



Fraunhofer FOKUS

Fraunhofer Institute for Open
Communication Systems

Kaiserin-Augusta-Allee 31
10589 Berlin, Germany

www.fokus.fraunhofer.de



Identification and Utilization of Components for a linked Open Data Platform

Evanela Lapi, Nikolay Tcholtchev, Louay Bassbouss, Florian Marienfeld,
and Ina Schieferdecker

Fraunhofer Institute FOKUS, Berlin, Germany
{firstname.lastname@fokus.fraunhofer.de}

IEEE METHOD Workshop, Izmir, 16 July 2012



What is Open Data ?

“ A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and share-alike. ”

<http://opendefinition.org>

Paradigm changes for Open Data

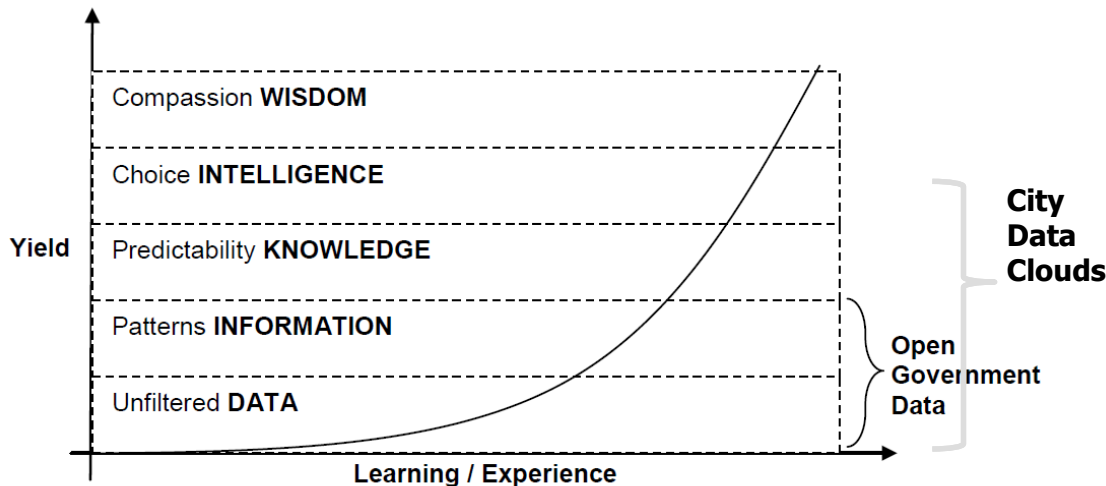
- **Public vs. confidential**
From: everything is confidential if not explicitly marked as public
To: everything is public if not explicitly marked as confidential
- **Publication vs. withhold**
From: amount of published data and time of publication is decided individually by administration, often access on request only.
To: all data, which is not subject to security or privacy constraints, are published proactively in a timely manner
- **Free vs. limited use**
From: published data is available for private use, all other usages need to be arranged separately
To: published data is available for any use, including reuse and further processing

Internet & Gesellschaft Co:llaboratory: Offene Staatskunst - Bessere Politik durch Open Government ?, Abschlussbericht, 1. Auflage, Berlin 2010.



Open Source, Open Standards have been, Open Data are upcoming

- European PSI directive in 2003: opening up government data to the public
- Sunlight Foundation, US: principles of open data



Public data shall be considered open if they are made public in a way that complies with the principles below:

1. Complete
2. Primary
3. Timely
4. Accessible
5. Machine processable
6. Non-discriminatory
7. Non-proprietary
8. License-free

Compliance must be reviewable.

Data in the conventional view on the knowledge hierarchy (CSC: Uncharted Water, The State of Open Data in Europe, Jan. 2011, adapted from Tuomi; I (1999) "Data is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organizational memory", Journal of Management Information Systems, 16, 3, 107-121.)

