Dataverse in Göttingen

data.goettingen-research-online.de

European Dataverse Workshop 2020 (Tromsø, 2020-01-23)

Péter Király, Göttingen eResearch Alliance

institutional background

University of Göttingen
+ 5 Max Planck Institutes
+ 3 other research institutes
+ 11 associate partners
= Göttingen Campus
goettingen-campus.de

Göttingen eResearch Alliance

State and University Library
+ GWDG (campus IT facility)
→ Göttingen eResearch Aliance
eresearch.uni-goettingen.de

Göttingen Research Online

GRO.data – Dataverse
+ GRO.pub – publications
+ GRO.instruments – large equipment portal
+ GRO.plan – RDM planning
= Göttingen Research Online

intended usage

- general repository
- for data publication
- open for all Campus members

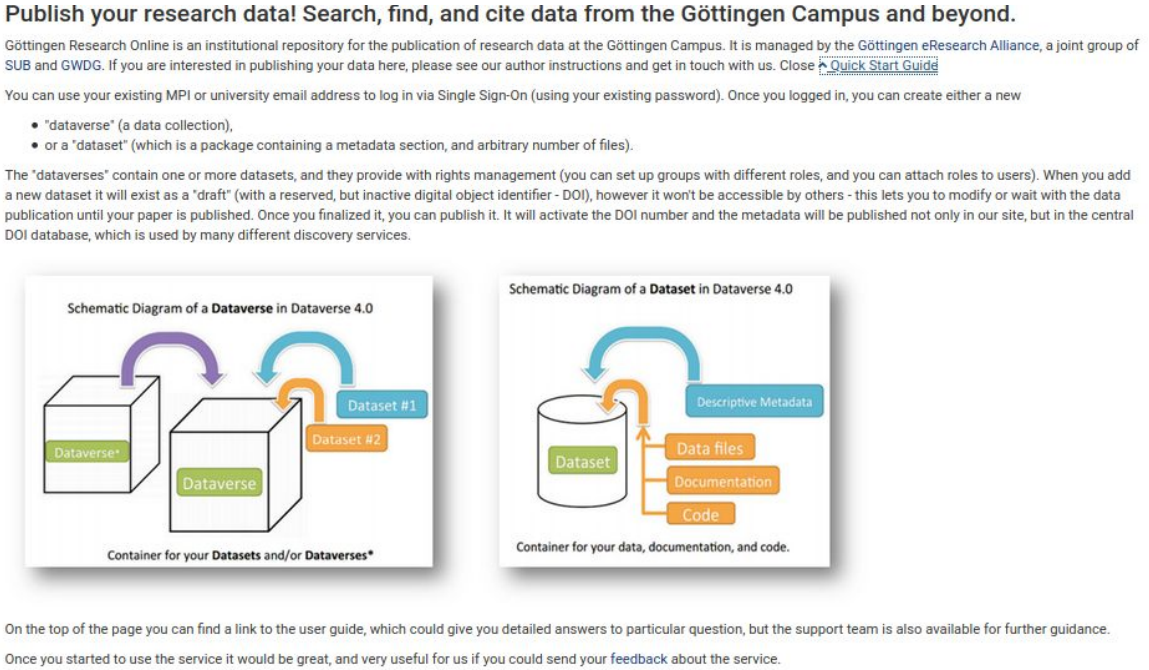
restrictions

- use GRO.data only if there is no special repository for your data
- for collaborative work on data use different platforms (e.g. file sharing, git, collaborative editing etc.)
- contact us if you have too much, extra large files

customization – frontend

Minimal outlook change

- logos, colors etc.
- quickstart quide
- HTML page manipulation with jQuery



Publish your research data! Search, find, and cite data from the Göttingen Campus and beyond.

Göttingen Research Online is an institutional repository for the publication of research data at the Göttingen Campus. It is managed by the Göttingen eResearch Alliance, a joint group of SUB and GWDG. If you are interested in publishing your data here, please see our author instructions and get in touch with us. Close ^ Quick Start Guide

You can use your existing MPI or university email address to log in via Single Sign-On (using your existing password). Once you logged in, you can create either a new

- "dataverse" (a data collection),
- or a "dataset" (which is a package containing a metadata section, and arbitrary number of files).

