

F indable A ccessible I nteroperable R eusable



Putting FAIR Principles in the Context of Research Information: FAIRness for CRIS and CRIS for FAIRness

Otmane Azeroual, German Centre for Higher Education Research and Science Studies (DZHW), Germany

Joachim Schöpfel, GERiiCO Laboratory, University of Lille, France

Janne Pölönen, Federation of Finnish Learned Societies, Finland

Anastasija Nikiforova, Institute of Computer Science, University of Tartu, Estonia & European Open Science Cloud TF “FAIR metrics and data quality”

DZHW
German Centre for Higher Education Research
and Science Studies

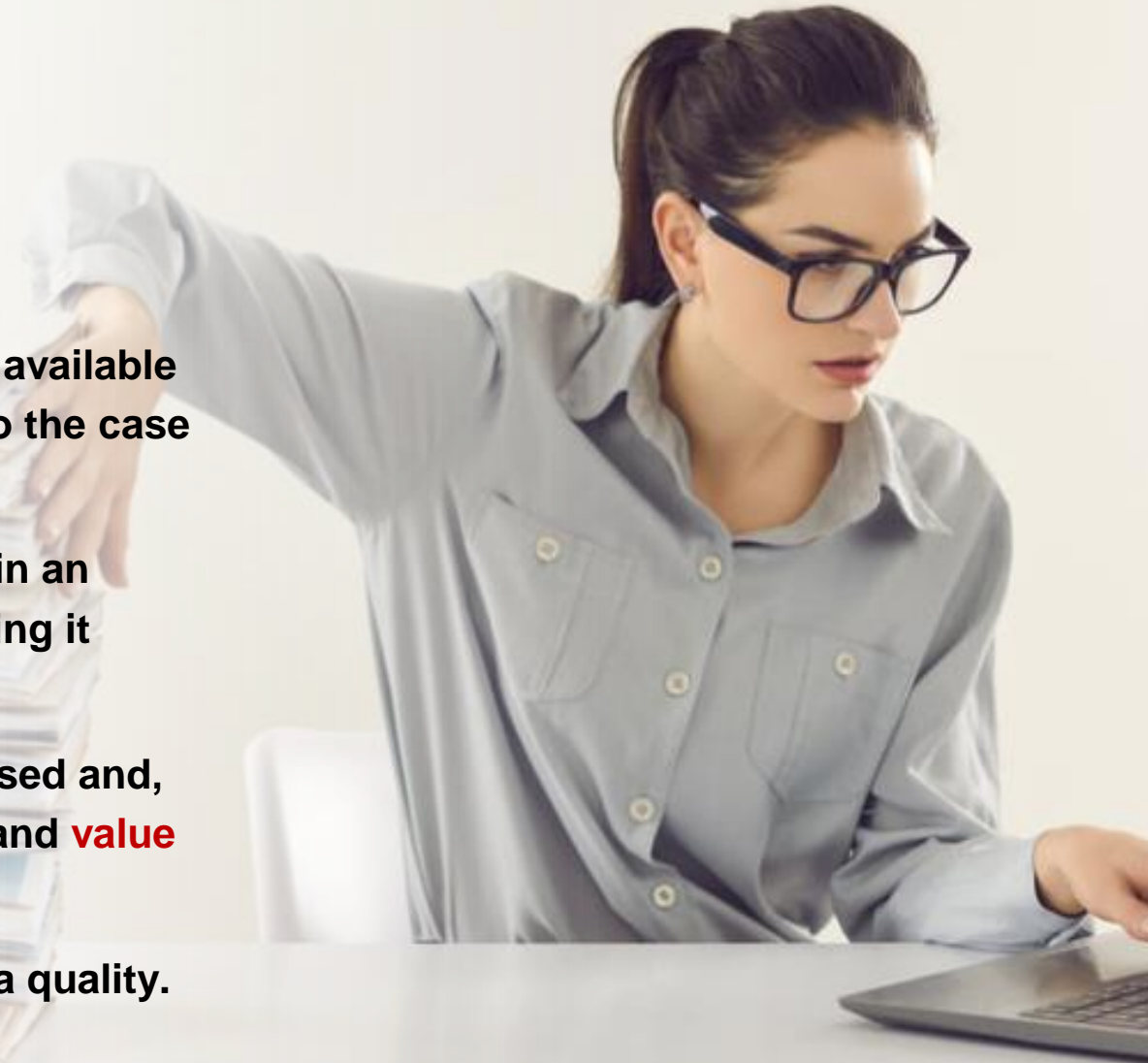
GERiiCO

Federation of Finnish
Learned Societies



BACKGROUND AND PROBLEM STATEMENT

- ✓ Today, more and more data – both produced, collected, and available from the past in paper form, are being digitized, which is also the case for the research domain.
- ✓ Digitization refers to making data available (and integrated) in an electronic and machine-readable format for further use, making it significantly more efficient.
- ✓ But, even if digitized, it is not clear whether these data are used and, more importantly, **whether the data are of sufficient quality, and value and knowledge could be extracted from them.**
- ✓ In other words, digitization does not necessarily involve data quality.
- ✓ **FAIR principles** represent a promising asset to achieve this!



