The ADEQUATe platform: data- and community-driven quality improvements

Search, Filter, Fork, and Link Open Data

Sebastian Neumaier, Lőrinc Thurnay, Thomas J. Lampoltshammer, Tomáš Knap
The project is based on two observations...

• An increasing amount of Open Data becomes available
• The integration of such open, freely re-usable data sources is a key success factor

... and two identified issues:

• The quality issues with metadata and the data itself
• The lack of interoperability between data sources

✓ 2 Austrian Use Case Portals: data.gv.at & opendataportal.at
✓ **Four focus group interviews, addressing different stakeholder**

- Lay-persons @ Vienna Open Data meet-up
- Open (Government) Data interested @ GovCamp Vienna
- Platform providers @ Austrian OGD Plaform meeting
- Scientists @ IRIS conference 2016

➢ 106 attendees

✓ **Online based survey**

- Data quality issues on Austrian data portals
- Sent to various channels: community mailing lists, social media groups, etc.

➢ 183 participants
Survey Results

Users want: ¹

- Consistent data formats, complete and correct descriptions, up-to-date data, clarity about legal aspects
- See issues, changes and development of datasets over time
- Find all resources for a certain entity (e.g. city), filter based on concepts
- Discuss and collaborate on datasets, fork/work on own versions

¹ We already filtered out requirements that are beyond our control, e.g., that users want the public data to be qualified as correct.
The Platform

Data Monitoring
QA & QI
Linkage & Search
Community Portal
(i) Data Monitoring

**Monitors** and **reports** the quality of the use case partner portals

Monitors and reports the quality of the use case partner portals.
(ii) Quality Assessment ...

Does the meta data contain information to contact the data provider or publisher?
- Contact: Yes

Is the used license conform to the open definition?
- License Opennness: Yes

Does the meta data contain information about format, size or update frequency of the resources?
- Preservation: 0.50
  - Information (size, format, mimetype,..) for preserving/archiving the dataset resource are missing

Are the available values of contact properties valid HTTP URLs?
- ContactURL: No
  - The publisher or contact URL is not a syntactically valid URI

Is the file format based on an open standard?
- Format Opennness: Yes

Does the meta data contain access information for the resources?
- Access: Yes

Are the available values of access properties valid HTTP URLs?
- AccessURL: Yes

Does the meta data contain information about the license of the dataset or resource?
- Rights: Yes
(ii) ... & Quality Improvements

**Available Services**

**CSV Clean**
Parse & clean
A service to parse and clean various types of CSV derivations (or character-separated-values files). The cleaned file is UTF-8 encoded and uses the ""," as value separator.

**CSV Profiler**
Metrics & statistics
This service analyses the input CSV and provides basic informations and metrics such as the encoding, column data types, and the completeness of columns.

**CSV MetaData Editor**
Describe & enrich
This is a prototypical form for generating metadata about a submitted CSV file. The metadata is compliant with the CSV on the Web metadata specification.

---

www.adequate.at
(iii) Linkage ...

**Odalic [1]:**
- Semantic interpretations of tabular data
- Annotations of (i) columns, (ii) cell values, and (iii) relations between columns
- Re-publishing as RDF

**Using ADEQUATe knowledge base**
(Austrian-specific knowledge)

[1] Tomáš Knap. Towards Odalic, a Semantic Table Interpretation Tool in the ADEQUATe Project. In LD4IE Workshop 2017
(iii) ... & Search

**PoolParty GraphSearch:**

- Full text search
- Faceted search, based on entities in ADEQUATe knowledge base
- **TODO:** Integrating Odalic annotations in the GraphSearch

---

1 https://www.poolparty.biz/poolparty-semantic-graph-search-server/
Exemplary User Journey
Katalog
Hunde pro Bezirk Wien

Anzahl der Hunde pro Bezirk und Rasse

Daten und Ressourcen

|Titel und Beschreibung Englisch| Number of dogs per district and race. |
|Veröffentlichende Stelle| Stadt Wien |
|Kontaktseite der veröffentlichten Stelle| https://open.wien.gv.at |
|Veröffentlichende Stelle - E-Mailkontakt| open@post.wien.gv.at |
|Datenverantwortliche Stelle| Magistratsabteilung 6 - Rechnungs- und Abgabenwesen |
Stadt Wien

Hunde pro Bezirk Wien

The information below and the linked dataset(s) are provided by the ADEQUATe Open Data project. The information and data were harvested from data.gv.at, analysed and may have been automatically and manually changed.

Description

Anzahl der Hunde pro Bezirk und Rasse

Published
📅 12/15/2015
📅 1/5/2018
📍 Stadt Wien
🕴 Magistratsabteilung 6 - Rechnungs- und Abgabenwesen
✉️ post@ma06.wien.gv.at
☑ Status of the maintainer's datasets

Keywords
#statistik  #hunde

Downloads

Anzahl der Hunde pro Bezirk
🔗 CSV  · 📑 369.2 kB  · 🎈 cc-by-at-30  · 📅 12/15/2015  · 📝 DefaultDelimeter, DropCommentLines

Original from www.wien.gv.at: Anzahl der Hunde pro Bezirk

Semantic linking
<table>
<thead>
<tr>
<th>Name</th>
<th>Last commit</th>
<th>Last Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>cleaned</td>
<td>1750</td>
<td>4 months ago</td>
</tr>
<tr>
<td>resources</td>
<td>1750</td>
<td>4 months ago</td>
</tr>
<tr>
<td>README.md</td>
<td>1811</td>
<td>a month ago</td>
</tr>
<tr>
<td>metadata.jsonld</td>
<td>1815</td>
<td>6 days ago</td>
</tr>
<tr>
<td>original.json</td>
<td>1802</td>
<td>3 months ago</td>
</tr>
</tbody>
</table>
Outlook & Open Challenge

Platform, source code, and data online available:

• www.adequate.at

Outlook:

• Integration of other QI tools, e.g. Open Refine
• Continuation in substantively related Austrian project CommuniData: www.communidata.at

Open Challenge:

• We have the tools, services & platform...
How to get the users?
Outlook & Open Challenge

Platform, source code, and data online available:

- www.adequate.at

Outlook:

- Integration of other QI tools, e.g. Open Refine
- Continuation in substantively related Austrian project CommuniData: www.communidata.at

Open Challenge:

- We have the tools, services & platform... How to get the users?

Thank you!
sebastian.neumaier@wu.ac.at
https://sebneumaier.wordpress.com/
twitter: @sebneum