Collecting author affiliation data for Flemish non-Web of Science SSH publications: process, results and lessons learned.

PETER ASPESLAGH // ECOOM-UNIVERSITY OF ANTWERP

Context.

In 2019, a parameter measuring international collaboration was added to the Flemish performance-based research funding system. Its implementation required author affiliation data for all co-authored publications counted in the funding system. However, contrary to the Web of Science (WoS), affiliations were not available in the Flemish Academic Bibliography for the Social Sciences and humanities (VABB), thus requiring a multifaceted data retrieval approach for 23,079 records. We zoom in on the data collection process, provide insight in international academic collaboration for Flemish non-Web of Science SSH publications, finishing with a look at an extended organization database generated by the project. All aspects teach important lessons.

Data collection process.

DATASET & DATA REGISTRATION.

>> 23,079 non-WoS publications with at least one co-author (2011-2020) >> Data registered using online platform; coded on publication level >> Unique identifier assigned to each affiliated organization (ROR ID, if available); other metadata (DOI, abstract, ORCID, ...) added or corrected

DATA RETRIEVAL STEPS.



Figure 1: Situating the subset targeted for author affiliation data retrieval.

Is affiliation data included in non-Web of Science citation databases (Crossref, ...)?



>> Less than 20% of the publications available in Crossref or Scopus >> Rest: manual data retrieval, applying a step-by-step approach >> Over 95% of the publications found, over 92% contained affiliations

COMPLEX PROCESS HIGHLIGHTS MISSING LINKS AS BARRIERS FOR INSIGHTS.

>> Affiliation data unnecessarily locked **behind paywalls** >> Google Books randomly omits elementary metadata >> Publication of affiliations lacking on websites of journals/publishers >> Transfer of metadata to e.g. Crossref far from universal practice

New perspective on international collaboration in Flemish SSH.

PERCENTAGE OF INTERNATIONAL COLLABORATION.

>> Affiliations: **21,279 VABB** versus **37,859 WoS** publications >> International collaboration: 48.1% VABB and 60.0% WoS publications

BY DISCIPLINE.

>> In general, **WoS** publications have **a higher percentage of** international collaboration than VABB publications >> Selected number of SSH disciplines: share higher in VABB subset

RELEVANCE.

>> Co-authored SSH publishing is relevant beyond a single country, regardless of WoS indexation

>> Future research: analysis on organization level



Figure 2: Step-by-step approach for author affiliation data retrieval.



Figure 3: Percentage of international collaboration per discipline and database of origin compared to the total number of publications in the dataset (Web of Science + VABB).

Addresses	Labels		Categories				Keywords		
Chalmers University of Teo	chnology			Copy Ex	ID Copy S	ys ID	Add keyword		
General data	Contact data		ROR Data & E	COOM	Enriched dat	a			
Name: Chalmers University of Technology Extra information: xternal database: ror xternal database ID: https://ror.org/040wg7k59 ocation on Google Maps	[Contact 1] [Main] [Edit] Municipality: Gothenburg [Geo [Google Maps] Country: SE	names]	General Established: Type [Overview]:	1829 Type Educati	on				
DOB Deletions & ECOOM Deletions		Quantan	Labels [Overview]	Туре	Language	Label		Source	
		Overview	[+]:	alias	NA	Chalmers		ROR	
Child(ren) of Chalmers University of Technology	Fro	m To		label	fin	Chalmersin teknillinen korkeakoulu		ROR	
Wallenberg Wood Science Center (organisation)				label	swe	Chalmers tekn	iska högskola	ROR	
			Links	Title		Li	nk		
Related publications due to affiliations			[Overview]:	Chalme	rs University of Te	echnology ht	tp://www.chalmers.se/en/Page	es/default.aspx	
2 publication(s)			External identi	fiers [Ov	erview] [+]				
Timeline		Overview	Source			Ide	ntifier		
Date Event			CRIS Universiteit A	ntwerpen		XA	5231		
1937 [OrgReg Founded] Established in 1829 and until 1937 was private institution, when it became a state owned university.			EU Participant Identification Code (PIC) FundRef				999980373 501100006552		

Development of extended organization database.

NEED FOR EXTENSION.

>> ROR: no identifier for all affiliated organizations (unavailable for more than one third (34.7%, n=1,967) of the 5,671 different organizations) >> Creation of a system to **extend a local copy of ROR with new** organizations, using internal identifiers and coded according to ROR data scheme; extra modules available (timeline, keywords).

INSTRUMENT SUI GENERIS.

>> Organization database can also serve other purposes. ROR is common instrument, but lacks data about lesser known organizations. An extended database fills a void: unique identifiers for missing organizations and, if relevant, enrich ROR with local data for international use. >> Partners can provide additional & updated data; arsenal of external identifiers is being extended, both by manual addition and by retrieving relevant identifiers (cf. EU PIC code) from other sources (cf. Wikidata).



Figure 4: Extended organization database: components.

Key points.

>> Large author affiliation data collection operation for non-Web of Science publications requires multifaceted action. The operation highlights several missing links in metadata availability.

>> International collaboration in Flemish SSH publications is relevant beyond indexation in the Web of Science.

Extended organization database, created as tool for the data collection process, evolved to instrument sui generis.

Contact.

>> E-mail: peter.aspeslagh@uantwerpen.be >> ORCID: 0000-0002-2753-3144