

Subject to Change - Ongoing automatic catalogue enrichment with harvested subject headings and codes

Dr. Madeleine Boxler

Subject Librarian, Head Language & Literatures

Marcus Zerbst

Systems Librarian

Subject Cataloguing – Quo vadis?

Frankfurt, RSC Satellite Meeting, 4.11.2016



Kantons-, Stadt- und
Universitätsbibliothek



Content

Background

- Starting point, requirements for subject indexing at ZB (Zentralbibliothek Zürich)
- Previous project «Digitaler Assistent»

FRED

- FRED = FREmdDaten-Anreicherung von Sacherschliessungsdaten
 - data import / enrichment for subject indexing terms
 - Technical process, current application
- Implications on subject indexing at ZB

Outlook: Next steps for subject indexing

Starting point: requirements for subject indexing at ZB

- Rationalization of subject indexing
so that subject librarians can devote more time to new tasks
- Better retrieval and visibility in catalogue
(in particular e-books and subject access in other languages)

Looking back: „Digitaler Assistent“ (DA)

- Automated subject indexing in practice at the ZB since 2013
- Initial project «Digitaler Assistent»
 - ... offered suitable subject terms (GND) for manual selection / approval,
 - ... generated suggestions on the basis of
 - a) GND-data from other systems (data import),
 - b) translations of terms from other indexing systems,
 - c) statistical analyses based on of table of contents,
 - ... worked via a separate web-based system (data import into Aleph during the night).
- FRED focuses on a) and performs direct data import into Aleph

