

National bibliographic data for studies of social sciences and humanities: towards interoperability

Peter Aspeslagh

Centre for R&D Monitoring (ECOOM), University of Antwerp

Amsterdam/Online, June 30, 2020

Towards interoperability

- “ National bibliographic databases: opportunities for sciences studies/research evaluation when integrating multiple data sources
- “ Persistent identifiers, but lack of interoperability
- “ Example: Academic Book Publisher Register (ABP)

Interoperability

- “ **Definition:** Ability of two or more systems to use exchanged information
- “ **Key:** persistent identifiers (ISSN, ISBN, DOI, ORCID, GRID, ROR, ...)
- “ **Challenges:**
 - “ Not all entities can be associated with a particular identifier
 - “ Context-dependent; not all entities can be translated into one single identifier

Academic Book Publisher Register (ABP)

- “ International register of academic book publishers respecting the highest academic standards (cf. DOAJ)
- “ **Goal:** Supporting research evaluation, but also auxiliary database for quantitative studies of book publishing
- “ **Dependencies:** Based on data from national registers of publishers (national bibliographic databases); integrating in deduplicated master list of publishers
- “ **Issues:** delineation & connecting lists

ABP: delineation – what is a publisher?

First published 2020
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2020 Gabe Ignatow

- “ Published by [name]
- “ [name] is imprint of [other name]
- “ [other name] is part of [yet another name]

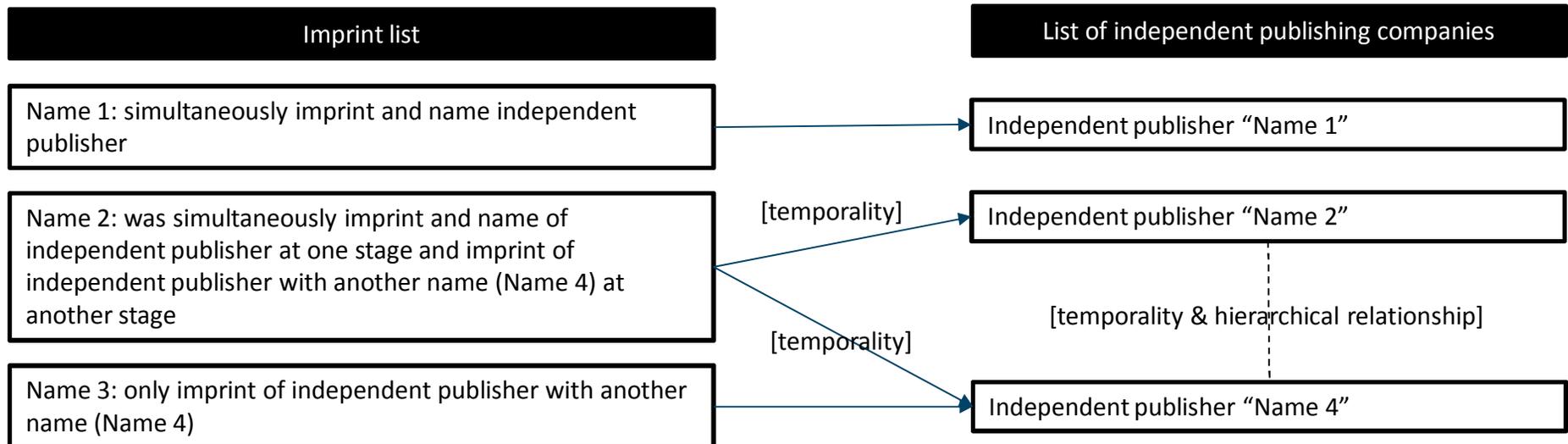
ABP: delineation – what is a publisher?

- ” **Conceptual:** multiple meanings/terms
 - ” “Publisher”: “companies/persons that prepare[s] and issues books, journals, or music for sale”
 - ” “Imprint”:
 - ” “[a] printer’s or publisher’s name, address and other details in a book or other publication”
 - ” “[a] brand name under which books are published, typically the name of a former publishing house that is now part of a larger group”
- ” Register of imprints? Register of publishers?
- ” **Temporality:** what if publisher becomes an imprint?

ABP: delineation – what is a publisher?

Possible solution: two distinct but connected lists

- “ Imprint list
- “ List of independent publishing companies



Dynamic model:

- “ Provide list of “publisher” names that are commonplace in scholarly practice [imprint list], but with different specifications
- “ Keep track of complex relations between imprints and publishing houses via parallel list of independent publishing companies
- “ Represent temporal and hierarchical changes (mergers, acquisitions, discontinuation)
- “ Translation in data model: addition of status variable to master list of publishers

ABP: connecting datasets

- “ ABP: connecting local publisher lists to master list
- “ **Particular situation:**
 - “ Local lists have diverse structures with varying set of variables
 - “ No common keys
- Customized approach necessary
- “ **Double strategy:**
 - “ Connection by text matching
 - “ Cross-validation
- “ **Procedure:**
 1. Exploration of structure of local lists
 2. Assigning internal ABP identifiers to local lists
 3. Name matching
 4. Cross-validation (ISBN-prefixes, country of headquarters, ...)
 5. Integration of matched local lists with master list items

Discussion

- “ Interoperability as challenge when using national bibliographic databases for SSH studies
- “ Emerging infrastructure cannot tackle all challenges
- “ Implications for design and comprehensiveness of current and future databases



Thank you

peter.aspeslagh@uantwerpen.be