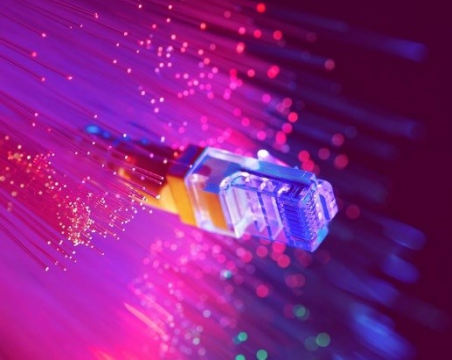




Towards the integration of European research information

Hanna-Mari Puuska, CSC – IT Center for Science Ltd., Finland
Janne Pölönen, Federation of Finnish Learned Societies
Tim Engels, University of Antwerp, Flanders
Gunnar Sivertsen, NIFU, Norway



CSC – Finnish research, education and public administration ICT knowledge centre

Contents of the presentation

1. Background: Integrated European Research Information - On the agenda, still unsolved...
2. Challenges of integration
3. Case Finland: VIRTa Publication Information Service
4. ENRESSH pilot project on applying VIRTa concept to European level

Integrated European Research Information - On the Agenda, but still unsolved...

- Science policy and research evaluation at all levels of the European Research Area need support from reliable, comparable, and comprehensive information on research activity, productivity and quality.
- Commercial databases, Web of Science or Scopus, do not provide complete coverage of research output in any field, and in SSH fields they suffer from severe lack of coverage of publications in books and languages other than English.

On the agenda, but still unsolved...

A report in 2010 to European Science Foundation and the British, Dutch, French, and German research councils

**Towards a Bibliometric Database for
the Social Sciences and Humanities –
A European Scoping Project**

A report produced for DFG, ESRC, AHRC, NWO, ANR and ESF

8 March 2010

Recommendations:

- Either negotiate expansions and/or a new database with the suppliers of *Web of Science* and/or *Scopus*
- Or integrate national and institutional documentation systems

On the Agenda, but still unsolved...



A report in 2014 to European Parliamentary Research Service



Measuring scientific performance for improved policy making

STUDY



Science and Technology Options Assessment
European Parliamentary Research Service
European Parliament
April 2014
PE 527.383



The key recommendation:

- The development of a European integrated research information system having features of a distributed infrastructure, inter-connecting the existing national research information systems, facilitated by CERIF – the European standards for research information systems.

On the Agenda, but still unsolved...



At the recent Strategic Membership Meeting of euroCRIS (November 2016), various international organisations agreed to explore the possibilities to cooperate in order to come to/create a European Research Information Infrastructure:

- EARMA (European Research Managers), Science Europe (funders), OpenAIRE, German Research Academies, JISC, ORCID and euroCRIS...

Challenges of integration

- At least 19 European Union Member States had developed or were developing a national research information system (*report to European Parliamentary Research Service, 2014*)
- The main difficulty of standardization and interoperability of research information at the European level is the variety of national systems, processes and data models.
- Many countries are facing a similar problem at national level when they compile information from research organizations using various local systems, e.g.:
 - Norway: a nationally integrated CRIS
 - Denmark: data integrated from local PURE systems to a national database
 - Flanders: data integrated from local CRISes to a regional database

Case Finland

- The Finnish Ministry of Education and Culture compiles bibliographic information on all scientific publications from 56 HEIs and other research organizations
- The organizations use **various solutions** for publication data collection: CRIS systems, self-made systems, excel files etc.
- National data collection covers ~60 000 publications per year = books, journal articles, conference papers, non-scholarly publications
- The national data are mainly used in monitoring research and as criteria of HEIs' performance-based funding model
- Statistical data are available in: www.vipunen.fi
- Bibliographic data and other metadata with links to full texts are available in: www.juuli.fi

VIRTA as a Finnish solution for integrating publication information

- Finnish organizations store a copy of publication metadata from their institutional CRISes in **VIRTA Publication information Service**
- VIRTA is a data warehouse, “a data hub”, making publication information available for other services and providing up-to-date, comparative data on publishing activity nationally and institutionally
- The information can be transferred to research funders, publication or data repositories, infrastructure services, or other services used by researchers
 - E.g. publication information automatically to research funding reports in Academy of Finland’s online service - implementation in 2017