The "dataverses" contain one or more datasets, and they provide with rights management (you can set up groups with different roles, and you can attach roles to users). When you add a new dataset it will exist as a "draft" (with a reserved, but inactive digital object identifier - DOI), however it won't be accessible by others - this lets you to modify or wait with the data publication until your paper is published. Once you finalized it, you can publish it. It will activate the DOI number and the metadata will be published not only in our site, but in the central DOI database, which is used by many different discovery services.

On the top of the page you can find a link to the user guide, which could give you detailed answers to particular question, but the support team is also available for further guidance.

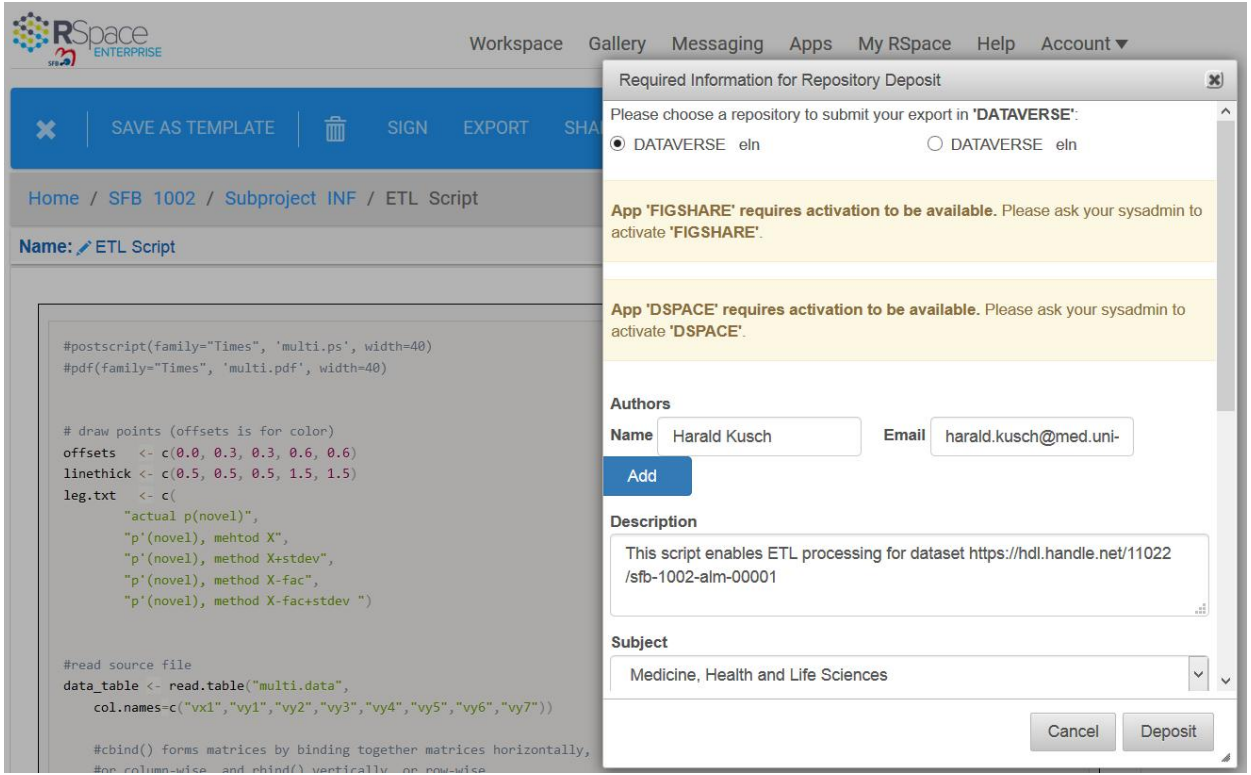
Once you started to use the service it would be great, and very useful for us if you could send your feedback about the service.