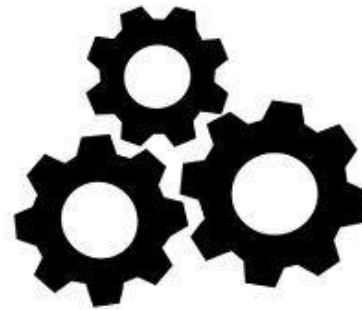
F_{indable}



A_{ccessible}



I_{nteroperable}



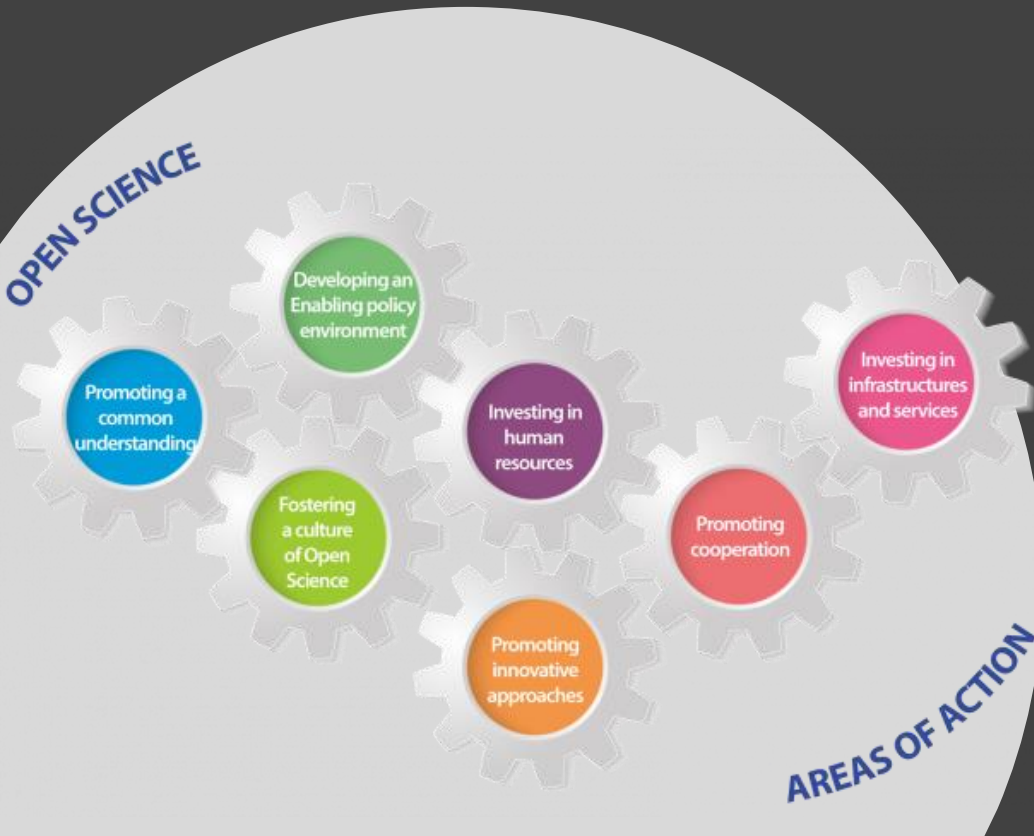
R_{eusable}





**EUROPEAN OPEN
SCIENCE CLOUD**

FAIR PRINCIPLES



- ✓ The FAIR principles were originally developed as **guidelines or recommendations** for the effective and efficient management of research data and stewardship as part of a new open science policy framework, with a “*specific emphasis on enhancing the ability of machines to automatically find and use the data*” in **data repositories** (Wilkinson et al., 2016)
- ✓ FAIR principles have become **central element in the debate and implementation of open science policies**
- ✓ Today they are **increasingly being applied to metadata, identifiers, catalogs, software, and larger infrastructures** such as European Open Science Cloud

FAIR PRINCIPLES

✓ Since their publication, they have rapidly proliferated and have become **part of (inter)national research funding programs**



EUROPEAN OPEN
SCIENCE CLOUD

✓ A special feature of the FAIR principles is the emphasis on the **legibility, readability, and understandability** of data

✓ At the same time, they pose **a prerequisite for data for their reliability, trustworthiness, and quality**

They are considered as important for research information and respective systems such as CRIS, BUT this topic is underrepresented → subject for research

FAIR Principles

Compliance



Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.

- ✓ F1. Resource is uploaded to a public repository.
