsetzung der Mehrsprachigkeit

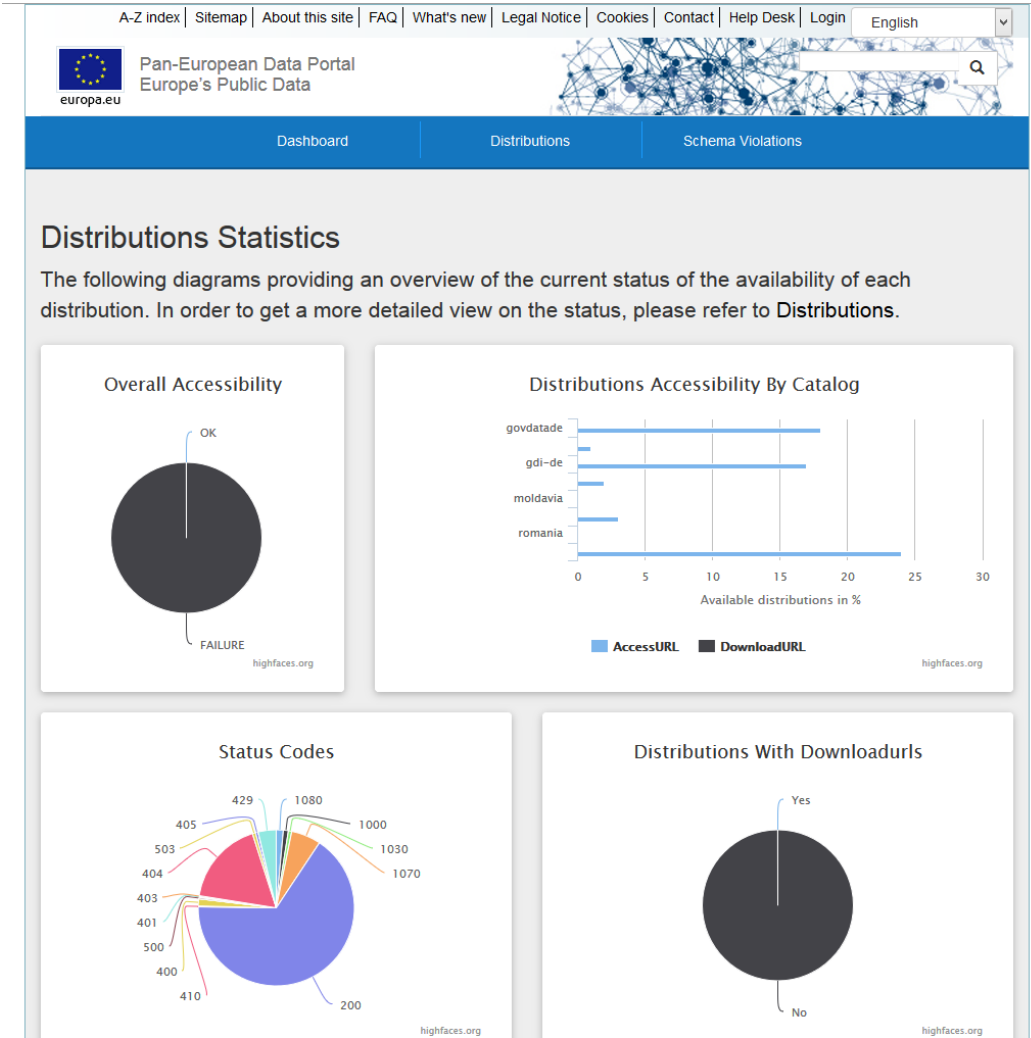
- Übersetzung in 24 offizielle EU-Sprachen
- Suche in allen Sprachen
- Größte Herausforderungen: Übersetzung der Metadaten und Integration in CKAN
- Basissprache: Englisch
- Einsatz von MT@EC
 - Service der EU für maschinelle Übersetzung
 - SOAP-Schnittstelle
 - http://ec.europa.eu/dgs/translation/translationresources/machine_translation/index_de.htm
- Erstellung einer CKAN-Erweiterung für Mehrsprachigkeit
 - Anpassung des Datenschemas (Ergänzung um ein Übersetzungsfeld „translation“)
 - Erweiterung des SOLR-Suchservers
 - Anpassung von CKAN-Kernfunktionalitäten, um Anpassungen entsprechend zu verarbeiten
- Implementierung einer Middleware zwischen CKAN und MT@EC
 - Nicht-blockierende automatische Übersetzung
 - Neue und aktualisierte Datensätze werden der Middleware übermittelt
 - Middleware fordert nicht vorhandene Sprachen von MT@EC an
 - Datensätze werden in CKAN zurückgeschrieben

Zeitplan



Erste Testversion







Vielen Dank!

Fabian Kirstein M.Sc.
Kaiserin-Augusta-Allee 31
10589 Berlin
Tel.: 030 3463 7718
Email: fabian.kirstein@fokus.fraunhofer.de

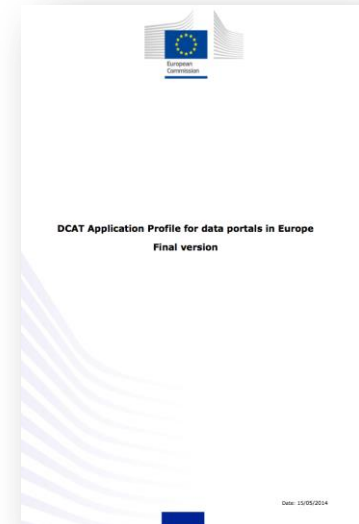


Backup



Data Catalogue Vocabulary – Application Profile (DCAT-AP)

- Spezifizierung auf Basis von DCAT zur Beschreibung von Datensätzen des öffentlichen Sektors in Europa
- Hauptanwendung von DCAT-AP ist die portalübergreifende und grenzübergreifende Suche nach Datensätzen
- Entwickelt von Arbeitsgruppe mit Mitgliedern aus 16 EU-Mitgliedsländern im Frühjahr 2013
- DCAT-AP wird im pan-Europäischen Datenportal zum Einsatz kommen
- Aktuell wird eine öffentliche Revision der Spezifikation durchgeführt: <https://joinup.ec.europa.eu/node/137964/>
- Bislang wurden 67 Änderungen vorgeschlagen, darunter z.B. Mechanismen zur Vermeidung von Duplikaten oder zur Identifizierung von maschinell übersetzten Inhalten



- „Open Data as part of the EU Digital Agenda: Pillar 1 Digital Single Market: Action 3: Open up public data resources for re-use“ -> „Public authorities produce large amounts of data that could become the raw material for new, innovative cross-border applications and services.“
- Leistungsangebot
 - Konzeption, Implementation & Betrieb von Open-Data-Plattformen sowie Strategieentwicklung & IT-Beratung zur Bereitstellung und Weiterverwendung von Open Data
 - Konzeption und Implementation von Werkzeugen zur Aufbereitung und Weiterverwendung von Open Data
- Beispielprojekte
 - EU Projekte: Pan-europäisches Open Data Portal, Open Cities, ENGAGE, Policy Compass, Share-PSI 2.0
 - Nationale & lokale Projekte in Deutschland: GovData – Das nationale Datenportal für Deutschland, Open Data Berlin, Transparenzportal Hamburg, Roadmaps, Strategieentwicklung & IT-Beratung für Open Data für die Städte Frankfurt und Köln sowie die Länder Niedersachsen & NRW