- Berlin Open Data Agenda, May 2011
- Berlin Open Data Strategy, February 2012



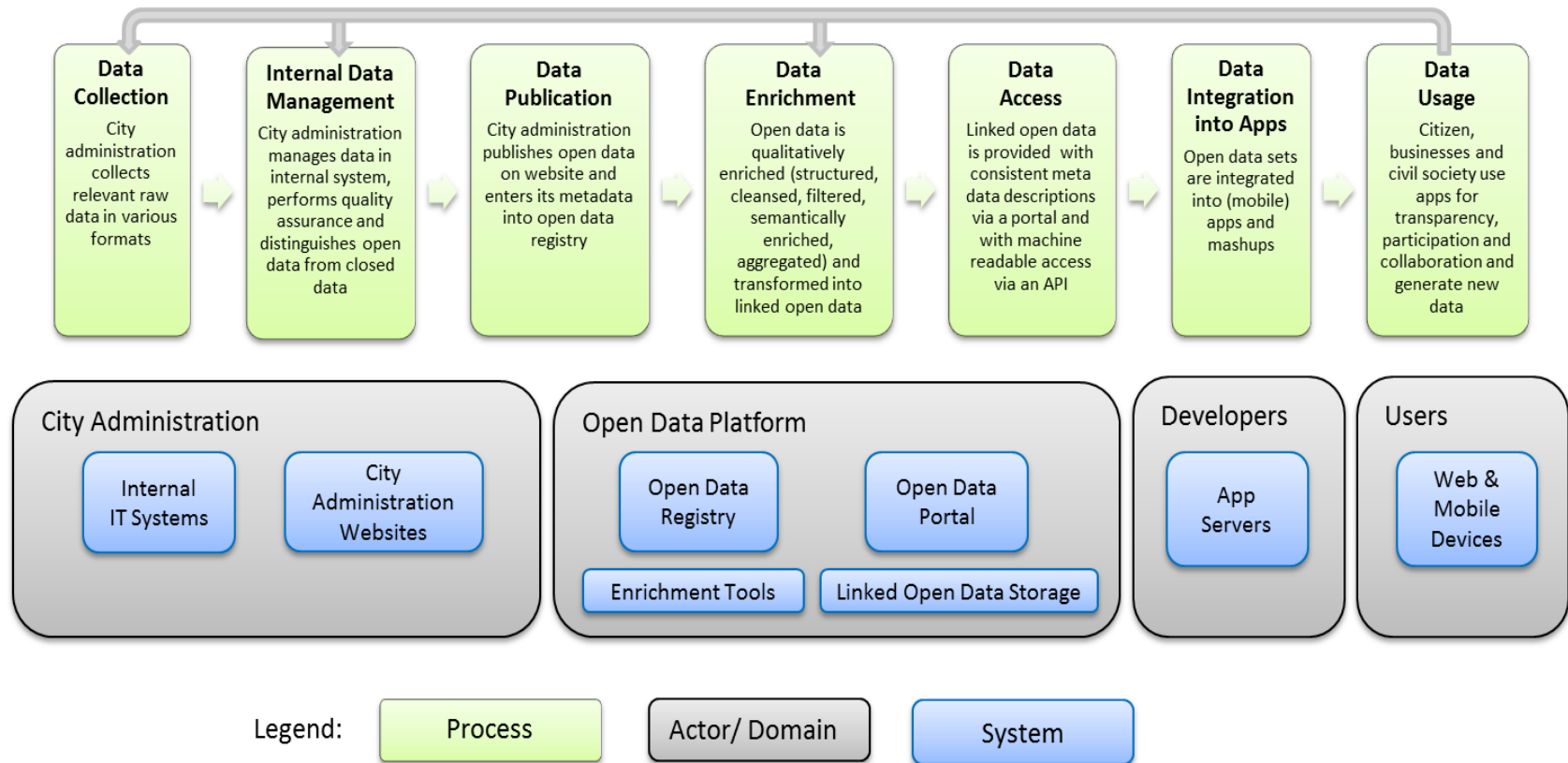
OPEN Innovation for future Internet enabled services in Smart CITIES

