



# European Publication Information Infrastructure - metadata transfers in European context

Joonas Nikkanen - Project Manager - CSC IT Center for Science  
[orcid.org/0000-0002-5036-6444](https://orcid.org/0000-0002-5036-6444), [linkedin.com/in/joonas-nikkanen](https://linkedin.com/in/joonas-nikkanen)  
November 20<sup>th</sup> 2019 - EuroCRIS Münster





BACKGROUND PROPOSED INFRASTRUCTURE WHAT'S NEXT?

Issues in data  
comparison of publications  
&  
ENRESSH  
point-of-view

STSM  
findings  
&  
Outline of  
infrastructure



BACKGROUND

PROPOSED INFRASTRUCTURE

WHAT'S NEXT?



Albeit European countries have invested heavily in the development of national research information infrastructures in the past years, there is not yet a way to do meaningful cross-country comparisons and international benchmarking for research publications across disciplines.



## Well-structured and commensurate data

### Publication/citation index databases

e.g. Web of Science,  
Scopus, Academic

### National CRISs or databases

e.g. CRISTIN, VIRTU,  
VABB-SHW

### Aggregating harvesters

e.g. OpenAIRE, Google  
Scholar, Dimensions

**Weak coverage**  
(in terms of disciplines  
and publication types)

**Good coverage**  
(in terms of disciplines  
and publication types)

## Less structured and miscellaneous data

**International  
National**



BACKGROUND



PROPOSED INFRASTRUCTURE



WHAT'S NEXT?

## ENRESSH

European Network for Research Evaluation

in the Social Sciences and Humanities

COST Action 15137

<https://enressh.eu/>

### WG 3.

The main objective of this Working Group is to reflect upon the standardization and the interoperability of current research information systems dedicated to the SSH research outcomes.

### Task 3.

Develop common rules and procedures for building the databases.



BACKGROUND



PROPOSED INFRASTRUCTURE



WHAT'S NEXT?



VIRTA-ENRESSH

Proof-of-Concept

<https://wiki.eduuni.fi/x/X37qAg>

[https://doi.org/10.6084/m9.figshare.  
5993506.v1](https://doi.org/10.6084/m9.figshare.5993506.v1)



BACKGROUND PROPOSED INFRASTRUCTURE WHAT'S NEXT?



Especially for SSH but not excluding  
other fields

Carried out between 3/2017-3/2018

Involved partners from Belgium,  
Finland, Norway, and Spain

Founded on the efforts made at  
national level in participating countries

The technical solution builds on the  
strengths of the Finnish VIRT  
Publication Information Service

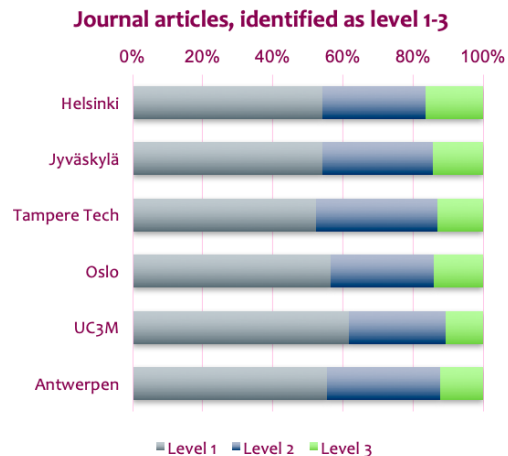
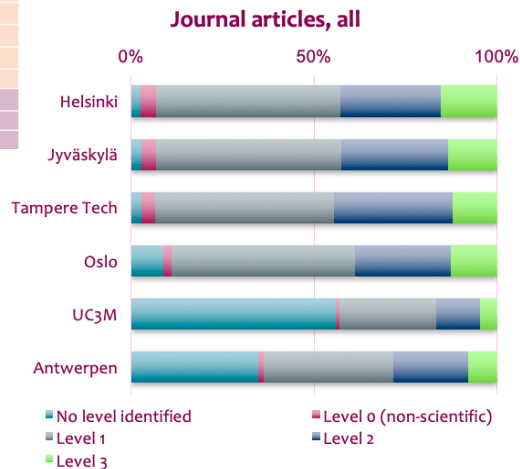
1

2

3

## BACKGROUND PROPOSED INFRASTRUCTURE WHAT'S NEXT?

Finland / Madrid		Flanders 1=peer-reviewed / 0 = non peer-reviewed	Norway
Peer-reviewed articles	A1 Journal article, original research	VABB-1: <a href="#">journal article</a>	1 3= <a href="#">Article in series</a> (ISSN)
	A2 Review article		
	A3 Book section	VABB-4: <a href="#">book chapter</a>	1 2= Article in book (no ISSN)
	A4 Conference proceedings	VABB-5: <a href="#">proceedings paper</a>	1
Non peer-reviewed articles	B1 Non-refereed journal articles	VABB-1: <a href="#">journal article</a>	0
	B2 Book section	VABB-4: <a href="#">book chapter</a>	0
	B3 Non-refereed conference proceedings	VABB-5: <a href="#">proceedings paper</a>	0
Monographs	C1 Book	VABB-2: <a href="#">monograph</a>	1 1= Monograph
	C2 Edited book	VABB-3: <a href="#">edited book</a>	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: <a href="#">monograph</a>	0
	E3 Edited popular book	VABB-3: <a href="#">edited book</a>	0





BACKGROUND



PROPOSED INFRASTRUCTURE



WHAT'S NEXT?



Issues in data comparability:

- Disciplines
- Inclusion criteria
- Semantics
- Publication types



BACKGROUND

PROPOSED INFRASTRUCTURE

WHAT'S NEXT?



