

Der nächste Schritt? Linked Open Data und Visualisierung beim Catalogue of Digital Editions

Greta Franzini Peter Andorfer

MWW-DARIAH-DE Workshop Suchtechnologien
May 23-25, 2018



Introduction

Digital edition

Reconstruction of a text which makes use of digital technologies to reproduce any level of detail of that text.

Catalogue of Digital Editions

- ▶ Created in 2011 for personal use only (Greta's PhD)
- ▶ Became a collaboration with the ACDH in 2016
- ▶ Team: Peter Andorfer, Ksenia Zaytseva and Greta Franzini
- ▶ No funding
- ▶ Accessible at: <https://dig-ed-cat.acdh.oew.ac.at/>

Related work

- ▶ *Patrick Sahle's Scholarly Digital Editions: An Annotated List*
 - ▶ **Scholarly** digital editions, i.e., strong **critical** component
 - ▶ ca. 420 entries (as of 22 March 2017)
- ▶ *RIDE | A review journal for digital editions and resources*
 - ▶ Digital **editions and resources** (e.g., text collections)

Research questions, method and goals

Research questions

- ▶ How are digital editions being **created** [3]?
- ▶ Are creators meeting the needs of **users**?¹

Method

- ▶ **Qualitative**: descriptive cataloguing according to 49 criteria
- ▶ **Quantitative**: collection of large body of evidence

Goals

- ▶ **Search** and **rich comparison** of digital editions
- ▶ **Popularization** of digital editions
- ▶ **Definition of (minimum) requirements** of digital editions [2]

¹Franzini, G., Terras, M., Mahony, S. (forthcoming 2018/2019) 'Digital Editions of Text: Surveying User Requirements in the Digital Humanities', *ACM Journal on Computing and Cultural Heritage (JOCCH)* - Special Issue on Evaluation of Digital Cultural Resources, 11(4)/12(1).

Potential

*Though limited to its own particular subset of digital publishing activity, [the] Catalogue comprises a dataset of some 230 digital editions, currently, with some fifty consistent and comparable pieces of data on each, that range from the edition's subject matter and URL, to its features, textual encoding scheme, and technological infrastructure. While it takes a more object, data-focused approach to reviewing and cataloguing the included editions, the Catalogue also uniquely offers the possibility of rich comparison and analysis across publications, even if that more subjective and evaluative work is yet to be done. **It may also someday provide a model to be applied to the evaluation of other types of digital publishing projects**, specifically like the Mellon-funded university press projects, the Getty's OSCI collaborative, and other open access, scholarly editions which have been the subject of our history here thus far.*

–Greg Albers, The J. Paul Getty Trust, 2017 [1]

Research support

Specific research that the Catalogue supports includes:

- ▶ Tracking the movement of documentary heritage across time and space
- ▶ Keeping a record of discontinued projects
- ▶ Finding reusable texts or corpora

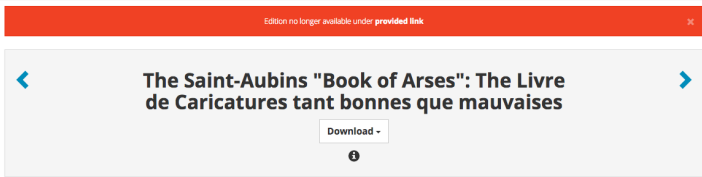


Figure: Digital edition no longer available under the previously provided link (404 error).

Web Application: How is it built?

- ▶ Relational database as storage (**MariaDB**) for web app
- ▶ **GitHub** used as actual storage for data as well as interface for data curation/creation
- ▶ **Django (Python)**
 - ▶ **django**baseproject
- ▶ Reuses open source code
- ▶ Licence: MIT (web app); Attribution 4.0 International (CC BY 4.0) (data)
- ▶ At ACDH-OeAW dockerized server-infrastructure
- ▶ Part of ACDH-OeAW backup strategy (daily in and out house backups)

Web Application: What does it facilitate?

Users can...