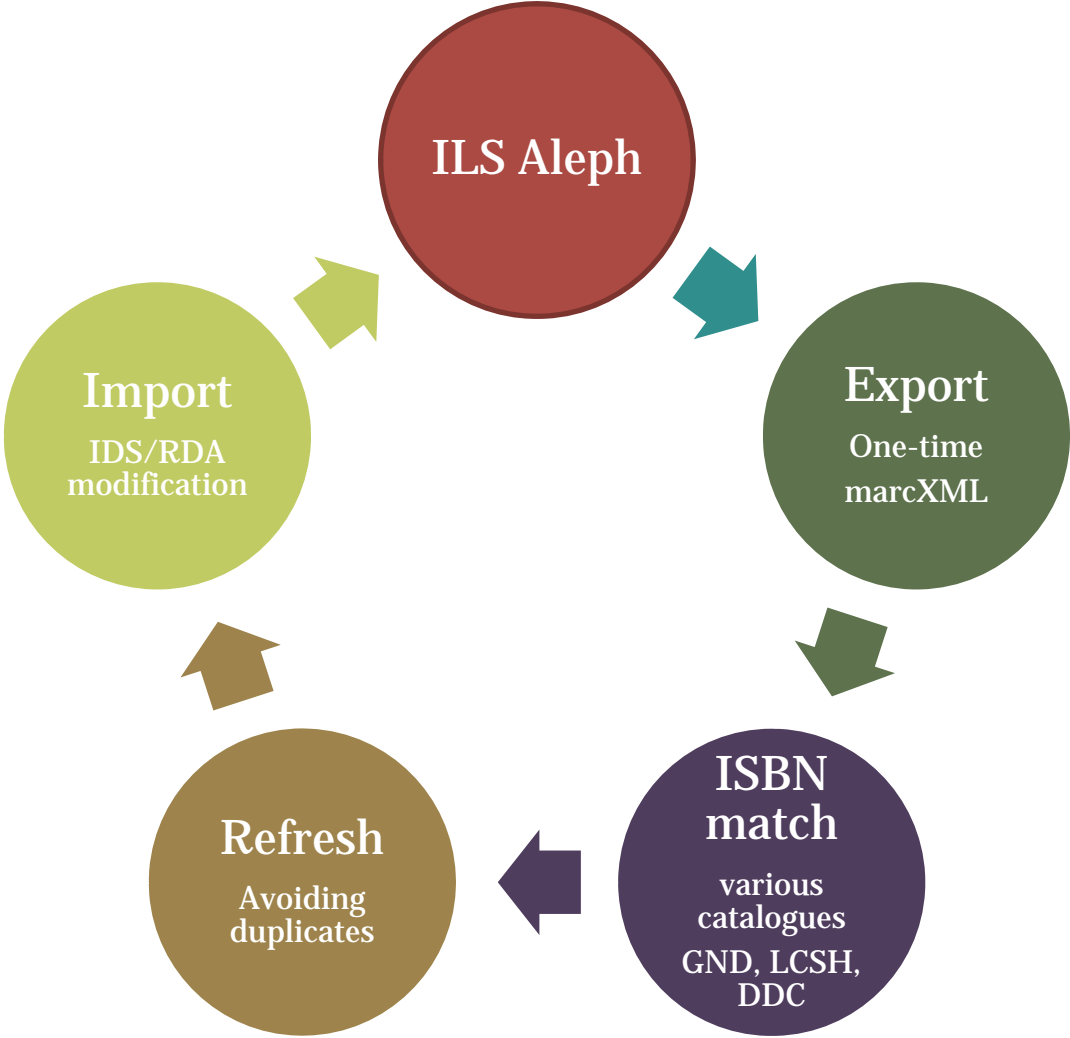
Project FRED

FREmdDaten-Anreicherung von
Sacherschliessungsdaten

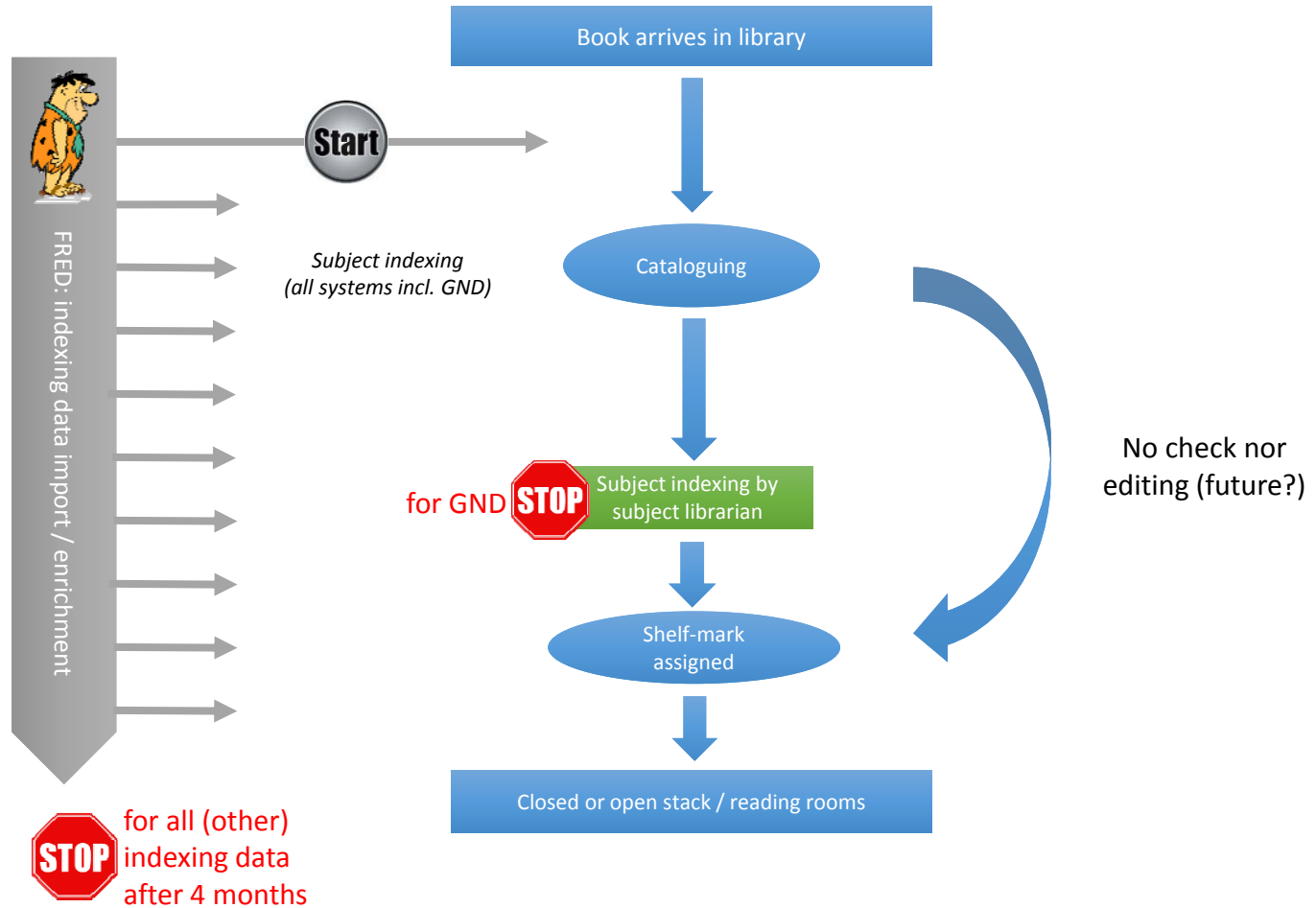
FRED in a nutshell

- Subject term data enrichment for all incoming books, including e-books
- Every day, automatically
- Matching via ISBN in other library / Verbund catalogues
- Enrichment with GND, LCSH, R ero/RAMEAU, Nuovo Soggetario, MeSH; and also DDC and RVK
- Co-operation with University Library Basel and Eurospider (commercial partner)
- First tests January 2016, Beta-test system March 2016, go-live September 2016

Process overview



Workflow chart



FRED Workflow

- Activation of import at item creation / order / arrival
- Ongoing enrichment in intervals until subject librarian handling or call number assignment respectively item status «available»
- Handling of subject librarian sets stop signal for GND data. Enrichment by other data continues.
- Subject indexing terms are imported directly to bibliographic library in Aleph, marked with cataloguer code.
- Option: Enrichment continues unmonitored for 4 months across all subject data types, is then stopped automatically.
- Separate start and stop for e-books

Refresh

Avoiding duplicates:

- Up-to-date complete bibliographic data is loaded to FRED database
- Comparison before transfer
 - „Are data to be transferred equal to current data? Then no import“.

Data equivalence check – what is equal?

- Major part of FRED's intelligence
- Check considers tag (e.g. 650) and type (e.g. "GND"), no indicators
- Limit on certain sub fields to match pre/RDA, e.g.:
655\$aAufsatzsammlung\$2gnd vs. \$2gnd-content
- Normalisation
 - Relevant sub fields to be identical after normalisation
 - Separated by hyphen "-", concatenated in original order
 - Lower case, standardized quotation marks
 - Standardized umlaute