customization – backend

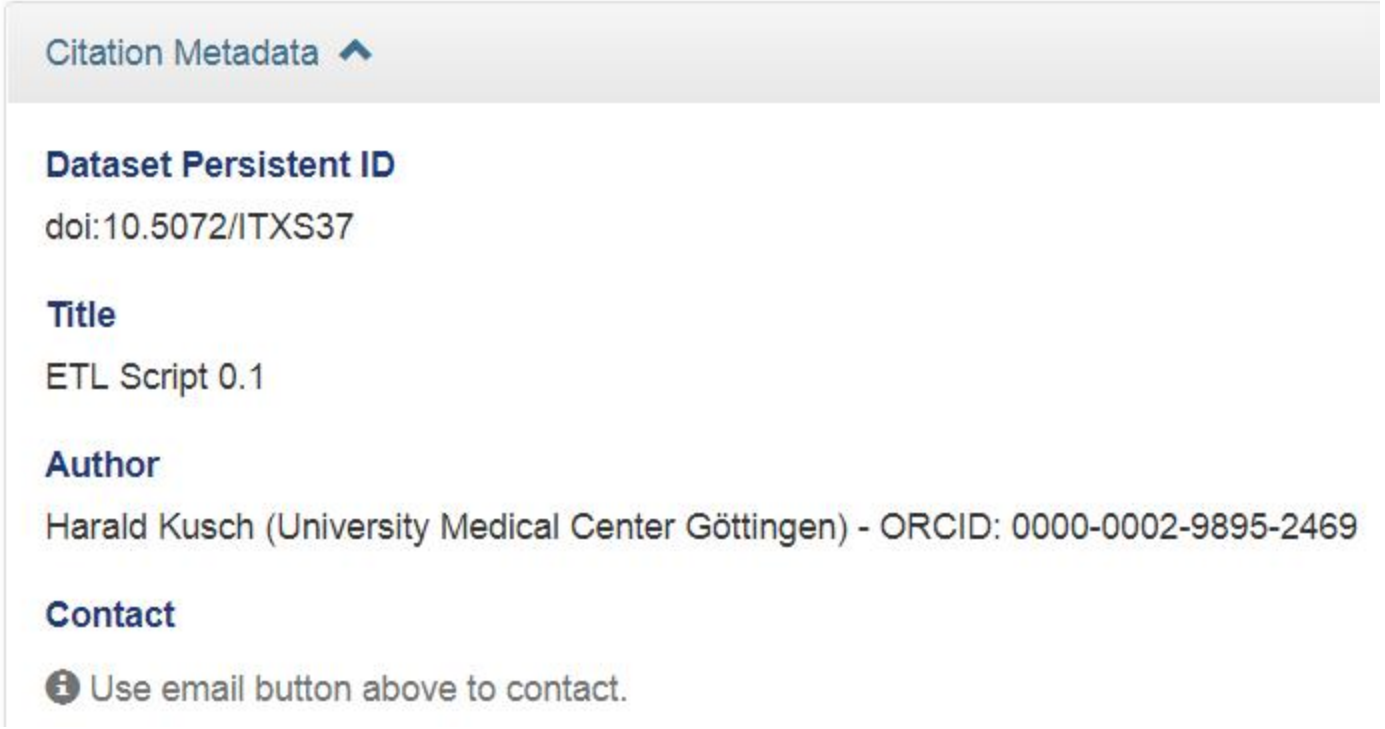
Backend

- latest version (after testing)
- Shibboleth
- adaptation to local IT infrastructure
  - VMware ESXi virtual machines
  - Ubuntu OS (instead of Red Hat)
  - Puppet based management
  - mounted disk (IBM Tivoli Storage Management)
  - archiving, daily backup
  - automatic update of OS
  - monitoring: munin, openITCockpit
  - fixed time window for deployment in production

RSpace electronic lab notebooks



RSpace electronic lab notebooks


Citation Metadata

Dataset Persistent ID

doi:10.5072/ITXS37

Title

ETL Script 0.1

Author

Harald Kusch (University Medical Center Göttingen) - ORCID: 0000-0002-9895-2469

Contact

Use email button above to contact.