Open Cities = Open Innovation + Future Internet + Smart Cities

- **5 Cities** – Helsinki - Berlin – Amsterdam – Paris - Barcelona.
- **5 Mechanisms** – Living Labs - Crowdsourcing – Open Data – Fiber to the Home – Open Sensor Networks
- **5 Objectives**
 - Research on Open Innovation in the Public Sector.
 - Living Labs in Smart Cities : Urban Labs
 - Role of Technological Platforms in Open Innovation
 - panEuropean platforms for Open Data & Open Sensors
 - Deployment of Future Internet Services (mobile & Fiber)

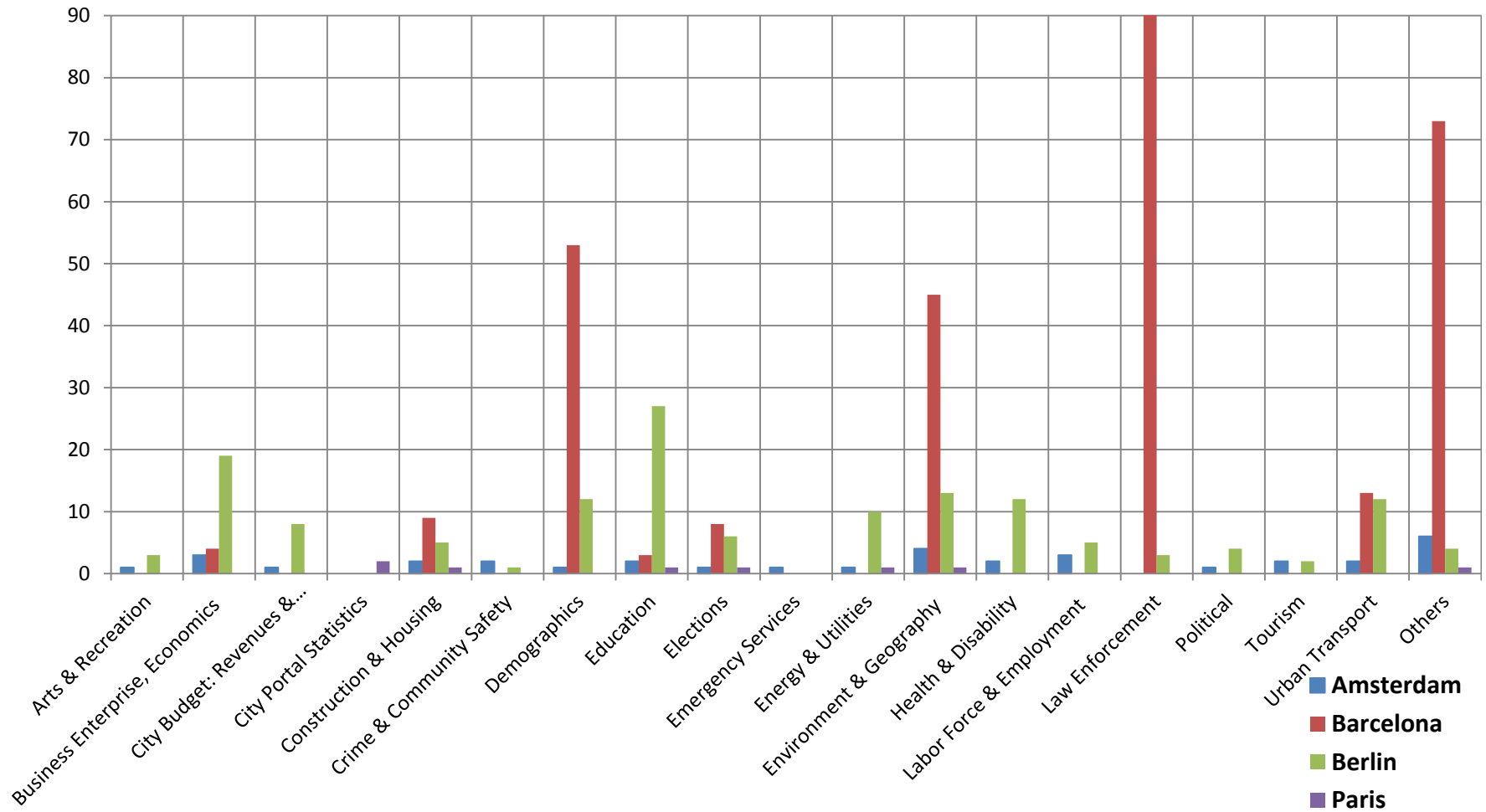
5 Cities x 5 Mechanisms x 5 Objectives

Generic Scenario for Open Data Provisioning

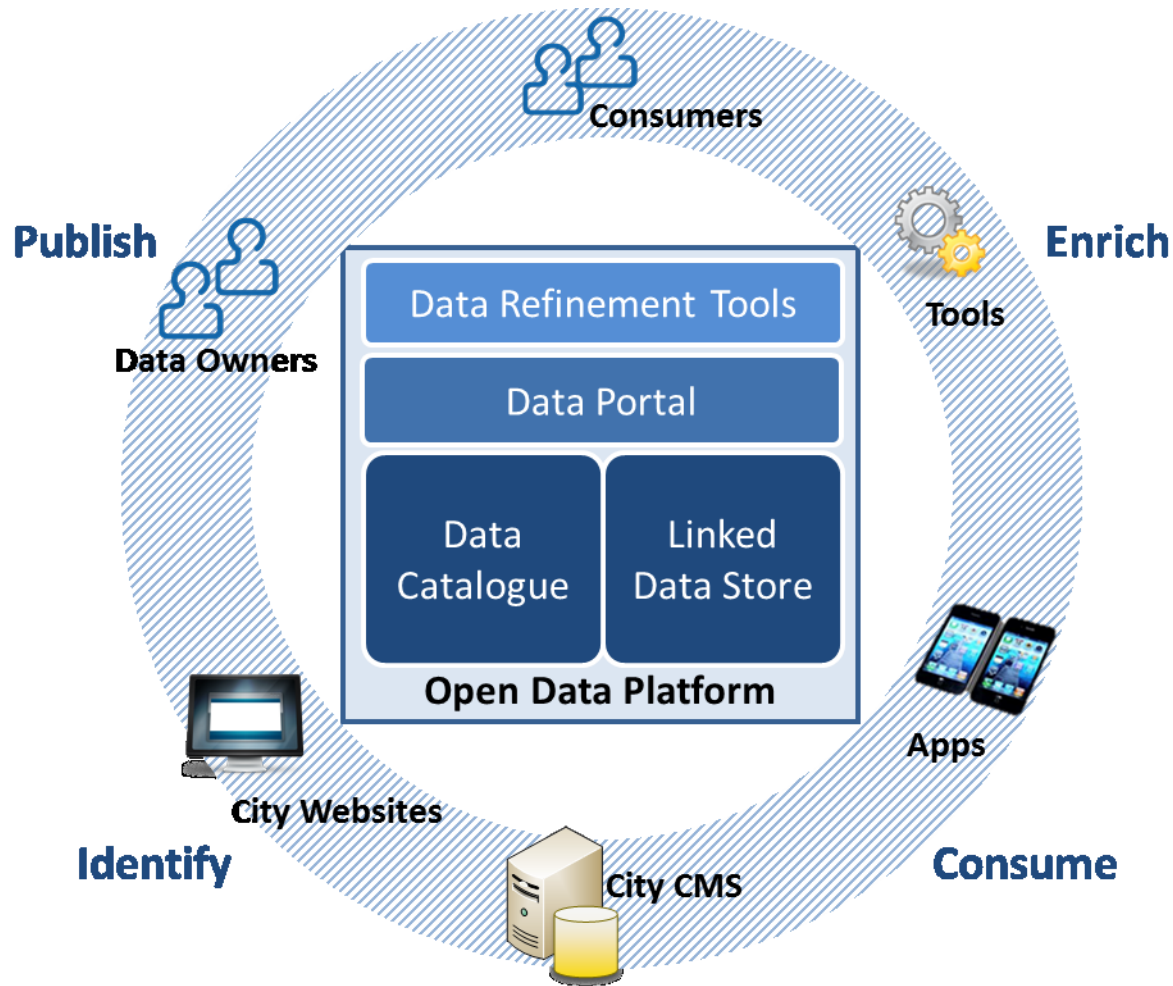


Open Cities Datasets Survey Results

Datasets per categories as reported by cities



The Open Data Process



Open Data Catalogue: Handling Metadata

- The key idea of an open data catalog is to provide data seekers/consumers with a “one-stop-shop” user experience.
- Main purpose of a data catalog is to gather all metadata regarding widely dispersed datasets at a central place.
- The metadata mainly includes attributes such as
 - the dataset’s
 - name
 - description and the URL of the actual resources i.e., files or service end points
 - ...
- The second important function is complementary to the gathering of information. It is to enable the federation of different data catalogues, such as catalogues operating on different administration levels, or topic specific ones.

