Let's talk about Data Quality

Karin Schmidgall (DLA Marbach)

International Conference Linked Open Data and Literary Studies

Nov 19, 2024

deutsches
literatur
archiv marbach

„So I actually found myself inside the holy of holies. It felt like being inside an enormous brain. Imagine being totally sorrounded by these shelves, full of books in their compartments, ladders all over the place, all those books stands and library tables piled high with catalogs and bibliographies, the concentrate of all knowledge, don't you know, and not one sensible book to read, only books about books"
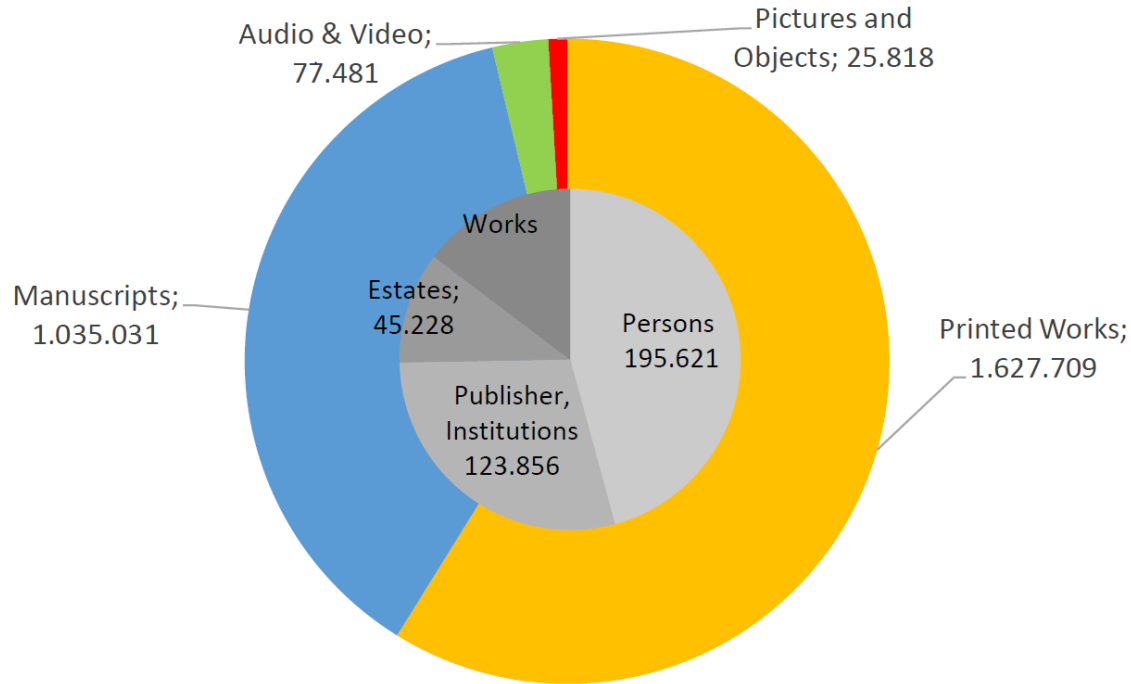
Source: DLA Marbach

Source: The man without qualities / Robert Musil. Transl. from the German by Sophie Wilkins. Bd. 1. New York: Knopf, 1996., S. 502

Catalogue statistic (Statistik 2023/10)

deutsches
literatur
archiv marbach

- Audio & Video; 77.481
- Pictures and Objects; 25.818
- Manuscripts; 1.035.031
- Works
- Estates; 45.228
- Persons 195.621
- Publisher, Institutions 123.856
- Printed Works; 1.627.709

Legend:
- Gedrucktes
- Handschriften
- Audio & Video
- Bilder & Objekte

What characterizes the DLA data?

– Density of the unique inventory
– Focus on modern German literature and literary studies.
– In the middle are authors, their works and publishers
– Depth of Cataloguing
– Variety of media
– High level of crosslinking

The DLA pursues a culture of open data and is a co-signatory of the WikiLibrary Manifesto, which aims to create and use a Linked Open Data network for art, culture and science