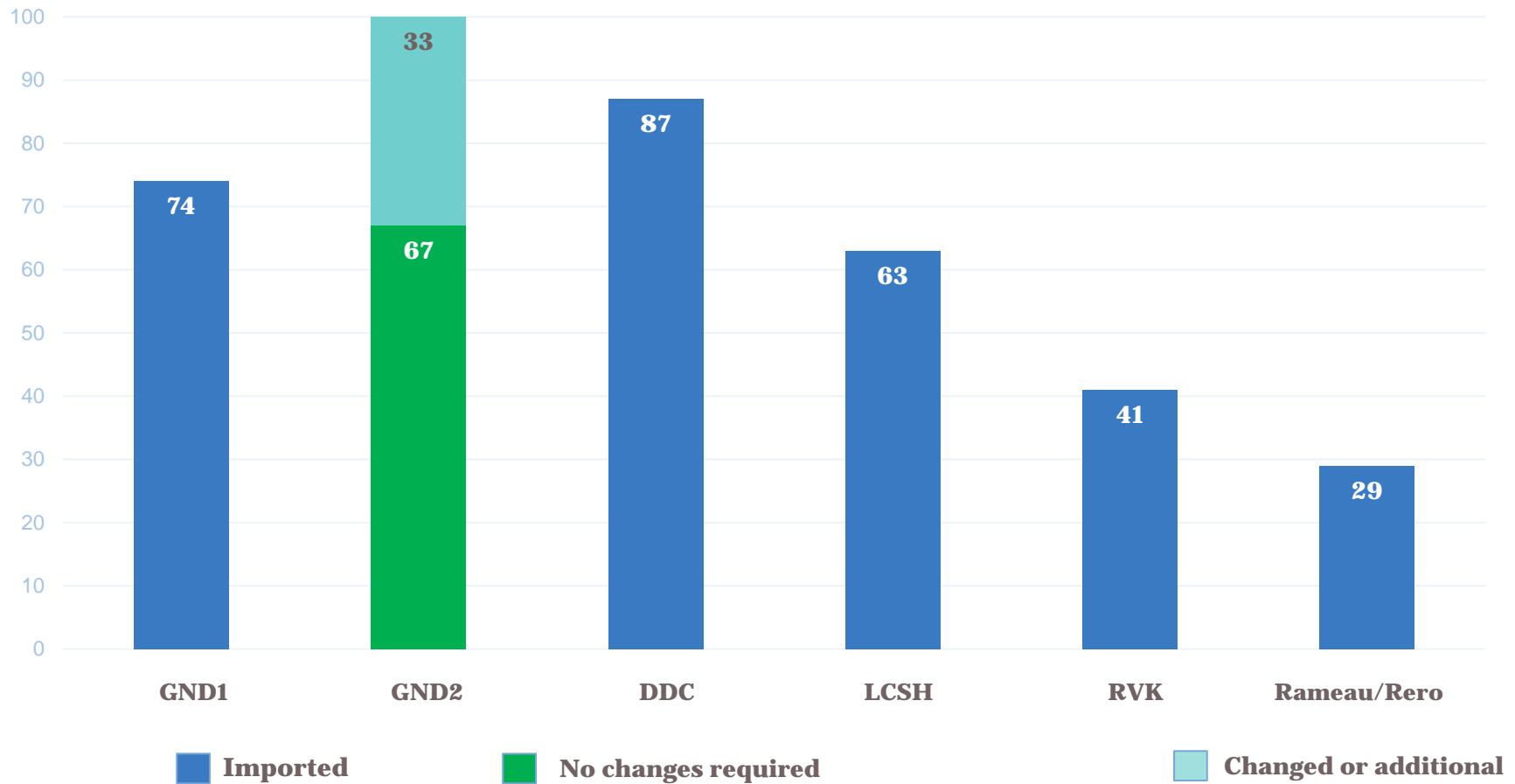
Data import process

- Import of retrieved data, possibly modified
- Script: MARC21 to IDS/RDA
 - Indicators, punctuation etc.
- Cataloger flag: CAT=ZB-FRED