Open Data Catalogue: Handling Metadata

- Following requirements for the catalogue software result from these use cases:
 - metadata should be stored and handled based on a well defined syntax and semantics, i.e. a documented schema
 - the catalogue software must offer both a user interface and a widely accepted application programming interface for access by other software like applications and data portals

- Regarding metadata semantics: DCat based on Dublin Core

Open Data Catalogue: Handling Metadata

- Comprehensive Knowledge Archive Network (CKAN) is the most popular existing solution and is developed by the Open Knowledge Foundation.
- CKAN fulfills the above mentioned requirements. In terms of
 - interfaces it offers a web access and a well documented JSON based REST API based
 - Metadata scheme inspired by CKAN
 - Additionally, CKAN users are invited to amend arbitrary

The screenshot shows the OpenCities Data Catalogue website. At the top, there is a navigation bar with the logo 'OPEN CITIES' and links for 'Home', 'Search', and 'Categories'. A search bar labeled 'Find datasets' and a 'Login' link are also present. Below the navigation bar is a large banner image of a sailboat. The main heading reads 'Welcome to OpenCities Data Catalogue!'. There are three main content blocks:

- Find data:** Includes a search bar with the text 'Find datasets' and a magnifying glass icon. Below it, it states 'OpenCities contains 784 datasets that you can browse, learn about and download.'
- Find datasets by City:** Includes a magnifying glass icon and lists datasets for Amsterdam, Barcelona, Helsinki, and Paris.
- Collaborate:** Includes a gear icon and lists city portals for Amsterdam, Barcelona, Berlin, Helsinki, and Paris.

 Below these blocks is a section titled 'Who else is here?' with three categories: 'Business Enterprise, Economics, and Trade' (35 datasets), 'Construction, Housing, and Public Works' (27 datasets), and 'Environment, Geography and Meteorological' (23 datasets). At the bottom, there is a row of colored squares and a logo.

Linked Data Store

- The management and storage of metadata enables the Open Data platform to only point to data resources, i.e. to integrate them as external links in the overall picture
- The multi-data store allows for managing datasets and responding resources in the emerging platform itself
- Standard Open APIs: pure database solution is not optimal given the small, but troublesome, differences in the implementation of standards throughout the database products
- Go for an RDF (Resource Description Framework) based storage
 - SPARQL queries
 - „Primitive“ XML parsers
 - Sophisticated RDF parsing frameworks
- For the technical realization of the component a linked database is required

Linked Data Store: Virtuoso

- RDF Graphs
- WebDAV,
- Open APIs based on
 - SPARQL and JSON

```

PREFIX td: <http://www.opendata.opencities.net/touristicdata#>
select ?title ?location ?address ?district ?desc ?public_transport ?close_by
|?dbpedia ?telephone ?openinghours ?disabled ?entrancefees ?URL ?pic ?email where {
?s td:hasTitle ?title .
OPTIONAL { ?s td:hasGeocoordinates ?location } .
OPTIONAL { ?s td:hasAddress ?address} .
OPTIONAL { ?s td:hasDistrict ?district} .
OPTIONAL { ?s td:hasDescription ?desc} .
OPTIONAL { ?s td:hasPublicTransport ?public_transport} .
OPTIONAL { ?s td:isCloseBy ?close_by} .
OPTIONAL { ?s td:hasDBpediaEntry ?dbpedia} .
OPTIONAL { ?s td:hasTelephone ?telephone} .
OPTIONAL { ?s td:hasOpeningHours ?openinghours} .
OPTIONAL { ?s td:isAccessibleForDisabledPerson ?disabled} .
OPTIONAL { ?s td:hasEntranceFees ?entrancefees} .
OPTIONAL { ?s td:hasWebsite ?URL} .
OPTIONAL { ?s td:hasPictures ?pic} .
OPTIONAL { ?s td:hasEmail ?email} .
filter regex (?title, "The two towers Garisenda e Asinelli", "i")
}