Publishing production before and after 1945
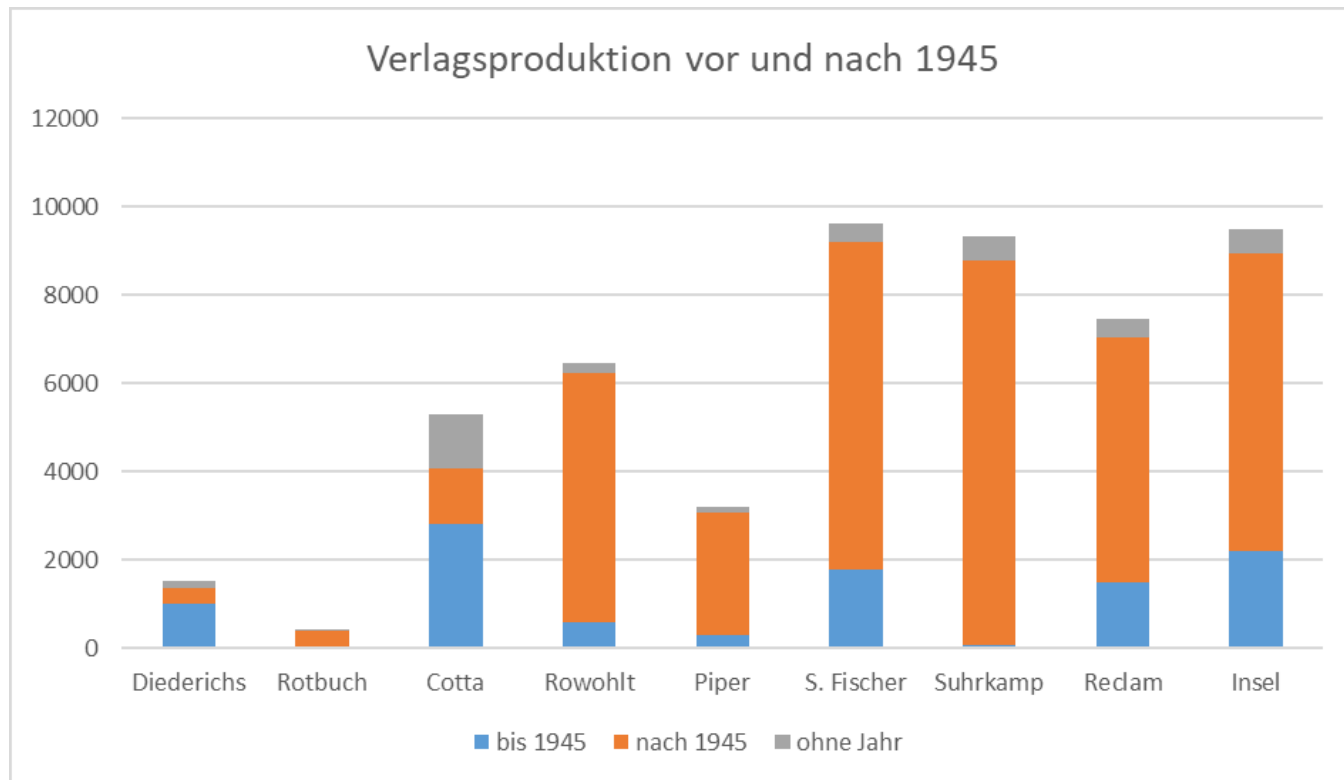
Verlagsproduktion vor und nach 1945

The visualisation based on DLA Catalogue Data. At the time of the query, a total of 52,573 data records (as of 11/2022)

