

DATA STREAMS IN LINKED.SWISSBIB.CH

THE SWISS METACATALOG IN THE LINKED OPEN DATA CLOUD

The project linked.swissbib.ch integrates the metacatalog swissbib into the semantic web, together with the creation of a Linked Data service. This implies the transformation of the metadata from different Swiss library networks into an RDF based format, its enrichment and its publication. Main challenges arise from the requirement of fully automated processes to allow daily updates.

DATA SOURCES

About 15 library networks or repositories deliver their data to swissbib.

UPSTREAMS

Data is delivered at a daily frequency.

Data reception: deduplication, clustering, ...
[MARC/XML]

Transformation
(with Java/XSLT)

[MARC/XML]
> [SOLR/XML]

Transformation
(with Metafacture)

[MARC/XML] > [JSON-LD]

Preprocessing

[JSON-LD] > [N-TRIPLE]

Interlinking
(with LIMES)

Enrichment

[N-TRIPLE] > [JSON-LD]

SOLR
index

Elasticsearch
index

URI ATTRIBUTION

The first of the four Linked Data principles is realised during the transformation. The (still not resolved) challenge consist of creating persistent URI upon the basis of records deduplicated every day.

DBpedia
dump

VIAF
dump

GND
dump

INTERLINKING

Swissbib data about persons are compared and enriched. To achieve this, it is transformed into the N-Triple format and dumps of the external reference databases are imported, enabling a more efficient process. Main constrains arise here from the sometimes poor data quality and the huge data quantity.

REFERENCES

- BERNERS-LEE, T. Linked Data. *World Wide Web Consortium* [online]. 2010. <http://www.w3.org/Design/Issues/LinkedData.html>
- REDMAN, T. C. *Data driven: profiting from your most important business asset*. Boston, Mass : Harvard Business Press, 2008. ISBN 978-1-4221-1912-9
- SOUTHWICK, S. B., LAMPERT, C. K. and SOUTHWICK, R. Preparing controlled vocabularies for Linked Data: benefits and challenges. *Journal of library metadata*. 2 October 2015. Vol. 15, no. 3-4, p. 177-190. DOI 10.1080/19386389.2015.1099983

DOWNSTREAMS

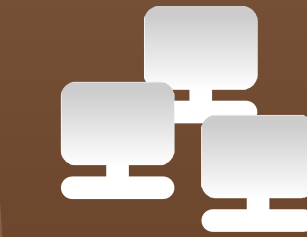
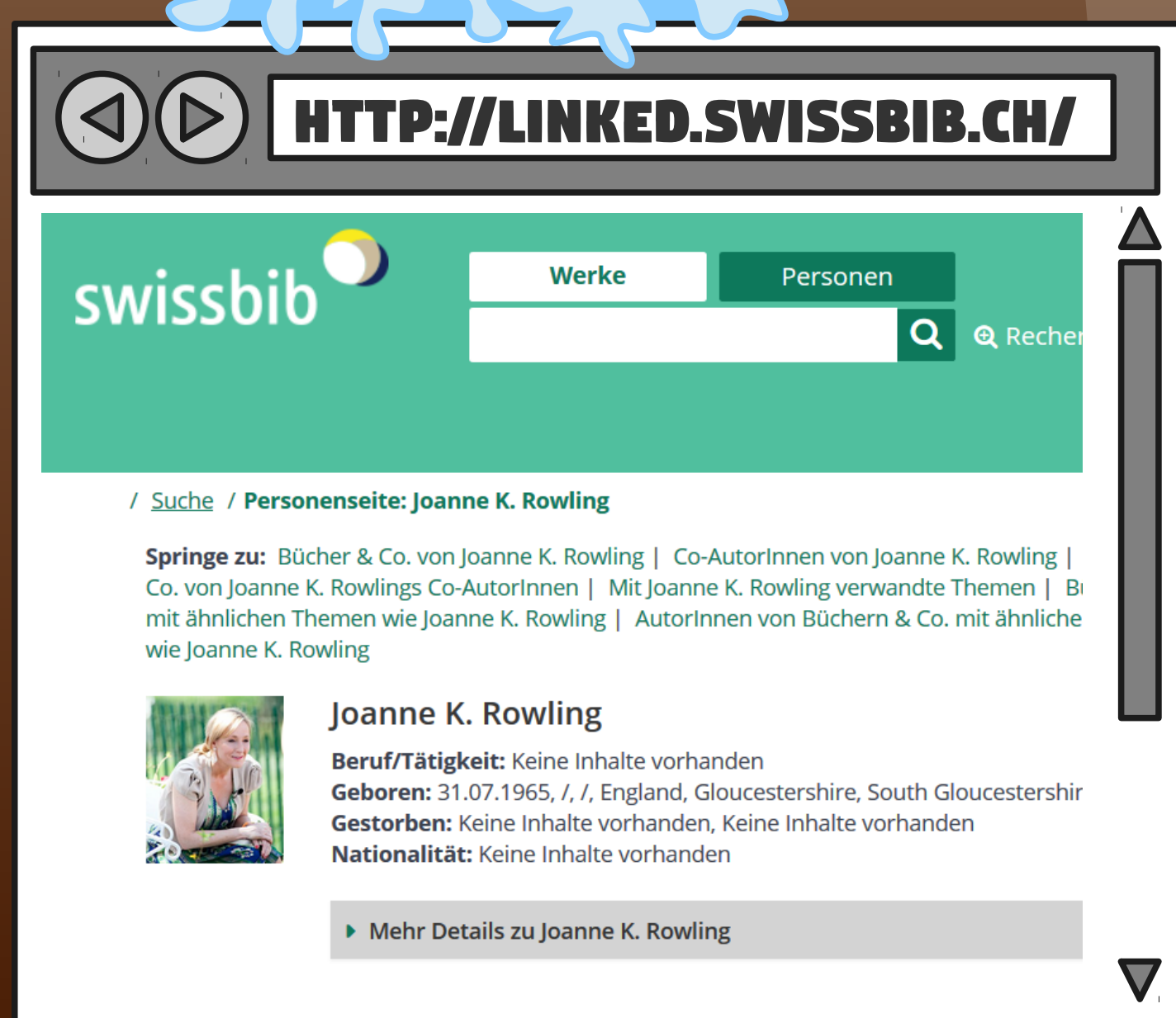
Data are made available for use and reuse.

CCO FILTER

This data are destined for massive reuse, but do nevertheless come from various providers with various terms of use. To avoid any problem of attribution for the reuser, only data licensed under CC0 are accessible via the API.

END USER

Besides the classic interface (not on this poster), the infrastructure provides an interface with a better user experience.



COMPUTER CLIENTS

A RESTful service relying on Hydra provides the data in various RDF serialisations for other applications to encourage data reuse and ensure a high level of interoperability.

h e g

Haute école de gestion
Genève

Universität
Basel

HTW Chur
Hochschule für Technik und Wirtschaft
University of Applied Sciences

gesis
Leibniz Institute
for the Social Sciences

Nicolas Prongué
& René Schneider