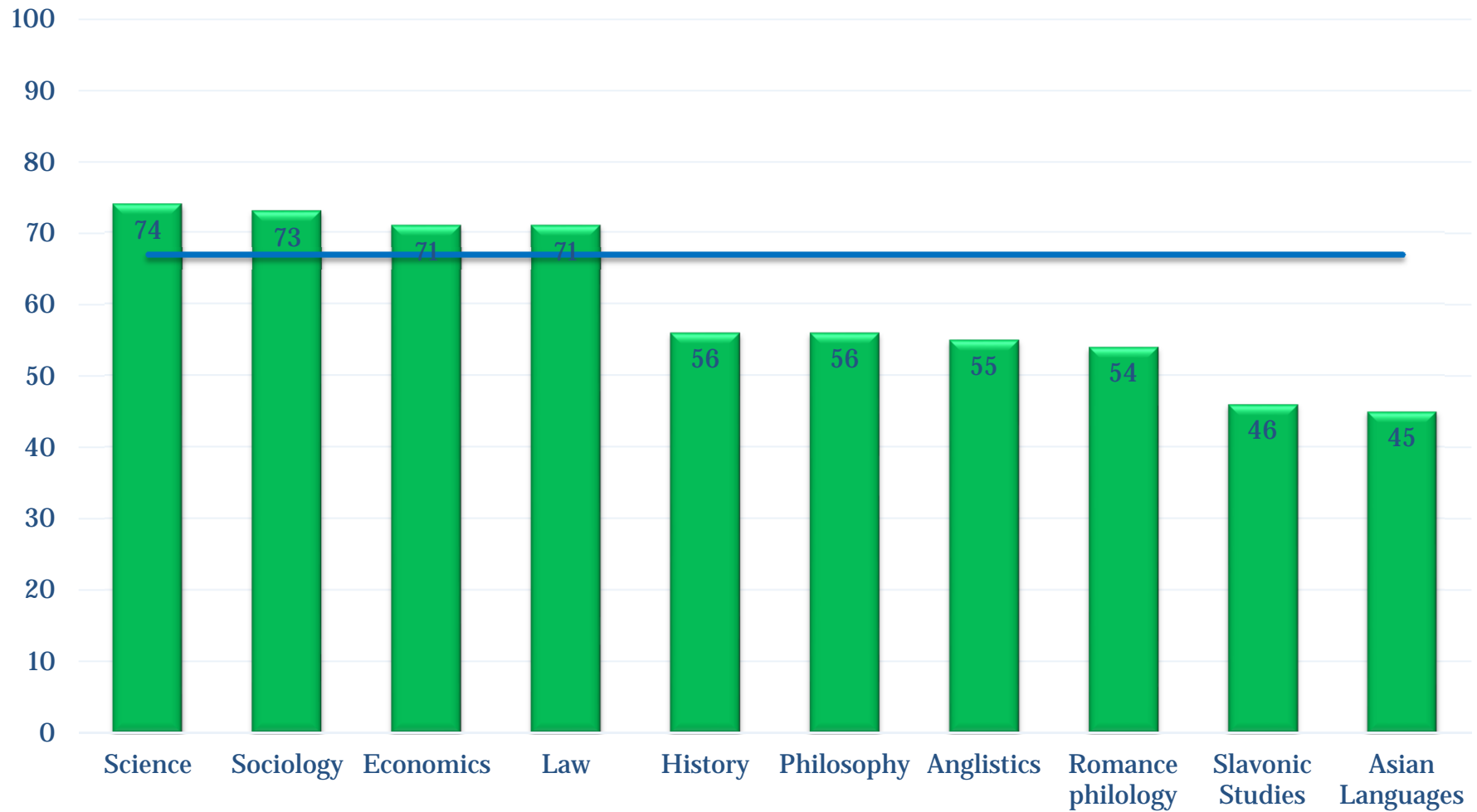
- Notification by email at field duplicates
 - Prior to import, logfile finds
- „Black list“ for enrichment stop per record
 - In case equivalence check fails (reasons partly unknown)

Print: All subject areas

Imported data, intellectual contribution of subject librarians (N=900)



Percentage of items where no changes/additions are required



Conclusion

- In addition to German subject terms FRED also enriches English and French data (varying quantity depending on record language).
- Significant results with notation codes enrichment (DDC, RVK)
- Workload of subject cataloguers is reduced because of automatic data enrichment.
- Not all subject cataloguers can equally benefit due to variations across disciplines.
- Great benefits with e-books

«Cataloguer's judgement» required

- Form subdivisions: overlapping or wrong
Chronological subdivisions: duplicate or overlapping
- Historical studies: qualifiers etc.
Example: «Griechenland» vs. «Griechenland (Altertum)»
- Philological studies: author instead of work, headings with/without qualifier «Motiv» etc.
Example: «Vögel» vs. «Vögel (Motiv)»
- Huge variations across disciplines
- *ZB has issued institutional guidelines*

Guidelines in use at ZB since Sept 2016

Only delete data which is really wrong, from the content/subject point of view

→ Example: wrong geographical term, person with wrong dates

Narrow subject terms are added manually, broad terms left standing.

→ Example : add «Kulturindustrie», leave «Kultur»

→ Example: add «Peterskirche (Rom)», leave «Rom»

Subject cataloguing - Quo vadis?

- New opportunities: SLSP – Swiss Library Service Platform
- Discovery systems improve. Implications on subject cataloguing?
- FRED currently is a good approach to reduce work load
- FRED currently is a good means for better retrieval of e-books in front ends
- FRED needs to be fed!