- ✓ F2. Metadata are assigned a globally unique and persistent identifier.



Accessibility

Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.

- ✓ A1. Resource is accessible for download or manipulation by humans and is ideally also machine readable.
- ✓ A2. Publications and data repositories have contingency plans to assure that metadata remain accessible, even when the resource or the repository are no longer available.



Interoperability

Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.

- ✓ I1. Resource is uploaded to a repository that is interoperable with other platforms.
- ✓ I2. Repository meta- data schema maps to or implements the CG Core metadata schema.
- ✓ I3. Metadata use standard vocabularies and/or ontologies.



Reusability

Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

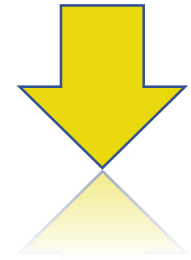
- ✓ R1. Metadata are released with a clear and accessible usage license.
- ✓ R2. Metadata about data and datasets are richly described with a plurality of accurate and relevant attributes.



HYPOTHESIS

Supporting the call for the need for a *"one-stop-shop and register-once-use-many approach"*, we argue that

CRIS is a key component of the research infrastructure landscape, directly targeted and enabled by operational application and the promotion of FAIR principles



We hypothesize that

the improvement of FAIRness is a bidirectional process, where CRIS promotes FAIRness of data and infrastructures, and FAIR principles push further improvements to the underlying CRIS

