

Proof of Concept of a European database for Social Sciences and Humanities publications: the VIRTAs-ENRESSH pilot

ENRESSH - Working Group 3

Presented by:

Janne Pölönen

Raf Guns

Joshua Eykens

What will follow

- 1. Introduction of the VIRT-ENRESSH POC:** on the origin and use of VIRT-ENRESSH in Finland, the different features of the system, and the potential at the European level
- 2. Implementation of the VIRT-ENRESSH POC:** presenting the different institutions that took part, the data formats and requirements, and the different publication types
- 3. Description of the data and the results obtained during the STSM**
- 4. Potential improvements:** differences in classification schemes and terminology, technical issues, language classification, etc.
- 5. Potential as science policy supporting tool:** international benchmarking and monitoring, research potential into publication patterns, etc.
- 6. Questions and discussion**

1. What is the VIRT-ENRESSH POC?

- First steps towards integrated bibliographic database at the European level
- Collaborative effort of ENRESSH working group 3
- Why?: Commercial bibliographic databases suffer from severe lack of coverage
- Comprehensive database for social sciences and humanities research output is a crucial component of research evaluation and/or funding distribution
- Great variety of institutional and national publication information systems → Challenge to integration

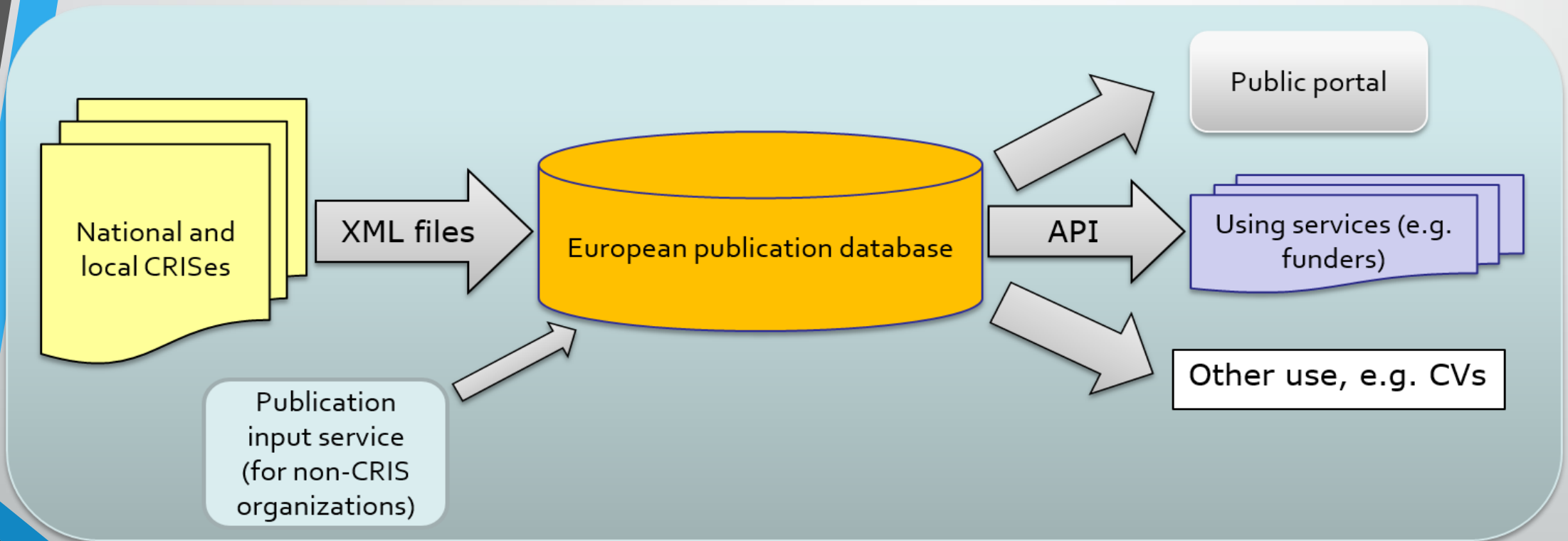
1. VIRT A in Finland

- The Finnish Ministry of Education and Culture compiles bibliographic information annually
- Supports the performance-based research funding system
- Collection of bibliographic information, commenced in 2011
- VIRT A is new integrated solution at the national level
- 54 organizations, 14 of which are universities
- +- 60,000 publications per year
- A data warehouse/"data hub"