Data contents of VIRTAs publication information service

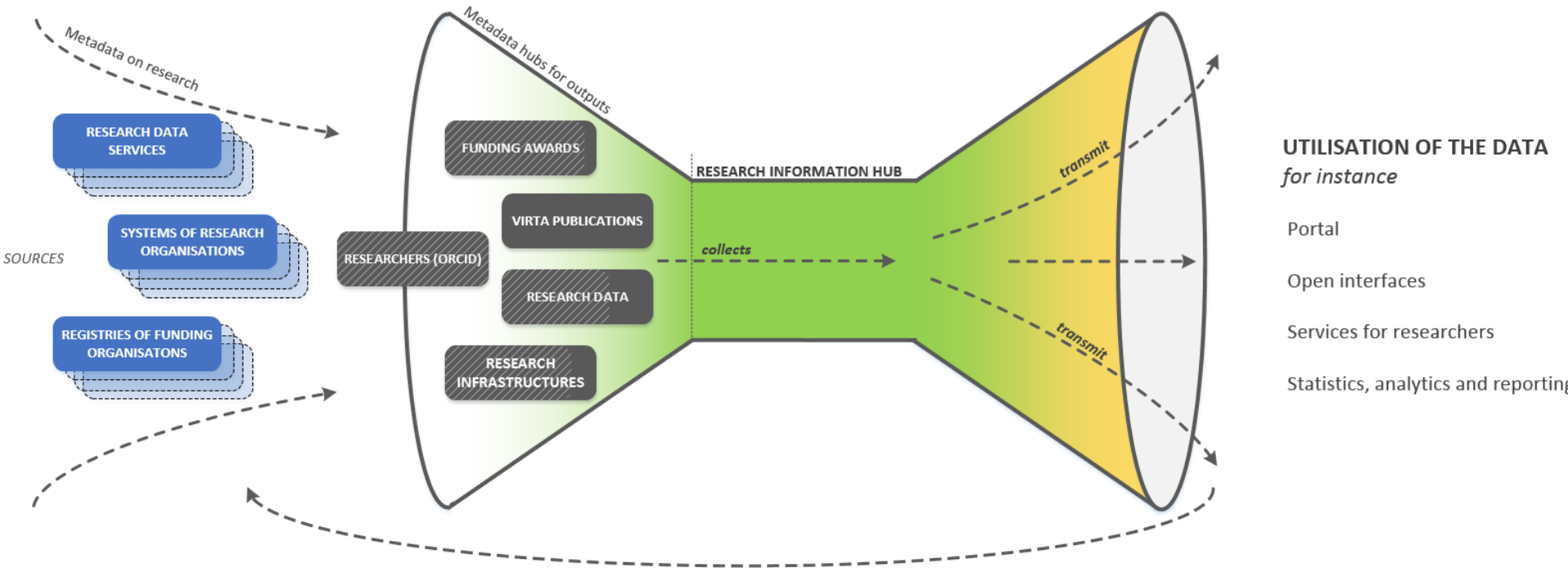
Mandatory in the Ministry's publication data collection				
Mandatory in data transfer ^a	Optional in data transfer ^b	Optional in Ministry's data collection	Other information not included in Ministry's data collection	Information generated automatically in VIRTAs
Organization ID Publication ID Publication year Title Authors Publication type Authors in organization	Number of authors Scientific field (1-6) International (yes/no) International co-publication (yes/no) Co-publication with a company Open Access ISBN* ISSN* Conference title* Publisher* *) Requirements depend on publication type	Volume Number Pages Article number Journal/series title Country Unit or department Place of publishing Host publication's title Host publication's editors Language DOI Permanent address Source ID Keywords ORCID	Project ID Funder ID	Reporting year Publication's VIRTAs ID Co-publication's VIRTAs-ID Publication Forum ID Publication Forum ranking Status of the publication

VIRTA Publication Information Service in a nutshell



Data sources	Local CRISes / publication databases of research organizations
Data format	XML files (CSV converter tool provided for small organizations)
Data exchange	From organizations via a secure and certified connection by using SFTP protocol and SSH authentication keys.
Updates	New publications and updates in local systems automatically to VIRTA e.g. once a day. The frequency depends on the organizations, minimum being once a year.
Temporal coverage	All data from previous years to present can be transferred.
Data validation	The data must include required fields and fulfil certain technical criteria. Duplicates, faults and inter-organizational co-publications identified automatically. Error reports for research organizations in online service.
Data use and availability	All metadata real time in JUULI portal: www.juuli.fi . The data are usable in various systems and services via REST API and OAI-PMH.

Towards the national research information hub 2020



ENRESSH pilot project:

Applying VIRTAcconcept to European level



- **ENRESSH** = “European Network for Research Evaluation in the Social Sciences and the Humanities” is a COST Action 2016-2020
- The ENRESSH Working Group 3
 - *“is focused on information systems, databases and repositories for publications and other outcomes of SSH research. The goals of the WG include comparison across SSH disciplines, countries and publication channels, as well as exploration of the feasibility of further streamlining (by developing shared procedures for building and maintaining databases and by designing a roadmap for a European database).”*
- ENRESSH and CSC have started a **pilot project** on examining the use of VIRTAc as a European solution
 - Adaptation of the procedure and technical solution of VIRTAc concept would establish a convenient and cost-efficient way for developing a **European Publication Information Service**

Principles for the European Research Information Service

CSC

- Common data content to be defined in collaboration with ENRESSH - “lowest common denominator” + additional optional information
- Data transfer from national or local CRISes in a standardized format (CERIF?) – annually or more frequently
- Any European country or single organization would be able to provide its data
- All information freely available in a public portal
- Via open APIs, the information usable for various purposes, e.g. funders or other existing initiatives such as OpenAIRE
- In the future, the concept is extendable to other European research output and activities, such as research data, research projects and research infrastructures

Principles for the European Research Information Service

CSC



ENRESSH pilot project: Roadmap

- **December 2016:** Proposal presented in VIRTAs steering group in the Finnish Ministry of Education and Culture
 - VIRTAs concept could potentially be extended to European level and the technical solution implemented by CSC
- **February 2017:** Proposal presented for the euroCRIS Board
- **Currently:** Discussion with potential collaborative and advisory stakeholders at European level
- **Spring/Autumn 2017:** Pilot between 4 countries
 - Finland, Flanders (Antwerpen), Norway (Oslo), Spain (Madrid)
 - Countries report their publication data from the years 2014-15 to VIRTAs
- **2018->** Implementation project ? - Funding needs to be arranged at European level



Hanna-Mari Puuska

Development Manager

hanna-mari.puuska@csc.fi

+358 50 3818 568



<https://www.facebook.com/CSCfi>



<https://twitter.com/CSCfi>



<https://www.youtube.com/c/CSCfi>



<https://www.linkedin.com/company/csc---it-center-for-science>