testing sessions

purpose: to test Dataverse installation

- multiple users in one room using Dataverse simultaneously
- no previous knowledge
- UI and API tests
- finding differences between existing services on the Campus

collaborations

- Dataverse.DE – one day discussion about the current state of Dataverse in German speaking countries
- Social Sciences and Humanities Open Cloud (SSHOC) (sshopencloud.eu)
- Global Dataverse Community Consortium (dataversecommunity.global/)
- the Harvard Dataverse Team (most frequently w/ Phil Durbin – thanks!)

contributions to community (so far)

- API documentation
- custom metadata block administration (utilizing Solr Schema API #5989)
- clean code
- logging
- participating in discussion in issue queue, mailing list, and IRC

eResearch Lab

purpose: to create an open space for

- sharing knowledge about eResearch
- building communities
- our idol: Berkeley Institute for Data Science (bids.berkeley.edu)

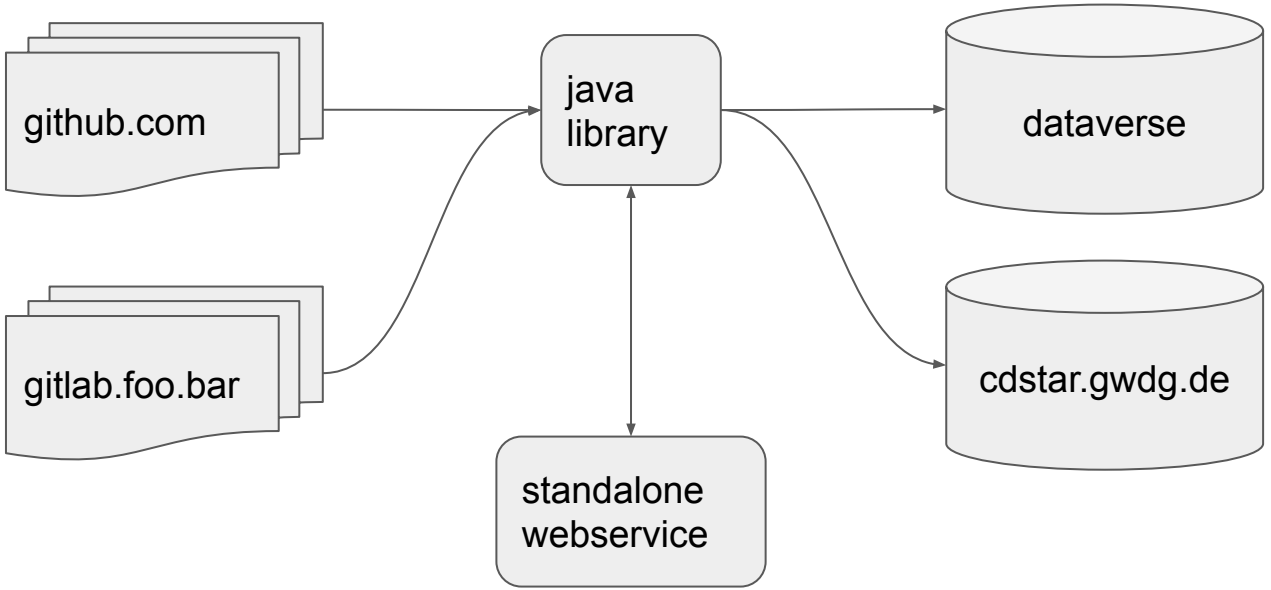
Alpha phase: monthly Dataverse sessions

- introduction (concepts, quick start)
- practice

There are always new use cases, interesting questions

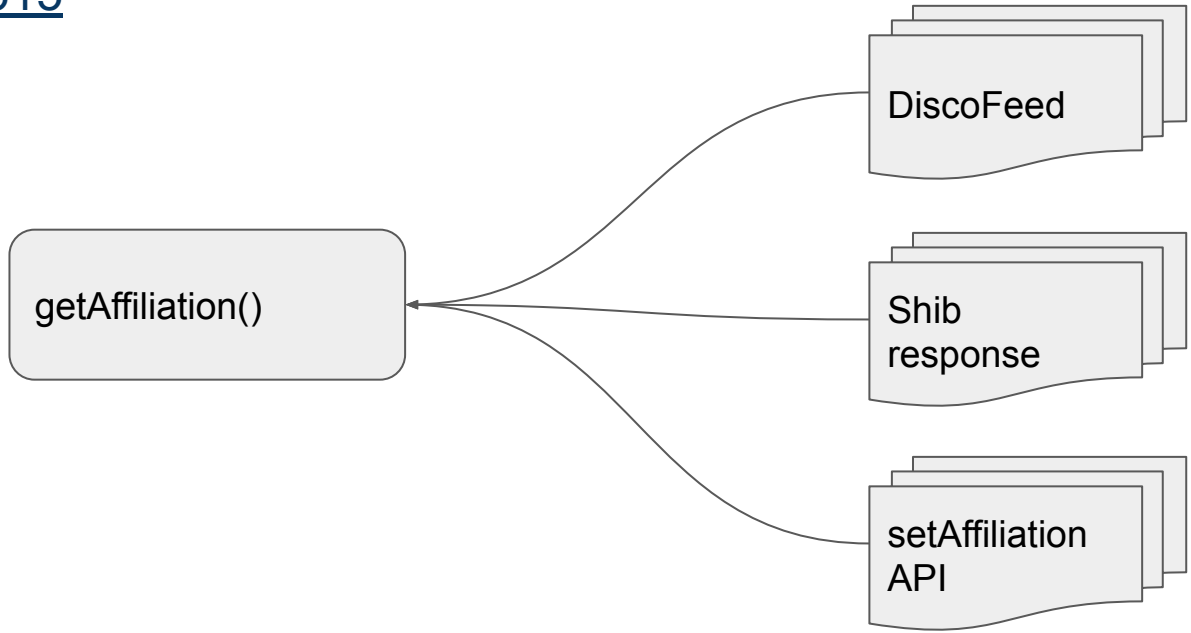
plans – Integration of Git repositories and Dataverse

- branches
- tags
- releases
- every objects (inc. wiki, issues)
- snapshot or autoupdate?