1. Potential at the European level

- Extension of the Finnish publication information service to other European countries and institutions
- European decentralized system: integration and visibility of data about and for the SSH and other fields of science
- Would provide complete overview on European research publications
- Including all types of research output
- Open Science Agenda (OECD, 2015)

1. Schematically



2. Implementing the VIRT-ENRESSH POC

- Initiative of ENRESSH working group 3 + CSC – IT Center for Science: ENRESSH Meeting in March, 2017
- Institutional data from 6 universities, from four European countries:
 - University of Helsinki, University of Jyväskylä, Tampere University of Technology, University of Antwerp, University of Oslo, University Carlos III Madrid (UC3M)
- Publication metadata for years 2014 – 2015
- OECD Fields of Science Classification (OECD, 2007)

2. Data format and requirements

- XML as data exchange format: Finnish VIRTAs format as starting point
- Simple CSV model was chosen for pilot
- CSC uploaded the files into VIRTAs
- Lowest common denominator for comparison of data contents
- 8 mandatory fields: Organization ID, organization-specific ID of publication, publication year, publication title, authors, publication type, fields of science of the publication, and organization of the authors
- 28 optional fields: ISBN, ISSN, journal title, open access status, ORCID, etc.

2. Publication types

- Publication type classifications differ, also between piloting institutions
- Analogous categories can be found in all countries (journal article, book/monograph, edited volume, conference proceedings, etc.)
- Validation and identification algorithms in VIRTAs are heavily dependent on publication types
- Finnish classification scheme was chosen as basis

2. Publication types

| Finland / Madrid | | Flanders 1=peer-reviewed / 0 = non peer-reviewed | | Norway |
|----------------------------|---|---|---|------------------------------|
| Peer-reviewed articles | A1 Journal article, original research | VABB-1: journal article | 1 | 3= Article in series (ISSN) |
| | A2 Review article | | | |
| | A3 Book section | VABB-4: book chapter | 1 | 2= Article in book (no ISSN) |
| | A4 Conference proceedings | VABB-5: proceedings paper | 1 | |
| Non peer-reviewed articles | B1 Non-refereed journal articles | VABB-1: journal article | 0 | |
| | B2 Book section | VABB-4: book chapter | 0 | |
| | B3 Non-refereed conference proceedings | VABB-5: proceedings paper | 0 | |
| Monographs | C1 Book | VABB-2: monograph | 1 | 1= Monograph |
| | C2 Edited book | VABB-3: edited book | 1 | |
| Professional | D1 Article in a trade journal | | | |
| | D2 Article in a professional book | | | |
| | D3 Professional conference proceedings | | | |
| | D4 Development or research report | | | |
| | D5 Textbook, professional manual or guide | | | |
| | D6 Edited professional book | | | |
| Popular | E1 Popularised article, newspaper article | | | |
| | E2 Popularised monograph | VABB-2: monograph | 0 | |
| | E3 Edited popular book | VABB-3: edited book | 0 | |