## ENRESSH

Short Term Scientific Mission – June  
2019

*Develop the existing VIRT-ENRESSH data model by **taking into consideration the CERIF data model** and achieve wider interoperability by **forming a minimum set of CERIF elements** needed in publication metadata transfers*

*Investigate the potential of VIRT-ENRESSH data model as a basis of research publication metadata transfers in European context and **outline how the research publication metadata could be transferred** by using the ENRESSH-VIRT infrastructure as an example of national metadata aggregator*



BACKGROUND PROPOSED INFRASTRUCTURE WHAT'S NEXT?



## Data model

Set of CERIF elements with control  
over which attributes need to be  
included, divided into three categories:

### Mandatory

- Publication
- Internal identifier
- Publication type
- Publication title
- Publication date
- Author
  - Organizational author and affiliation
- Discipline

### Conditional

- ISSN\*
- ISBN\*
- Source title\*
- Peer review\*

### Optional

- Audience
- DOI
- Volume
- Number
- Start page
- End page



BACKGROUND PROPOSED INFRASTRUCTURE WHAT'S NEXT?

## Publication channel databases

To harmonize publication level metadata, a publication channel database (e.g. Publication Forum, Nordic list, ERIH+) and book publisher database (e.g. IRAP) could be used to resolve classification and publication type issues.

Also provides a qualitative aspect to individual publications and assists in identifying predatory journals



BACKGROUND PROPOSED INFRASTRUCTURE WHAT'S NEXT?



## Metadata transfers

Technical solution based on the  
ENRESSH-VIRTA-POC with addition  
of the use of (CERIF compatible)  
OAI-PMH endpoints

Accommodates to variety of source  
systems with varying technical  
capabilities



## ENRESSH Short Term Scientific Mission – June 2019

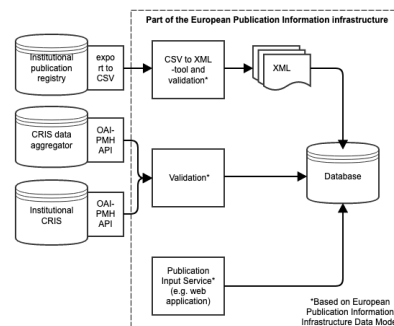
### Data model for European Publication Information Infrastructure

Following the ENRESSH-VIRTA-POC, an idea of required metadata model to be used on European level was discussed. This common standardization and data content would need to be defined to have real comparability between research outputs reported to institutional, national or even international databases. From the POC of 6 organizations and 4 countries, a certain set of classes, attributes and associations were observed that could make for a so-called "lowest common denominator" - a way to unify metadata from all sorts of source systems and thus achieve metadata that could be compared and analyzed across data from various countries in Europe. Thus the next step is to develop a data model specifically for the purpose of integrating institutional or national publication data from different countries. This needs to be done with an eye towards enhancing comprehensiveness, comparability and further use of the data. Although the data model and infrastructure should allow inclusion of all relevant scholarly outputs in different fields, it should also have enough metadata and structure to permit relevant subsets of publications to be used in comparisons and benchmarking.

As one deliverable of this STSM is the further analysis and draft of a data model for European Publication Information Infrastructure. The data model is to be as interoperable as possible, yet aiming to have as high quality metadata as possible.

Interoperability is of crucial importance when the source systems collecting metadata from research are numerous and vary heavily from institution and nation to another. The ontological approach also supports making data exchangeable with current research information standards such as EuroCRIS's CERIF data model. In an ontology-based

### Outline of the architecture for European Publication Information Infrastructure



**Results:**  
[https://wiki.eduuni.fi/x/S4t\\_Bg](https://wiki.eduuni.fi/x/S4t_Bg)

**Results:**  
[https://wiki.eduuni.fi/x/TYt\\_Bg](https://wiki.eduuni.fi/x/TYt_Bg)



BACKGROUND PROPOSED INFRASTRUCTURE WHAT'S NEXT?

### Well-structured and commensurate data

#### Publication/citation index databases

e.g. Web of Science,  
Scopus, Academic

#### European Publication Information Infrastructure

**National  
CRISs or databases**  
e.g. CRISTIN, VIRTa,  
VABB-SHW

#### Aggregating harvesters

e.g. OpenAIRE, Google  
Scholar, Dimensions

#### International National

### Less structured and miscellaneous data

**Weak coverage**  
(in terms of disciplines  
and publication types)

**Good coverage**  
(in terms of disciplines  
and publication types)

1

2

3

BACKGROUND

PROPOSED INFRASTRUCTURE

WHAT'S NEXT?

Emphasis on:

Controlled structure

Metadata quality

Inclusiveness

Comparability



BACKGROUND



PROPOSED INFRASTRUCTURE



WHAT'S NEXT?

ENRESSH coming to an end in April 2020

Work so far has been valuable and  
will continue in some shape or form

Collaboration in e.g. Nordic List development,  
NordRIS proposal for NordForsk

To continue work on the European infrastructure,  
an application for CEF-Telecom call  
was submitted last week



### Joonas Nikkanen

Project Manager  
Research Information Management and  
Interoperability  
Tel. +358 50 381 80 92  
[linkedin.com/in/joonas-nikkanen](https://www.linkedin.com/in/joonas-nikkanen)



[facebook.com/CSCfi](https://facebook.com/CSCfi)



[twitter.com/CSCfi](https://twitter.com/CSCfi)



[youtube.com/CSCfi](https://youtube.com/CSCfi)



[linkedin.com/company/csc---it-center-for-science](https://linkedin.com/company/csc---it-center-for-science)



[github.com/CSCfi](https://github.com/CSCfi)