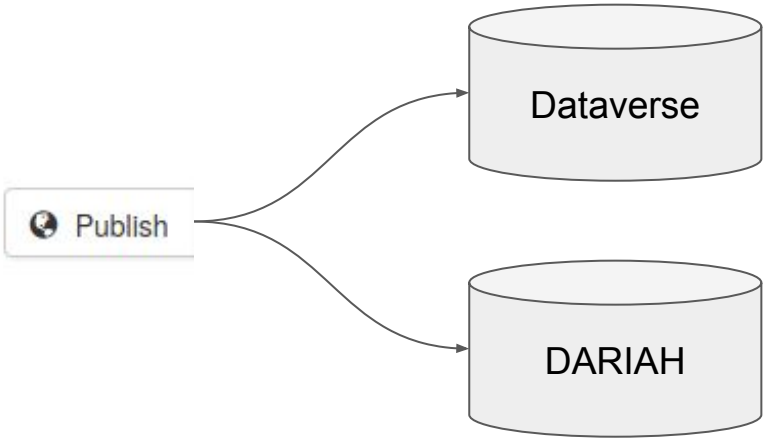


plans – Shibboleth

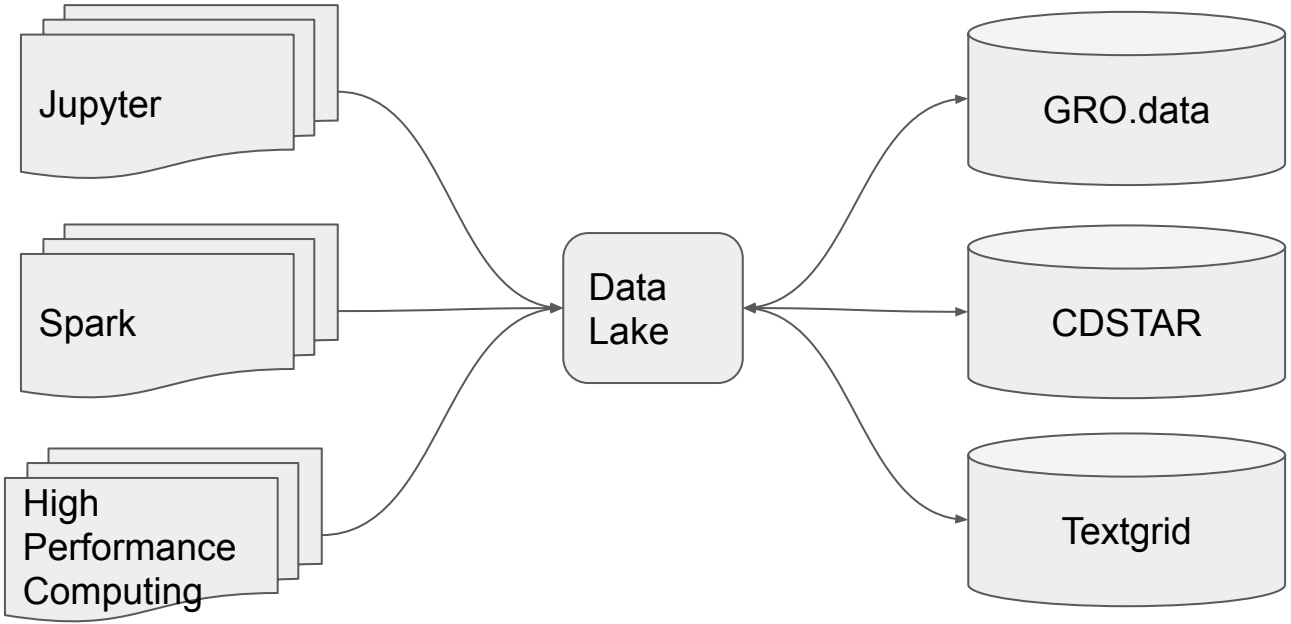
#6514, #6515



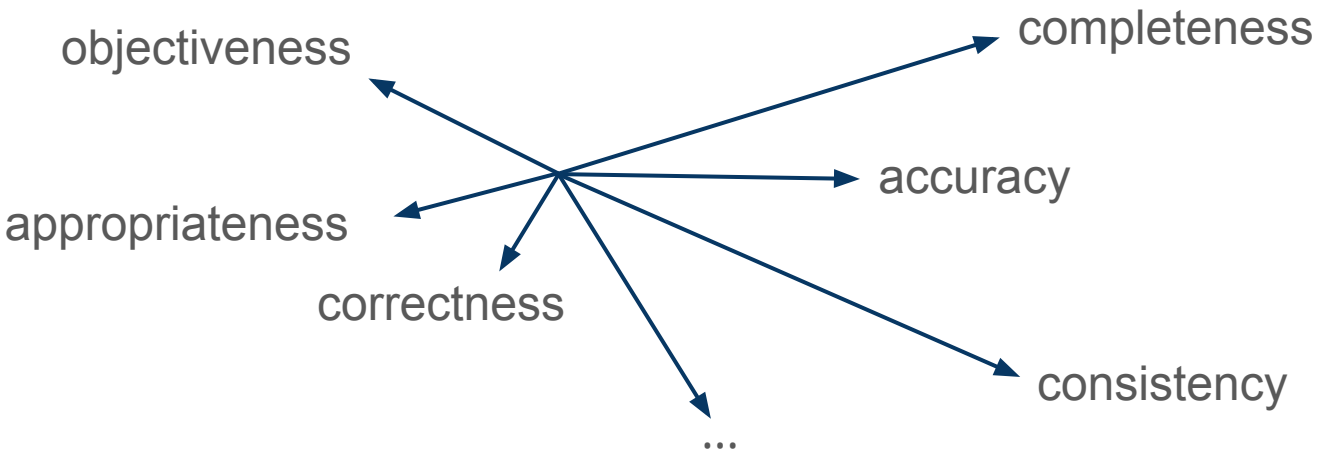
plans – Integration with DARIAH repository