Publishing production in the time of the 1st World War

Verlagsproduktion 1914 - 1918 in den Katalogdaten des DLA

Diederichs   Cotta   Rowohlt   Piper   S. Fischer   Reclam

The visualisation based on DLA Catalogue Data. At the time of the query, a total of 52,573 data records (as of 11/2022)
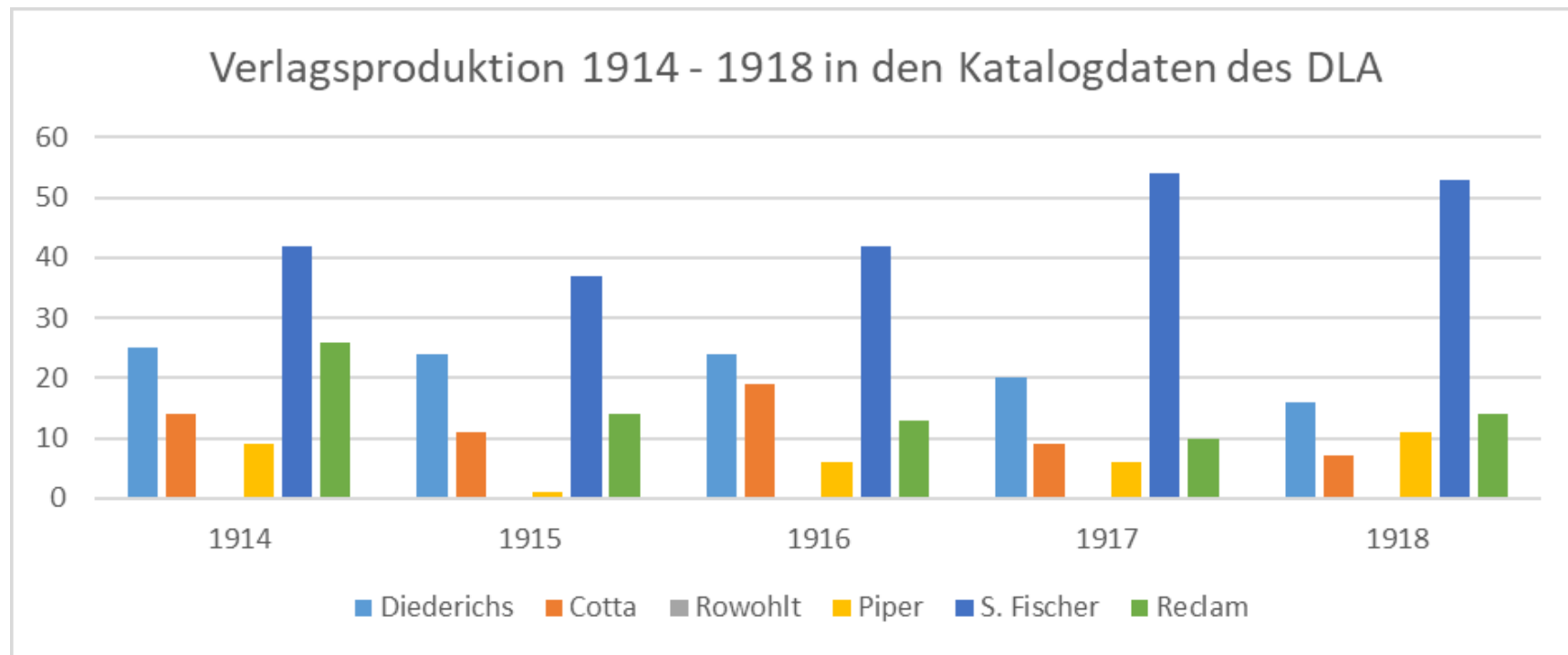
Publishing production during the National Socialist era

Verlagsproduktion 1933 - 1945 in Katalogdaten DLA

The visualisation based on DLA Catalogue Data. At the time of the query, a total of 52,573 data records (as of 11/2022)
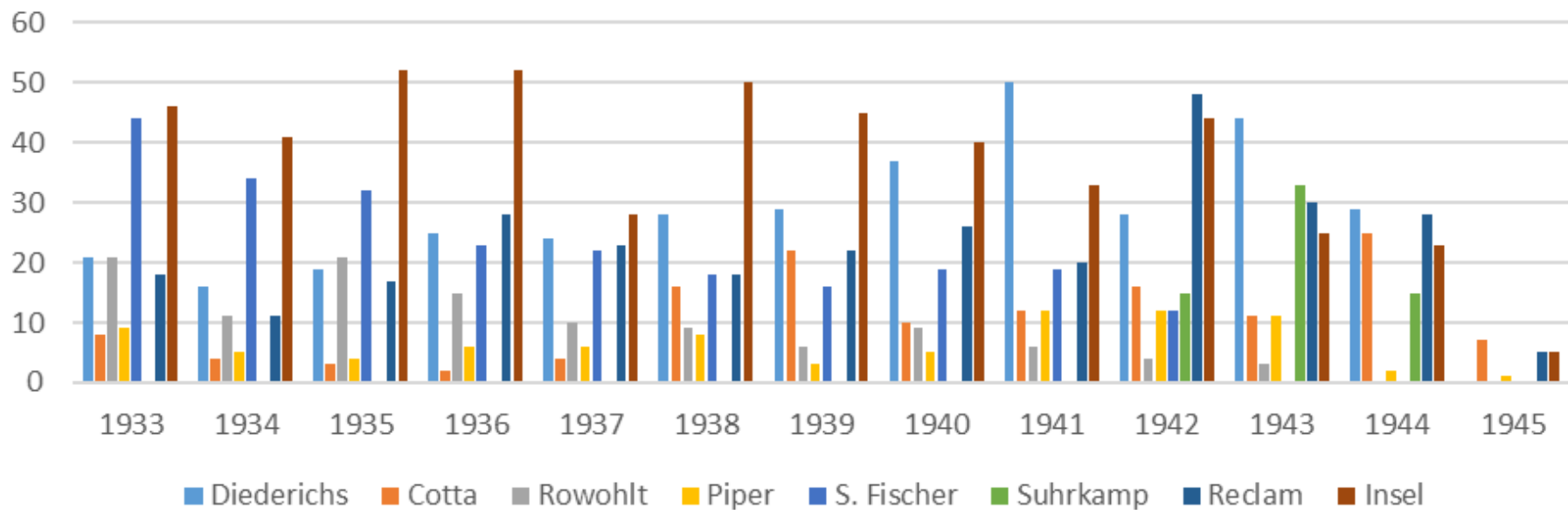
Percentage of publications by female and male authors