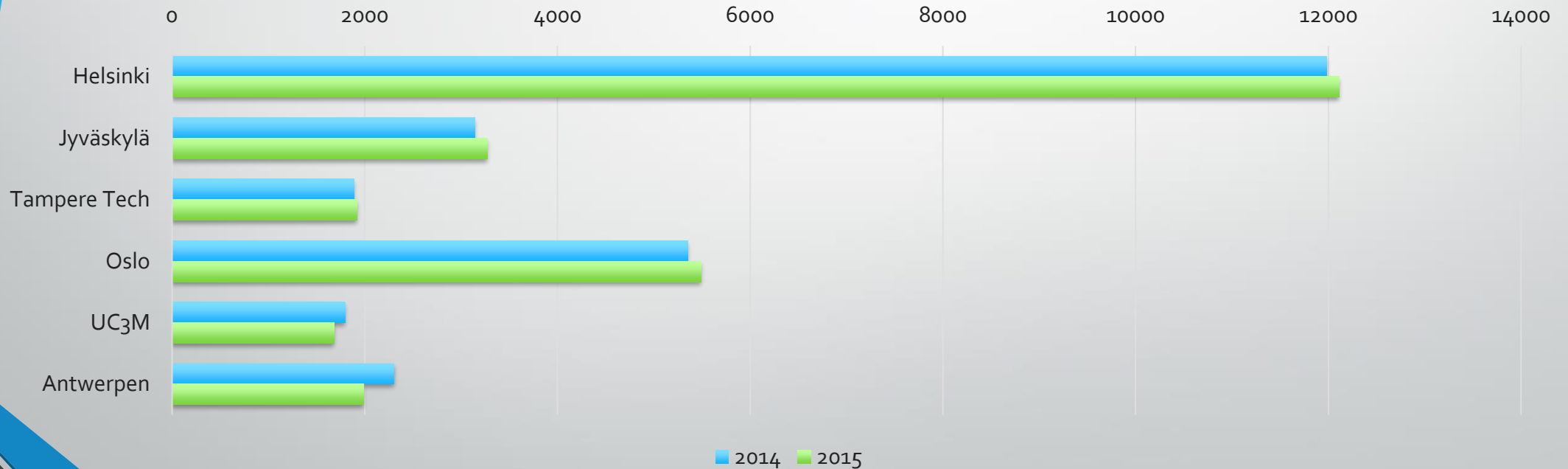
3. Description of the data

- Finnish universities and University of Oslo: all fields of science
- University of Antwerp and University Carlos III Madrid: only SSH publications
- Coverage of non-peer reviewed publications also differs between institutions
- All publication types, from all disciplines, and in all languages could be submitted
- For a preliminary analysis, peer reviewed publications (journal articles) were studied

3. Description of the data

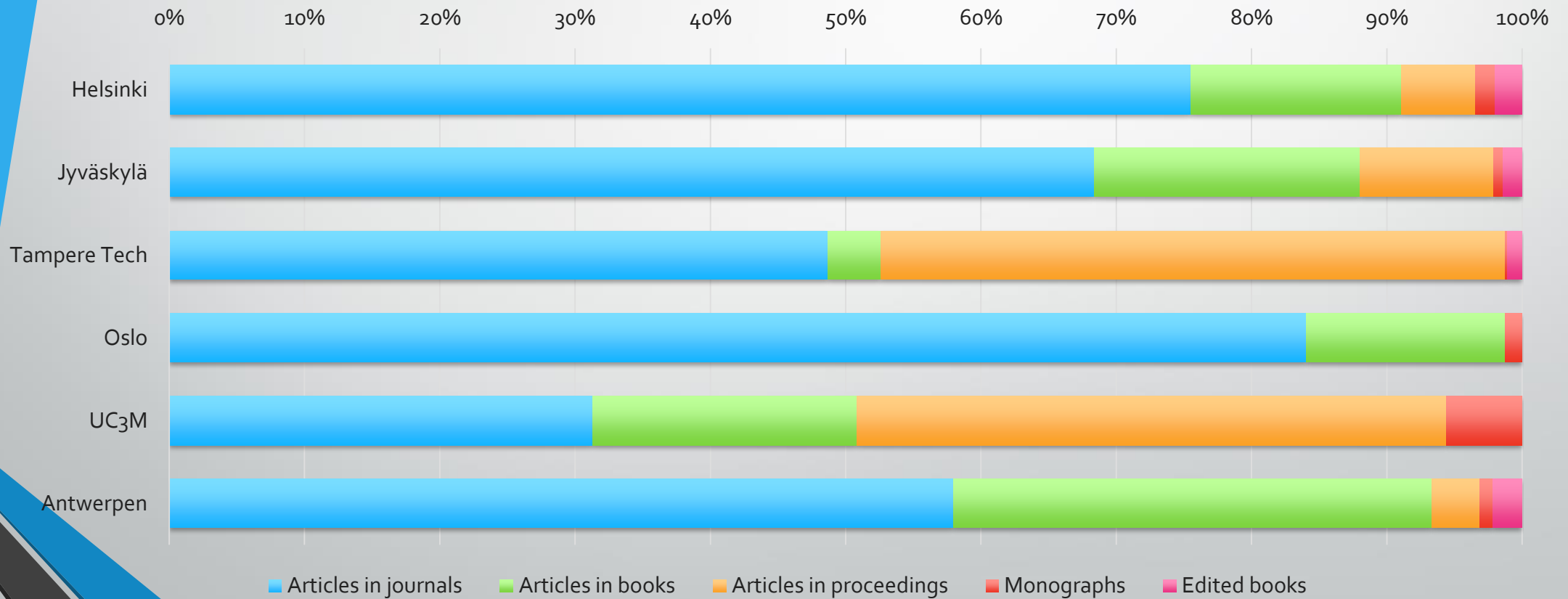
- Period 2014 – 2015: 52,948 publications of 6 research institutions