plans – Data Lake for analysis



plans – (meta)data quality assessment

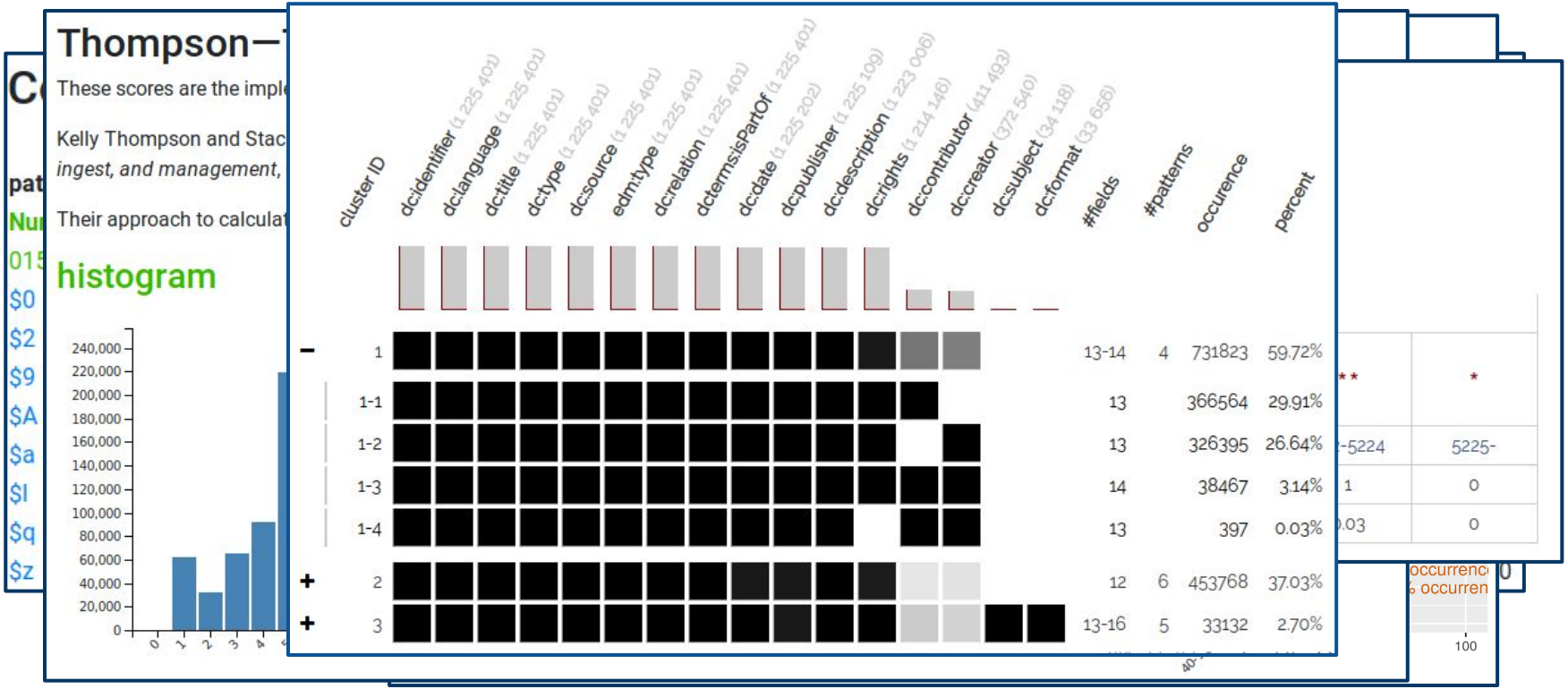


Bruce and Hillman (2004); Ochoa and Duval (2009); Palavitsinis (2014); Zaveri et al. (2015)

zotero.org/groups/488224/metadata_assessment

10.13140/RG.2.2.33177.77920

plans – (meta)data quality assessment



10.13140/RG.2.2.33177.77920 data.goettingen-research-online.de – bit.ly/gro-data-tromso

eResearch Alliance

contact details

Göttingen eResearch Alliance: www.eresearch.uni-goettingen.de

GRO.data: data.goettingen-research-online.de

these slides: bit.ly/gro-data-tromso

emails: info@eresearch.uni-goettingen.de, pkiraly@gwdg.de

twitter: @kiru