```



Open Cities Open Data Platform

Features by platform user roles

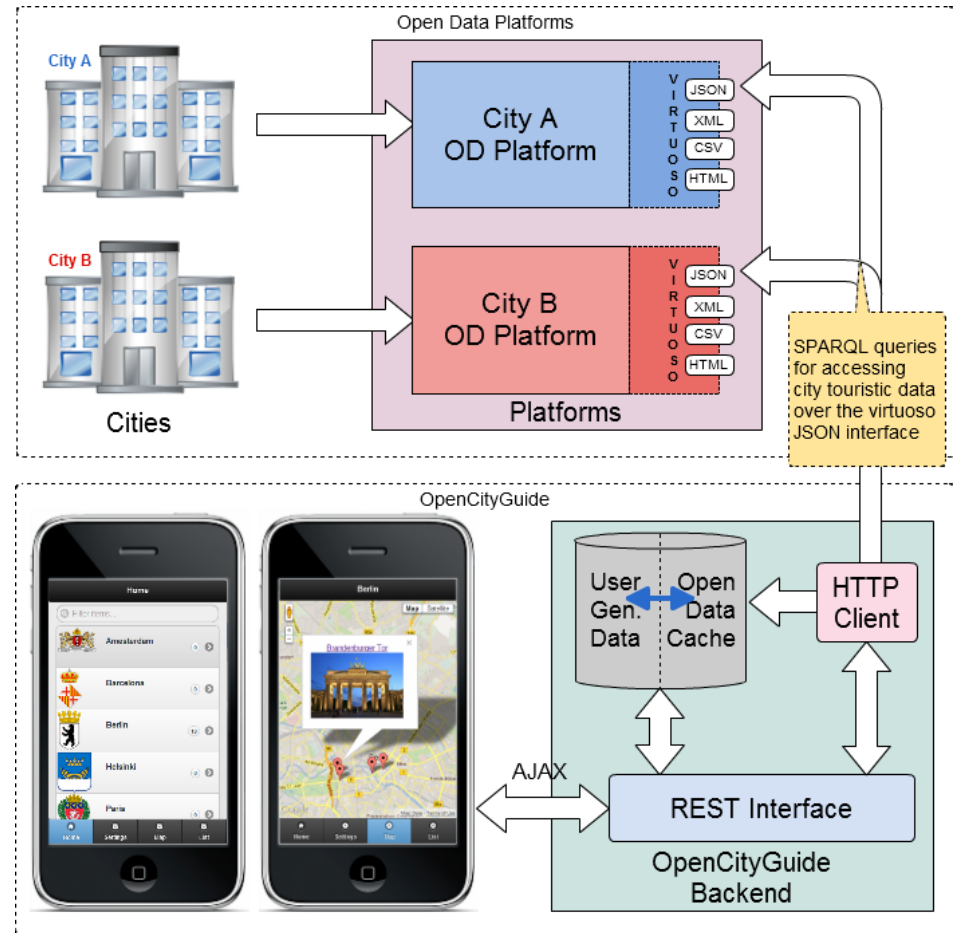
Platform User	Data Owner	Data Steward	Platform Admin
Search and browse datasets	Publish new datasets	Create linked data (Google Refine)	Identify and correct dead links
Download data	Maintain datasets	Define conversion rules to RDF (Google Refine)	Maintain categories and metadata nomenclature
Rate, comment, tweet, and discuss data	Hide datasets	Upload RDF-based linked data	Assure technical quality of datasets
Propose new datasets	Upload bulk datasets	Maintain RDF-based linked data	Administrate user accounts
Query (meta) data via API			
Get notification for new datasets and categories			