- ▶ **Search**, browse, **visualize** and filter the data around their research interests
- ▶ **Contribute** and edit data
- ▶ **Download** the data in bulk (JSON, CSV, BibTex, N3, RDF)
- ▶ **Download** custom data selections
- ▶ **Query** Linked Open Data (GeoNames, Dublin Core Metadata, Data Catalog, FOAF, GND) with inbuilt SPARQL endpoint interface

Linked Open Data

Implementation

- ▶ Mapping done 'by hand'
 - ▶ Python script
 - ▶ No extra (mapping) tool like e.g., Karma or ontoop needed
- ▶ Generated data dumps uploaded in dedicated triple store (Blazegraph)
- ▶ Dedicated triple store NOT public due to security issues (we try to avoid accepting DELETE/UPDATE requests)
 - ▶ User asks django-web-app; django-web-app passes requests forward to Blazegraph's query interface

Linked Open Data

Issues

- ▶ Relational database and triple store are out-of-sync
 - ▶ Solution 1: Import scripts triggers rdf-dump → blazegraph (update)
 - ▶ Solution 2: Remove triple store and data duplication

ToDoS

- ▶ Evaluate **d2rq**
 - ▶ pro: no triple store needed
 - ▶ pro: no data duplication
 - ▶ con: not much experience so far, e.g., federated queries possible? reasoning? scalability?
 - ▶ con: do some mapping again...

<https://sparql-dec.acdh-dev.oeaw.ac.at/> (not public yet)

Impact

Indexing and syndication

- ▶ Datenbank-Infosystem (DBIS): 300+ libraries
- ▶ OpenAIRE: Zenodo

International visibility

- ▶ User contributions (e.g., 40+ entries from Hungary)
- ▶ Refbacks (e.g., #dariahTeach MOOC)
- ▶ Press (e.g., Rai.tv Cultura & Scuola, iDNES.cz)
- ▶ DH Awards 2016: 2nd Runner Up as *Best Data Visualization*

Next steps

Upcoming developments

- ▶ **Handle.Net** IDs per Catalogue entry for stable addressability
- ▶ Collaboration with **Digitized Medieval Manuscripts App**

Desiderata

- ▶ Compliance with library standards
- ▶ LOD mapping to VIAF, ORCID, LoC, EDM, RDA
- ▶ Use of **RambleOn** to visualise movement of heritage
- ▶ Connect to **Palladio** via SPARQL
- ▶ Small grant to help with data creation and curation

Credits

Peter Andorfer

- ▶ peter.andorfer@oeaw.ac.at
- ▶ <https://orcid.org/0000-0002-9575-9372>

Greta Franzini

- ▶ g.franzini.11@ucl.ac.uk
- ▶ <https://orcid.org/0000-0003-1159-5575>

Ksenia Zaytseva

- ▶ ksenia.zaytseva@oeaw.ac.at

<https://doi.org/10.5281/zenodo.1250796>

References



G. Albers.

Bringing Books Online.

MW17: Museums and the Web 2017, 2017.

[https:](https://mw17.mwconf.org/paper/the-next-generation-of-online-publishing-building-on-what-weve-learned-together/)

[//mw17.mwconf.org/paper/the-next-generation-of-online-publishing-building-on-what-weve-learned-together/.](https://mw17.mwconf.org/paper/the-next-generation-of-online-publishing-building-on-what-weve-learned-together/)



F. Fischer.

Digital Corpora and Scholarly Editions of Latin Texts: Features and Requirements of Textual Criticism.

Speculum, 92:S265–S287, 2017.

[https://doi.org/10.1086/693823.](https://doi.org/10.1086/693823)



G. Franzini, M. Terras, and S. Mahony.

A Catalogue of Digital Editions.

In E. Pierazzo and M. J. Driscoll, editors, *Digital Scholarly Editing: Theories and Practices*, Digital Humanities, pages 161–182. Open Book Publishers, Cambridge, 2016.

[https://doi.org/10.11647/OBP.0095.09.](https://doi.org/10.11647/OBP.0095.09)