Total 662.540 Catalogue Records

109.310; 16%

553.229; 84%

■ Autoren ■ Autorinnen
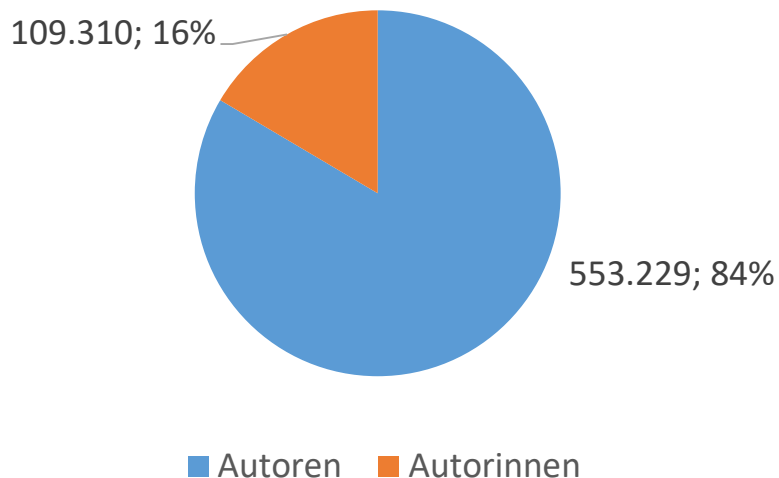
The figure shows the gender distribution in relation to the total number of catalogue records considered: 553,229 publications by 21,830 male authors and 109,310 publications by 11,580 female authors. On average, there are 25 publications per male author in the catalogue and 9 publications per female author
More informations: https://breiterkanon.hypotheses.org/3989 (urn:nbn:de:hebis:30:3-816376)

Catalogue Systems

The secret of a good librarian is that he never reads anything more of the literature in his charge than the titles and the titles of contents. ‚Anyone who lets himself go and starts reading a book is lost as a librarian.' he explained. ‚He's bound to lose perspective.'

‚So' I said, trying to catch my breath. ‚you never read a single book?

‚Never. Only the catalogs'

‚But aren't you a Ph.D.?'

‚Certainly I am. I teach at the university, as a special lecturer in Library Science. Library Science is a special field leading to a degree you know.' he explained. How many systems do you suppose there are, General, for the arrangement and preservation of books, cataloging of titles, correcting misprints and misinformation on title pages and the like'

‚I must admit that when he left me there alone, after that, I felt like doing one of two things: bursting into tears, or lighting a cigarette – neither auf which I was allowed to do there'

Source: The man without qualities / Robert Musil. Transl. from the German by Sophie Wilkins. Bd. 1. New York: Knopf, 1996., S. 503

Transfer of catalogue cards into data records

Source: DLA Marbach

Data curation - a future task of libraries and research, together

Data curation of catalogue data is a permanent task for libraries, which serves the findability, comprehensibility and usability of data records. The question of when a catalogue entry is qualitatively sufficient must be asked again and again. The answer may vary depending on the project. Archival and library standards should open up in order to be able to react to impulses and research interests and to generate new views of the data. Specifically:

- Retrospectively revise and enrich catalogue data, especially old data, using manual or semi-machine matching methods

- Continuously develop rules and regulations for catalogues in line with research needs

- Feed research results (on collections and individual objects, findings from data analyses) back into the catalogue data

deutsches
literatur
archiv marbach

Questions are gladly answered by:
Karin.Schmidgall@dla-marbach.de