Open Cities – Mobile Open Data Mashup

OpenCityGuide Architecture

OPEN CITIES

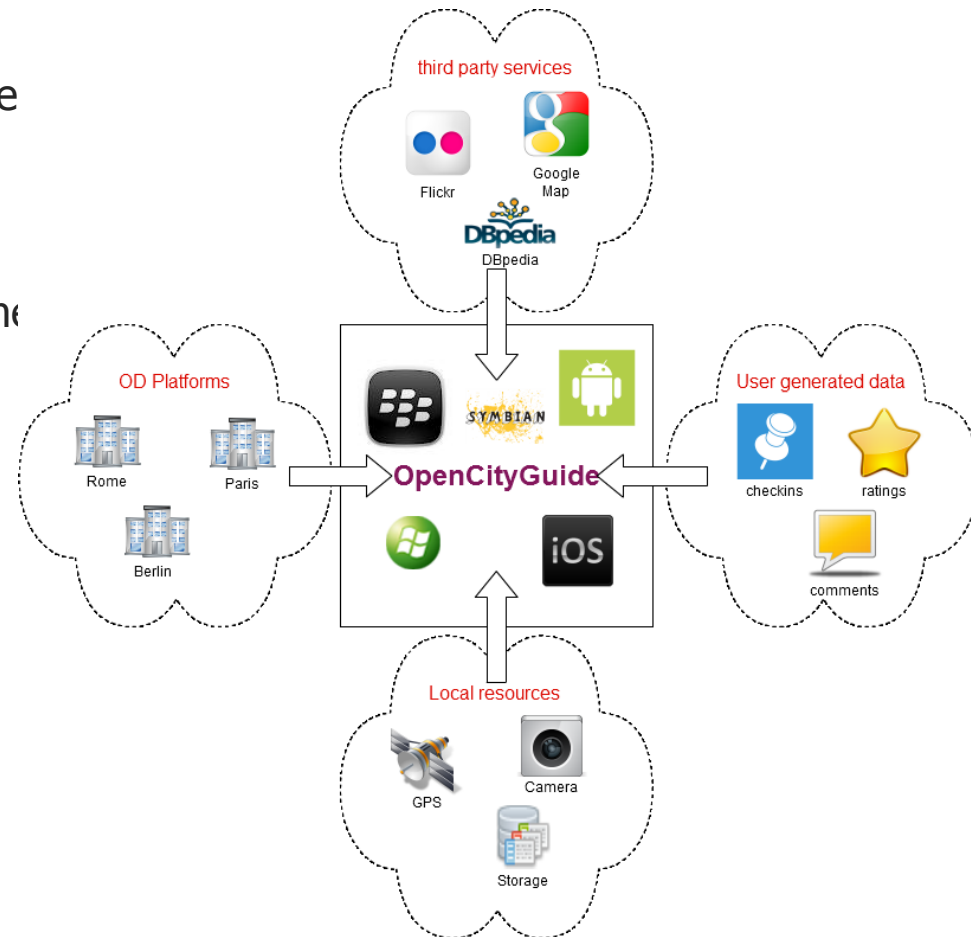
- Cross platform mobile application which make use of touristic data
- OD platforms data access via SPARQL and REST Web services
- The Data format - JSON



Open Cities - Mobile Open Data Mashup

OpenCityGuide Mashup

- OpenCityGuide aggregates data and use services from different sources
 - Open Data Platforms
 - Aggregate touristic data from the different open data platforms of the involved cities using Virtuoso
 - Third party services
 - Flickr
 - Google Map
 - DBpedia
 - User generated data
 - Comments
 - Ratings
 - Local resources
 - Geolocation
 - Local Storage
 - Camera



Summary & Outlook

- Open Data Platform for the city of Berlin and for Germany, various European projects on Open Data
- New Open Data Platform based on lessons learned
- Standard APIs available for application development
- Expertise and practical experiences in developing and deploying Open Data offers

- Improvement of the platform and investigation of its application to the area of eMobility
- Investigation of trust issues around the Open Data process
 - Development of a trust ontology for Open Data
- Various new features are planned
 - Crawling features – automatic identification of potential Open Data sets
 - Different types of metrics for judging on the quality of Open Data sets