Number of publications, total



3. Description of the data

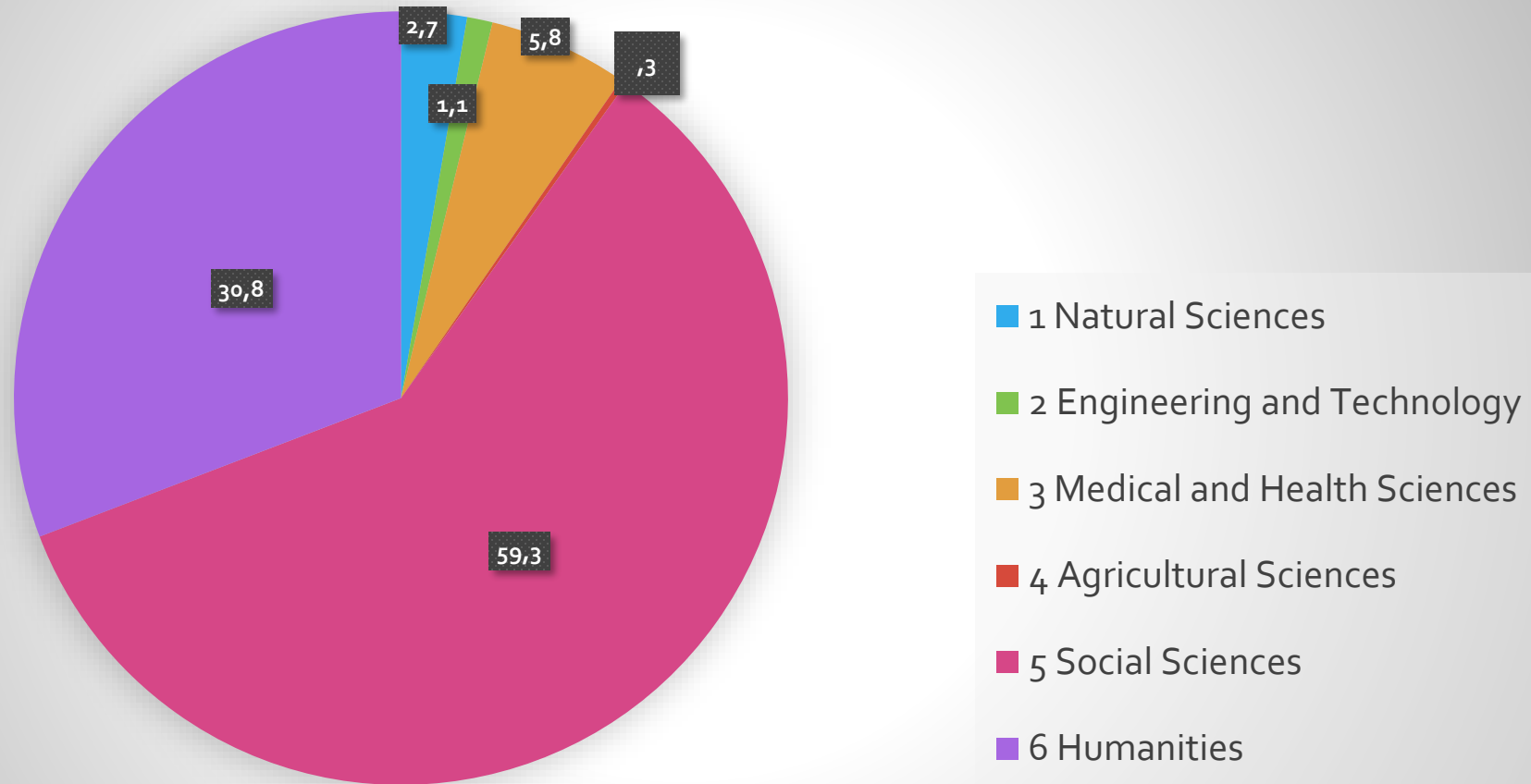
Publications by types



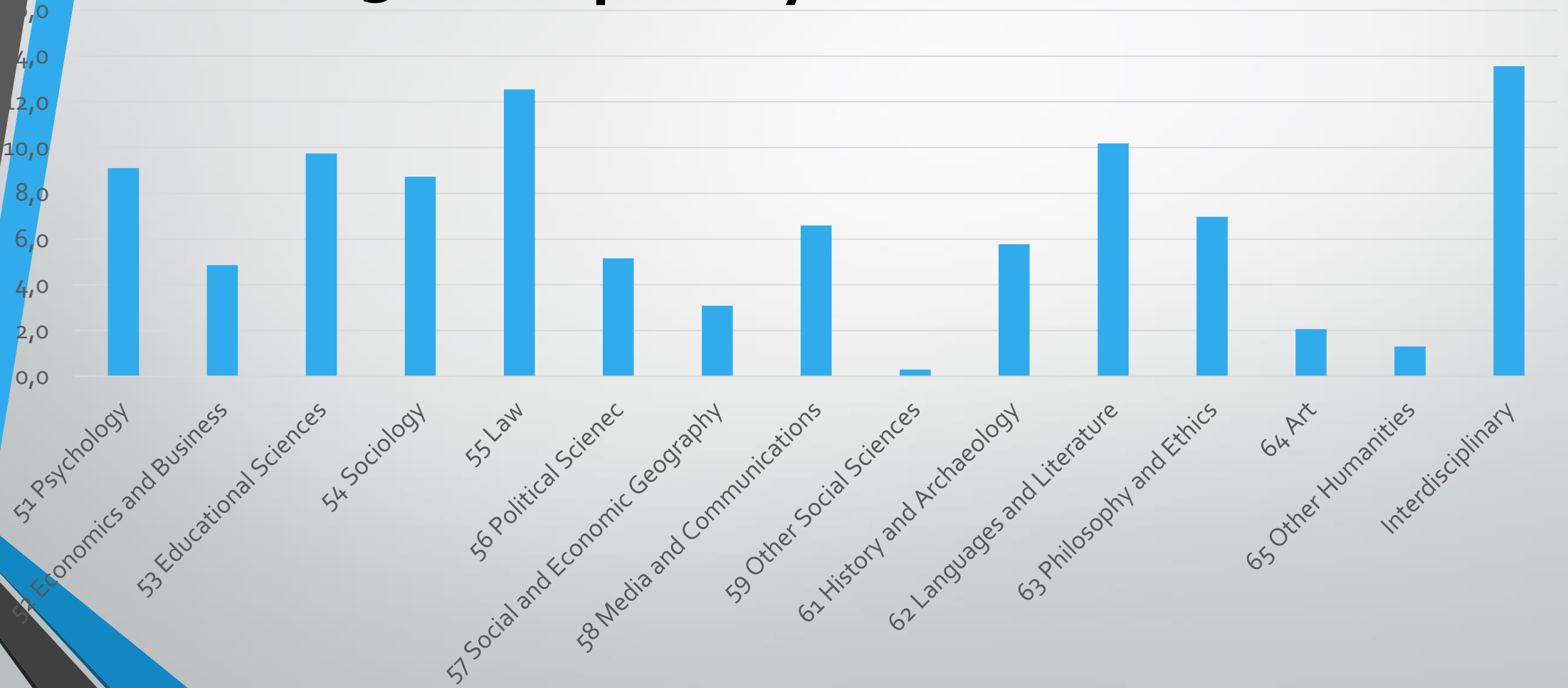
3. Results obtained during STSM

- Study of articles published in peer reviewed academic journals: SSH only
- Main aim of the analysis was to look at the shared use (co-occurrence of journals) by authors from different research institutes
- A list of 7,791 original research papers
- Additional information on the journals was added using different sources (fields of science, language)
- Articles were distributed over 3,575 different journals

3. Disciplinary fields of journals



3. Disciplinary fields of articles



3. Pathways for further analysis

- More detailed bibliometric research on all publication types
- New possibilities for sociology and history of science research
- Investigations of new modes of knowledge production (mode 2) (Nowotny et al., 2003)
- Changes in publication language
- Scattering of publications over different journals and changes over time
- Shares of book publications compared to journal articles
- Changing collaboration patterns between different research institutes

4. Potential improvements

- Finding common ground for differences in classification types and terminology (e.g. academic/scholarly, peer-review)
- Technical issues (adaptation of identification algorithms)
- Language classification
- Gradually implementing an ontology based management approach (CERIF interchange format)
- Further streamlining integration of data into VIRTAs
- Refinements and documentation of the OECD-FOS-Scheme

5. Potential as science policy supporting tool

- International benchmarking and monitoring across institutions and countries, and research into publication patterns
 - Share of journal, conference and book publications
 - Publication fields and languages
 - WoS and Scopus publications
 - Co-authorship and inter-university collaboration
 - Open access publishing and self-archiving
 - Impact factors and national ratings



Questions and Discussion