A Library Data Management Platform Based on Linked Open Data
D:SWARM
A Library Data Management Platform Based on Linked Open Data

❖ Back in Those Days
❖ The Age of Discovery
❖ Library Data Management
❖ Qualify, Link and Free Your Data: D:SWARM
❖ Live Demo
Back in Those Days ... Data Heterogeneity

- Multiple individual data silos
  - ILS, document repositories, databases, ...
- Data saved in heterogeneous formats
  - MAB, MARC21, ...
- Each data silo gets processed individually
  - Multiple admin interfaces
  - Multiple search interfaces
  - Data unrelated to one another
- Comprehensive view of resources almost impossible (for users and librarians)
The Age of “Discovery”

Data Normalization

- More comprehensive view of resources for users, but no real discovery/exploration
- Data gets normalized into one storage but not integrated
- Data available in record-oriented structures
  - External data (e.g. GND) has to be squeezed in the record
  - Metadata records are independent of each other
  - No explicit semantic quality of data
Library Data Management
What Libraries Actually Need

⭐⭐⭐⭐⭐ Library Data

- Get rid of data silos
  - Open formats for exchange
- Lossless data integration instead of reductive normalization
- Data integration with entity level granularity
  - Get rid of pre-compiled data records
- Focus on linking entities/objects:
  - Graph structures creating the knowledge graph
- Stick to quality policy of libraries
  - Versioning and provenance of data
What Should Library Data Actually Look Like?
Library Data Management
Whose Job Is Library Data Integration?

❖ Data integration should be done by **domain experts**
  • Librarians, not IT staff (IT always understaffed)
  • Programming skills should not be a requirement
  • Good user experience is a prerequisite for adoption

❖ Example driven modelling approach

❖ Value created in the community should be reusable
Library Data Management
What Tools Do We Need?

Our Approach: An Open Source Data Management Platform
Library Data Management

How Can Data Integration Be Done?
Qualify, Link and Free Your Data: D:SWARM
Who’s behind this Project?

- Collaborative development team of SLUB Dresden and Avantgarde Labs GmbH
- Started work in June 2013
- Funded from the European Regional Development Fund (ERDF)
Our Challenge: Existing Data Formats: MAB, MARC

- "selection of keywords"
  - Relevant MAB fields are 902x, 907x, 912x, 917x, 922x.
  - These fields have subfields a, b, c, ... coded with further information (type of keyword, person, time, place, concept...)
  - From field 902x to field 922x we have to check
    - If in subfield "a" there is one of these strings (800|801|820|830|845|850|860|870|880)?
    - If so, is there one of these strings (c|g|k|p|s|t|z) in subfield "b"?
    - If so, the value in subfield "c" qualifies as a keyword
    - Keyword needs to be trimmed (which is the easiest part)
Qualify, Link and Free Your Data: D:SWARM
Our Challenge: Existing Tools: Talend
Qualify, Link and Free Your Data: D:SWARM

Our Challenge: Existing Tools: Open Refine
Qualify, Link and Free Your Data: D:SWARM
What Is D:SWARM?

❖ Graphical web based **ETL modelling tool** that serves to:
  • import data from heterogeneous sources with different formats
  • map input to output schemata and design transformation workflows
  • load transformed data into property graph database

❖ With additional functionalities:
  • Exporting of data models as RDF
  • Sharing mappings and transformation workflows
Qualify, Link and Free Your Data: D:SWARM

How Does D:SWARM Work?

- Modelling GUI and job repository
- Execution environment
  - Operational data from heterogeneous data sources (ILS, OAI-PMH, CSV ...) get processed according to the transformation logics defined in modelling GUI
- Admin centre
  - Scheduling & execution planning
  - Monitoring of system (data ingest, processing, errors)
Why a Property Graph?

- Node (S) – Edge (P) – Node (O)
- Extension of RDF data model - each element can be endowed with additional information (key : value)
  - Version number
  - Provenance information
  - Type information
Qualify, Link and Free Your Data: D:SWARM
Intermediate Results as of November 2014

- Modelling GUI in 2nd version
  - Available file importer: XML, CSV, MABXML
  - Simple schema editor & graphic schema mapper
  - Transformation workflow designer & filter (Metafacture)
- Execution of mappings and transformations in modelling GUI
- Persistence in graph database (Neo4J)
- Exporter: Turtle, N-Quads, N3, ...

- Publication under Open Source licence (Apache 2):
  https://github.com/dswarm
Qualify, Link and Free Your Data: D:SWARM
Live Demo

http://demo.dswarm.org
Our Next Steps

- Provision of URI templates for resource matching and linking
- Scalable execution engine for production mode
- Extension of transformation function set
- Extension of importers
- Implementation of an administration centre
- Deduplication and FRBRization
- Integration of SLUBsemantics Enrichtment Service
- Implementation of sharing features
Qualify, Link and Free Your Data: D:SWARM
Your Next Steps

- Follow us on twitter.com/dswarm or www.dswarm.org or github.com/dswarm

- Try it out and get in contact with us
  - http://demo.dswarm.org
  - https://github.com/dswarm/dswarm-documentation/wiki
  - team@dswarm.org

- Help us prioritize our backlog
  - https://jira.slub-dresden.de/

- Fork us on github.com/dswarm