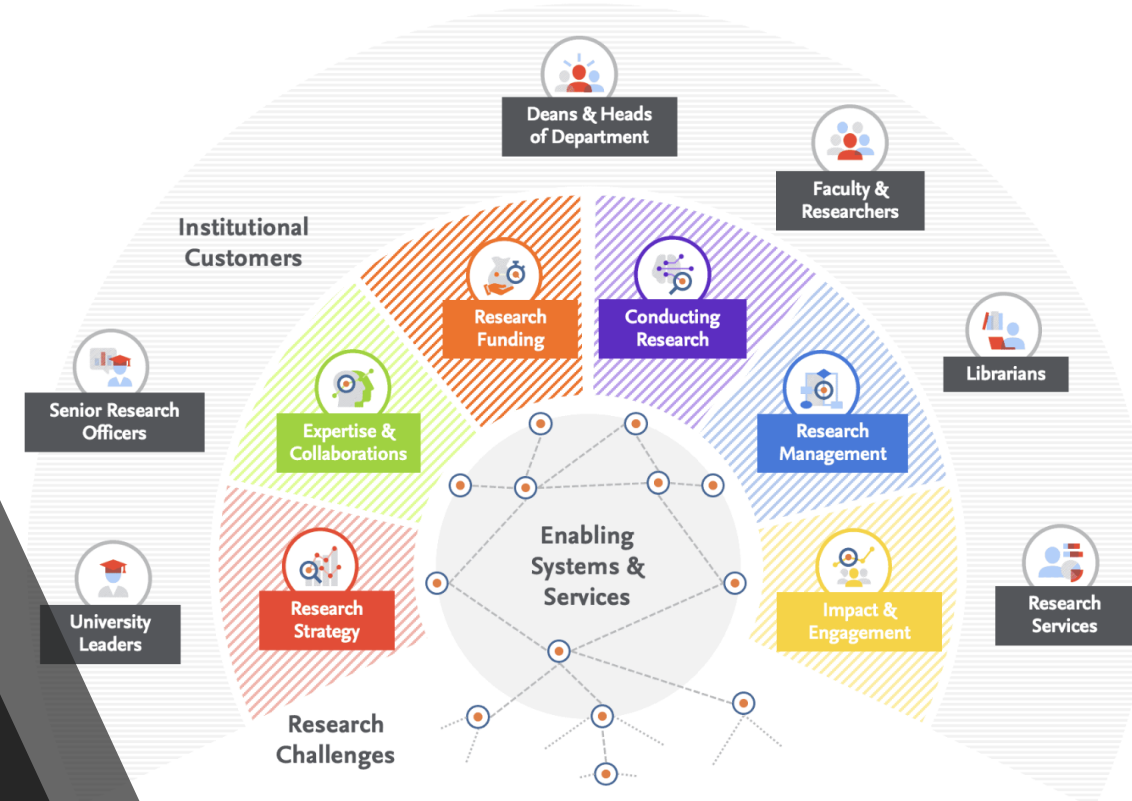
CURRENT RESEARCH INFORMATION SYSTEMS

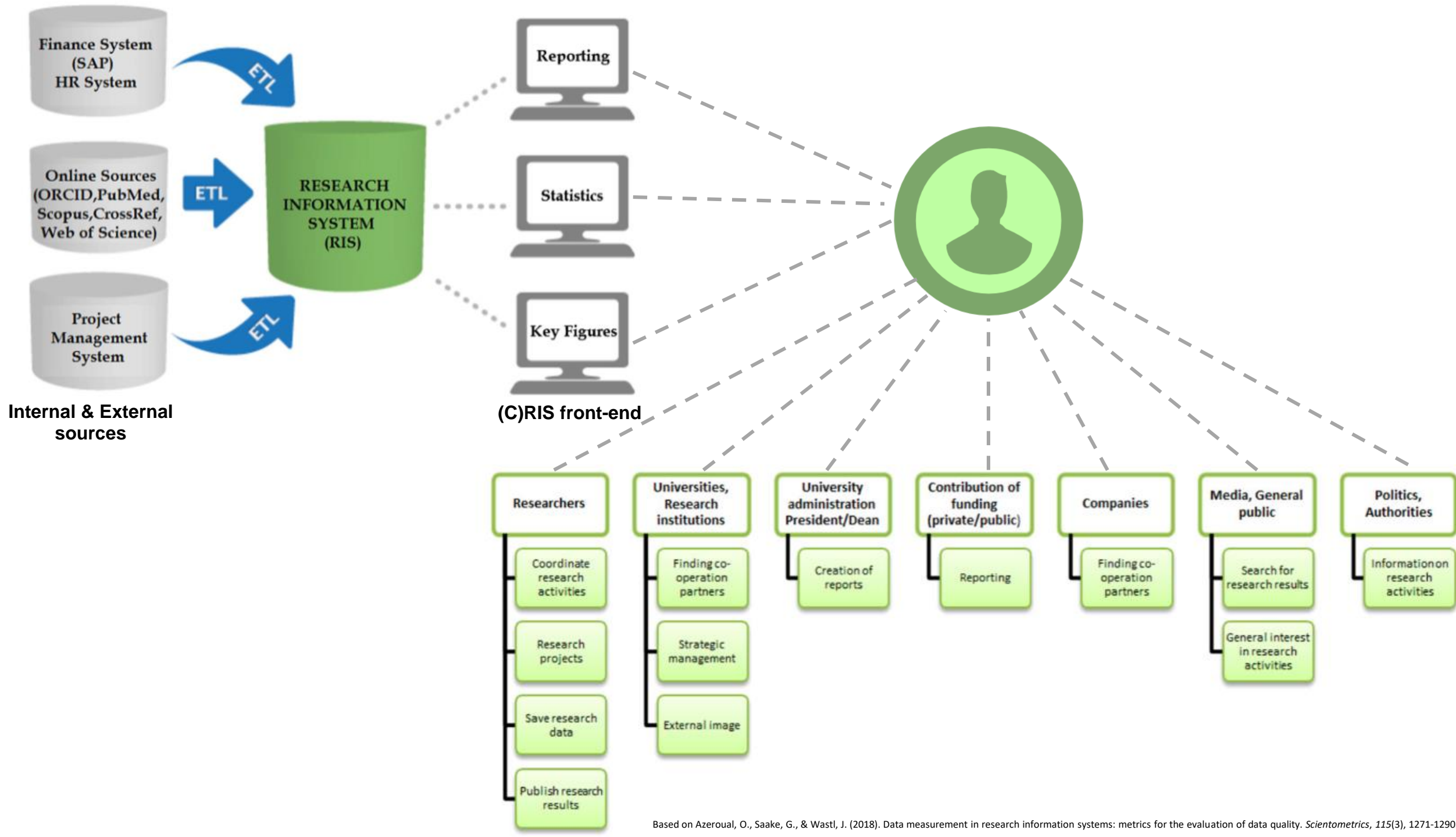
- ✓ Current Research Information Systems (CRIS, also RIS and RIMS – Research Information management Systems) are seen as:
“core elements of the technological solution since they provide rich additional meta-data on datasets and put the datasets and their meta-data into their proper context, and so significantly enhance the FAIRness of datasets”



(Terheggen and Smons, 2016)

- ✓ CRIS are not just data repository, but a key component of the research infrastructure landscape/ ecosystem, which is directly targeted and involved by the operational application and promotion of the FAIR principles.





CURRENT RESEARCH INFORMATION SYSTEMS

Three levels constitutes a set of propositions

Research information management systems are helpful to assess the FAIRness of research data and data repositories

Research information management systems contribute to the FAIRness of other research infrastructure

Research information management systems can be improved through the application of the FAIR principles

Top 10 RIMS/ RIS/ CRIS benefits





WHY DOES IT MATTER?

- ✓ The annual cost of not having FAIR data are estimated to a minimum of €10.2 billion per year and 16.9 billion in lost innovation opportunities (PwC EU Services, 2018)
- ✓ The actual costs are likely to be significantly higher due to unquantifiable elements such as the value of improved research quality & other indirect effects of FAIR data
- ✓ The impact on innovation, would account for over 60% of the likely costs, while the minimum cost encompassing indicators such as – time spent, license costs, research duplication, cost of storage and research retraction – for the remaining 40%
- ✓ These indicators, however, represent three areas applicable to all sectors (i.e., academia, private, public, non-profit) and can be described as
 - (1) impact on activities
 - (2) impact on collaboration
 - (3) impact on innovation



DOES IT MATTER?

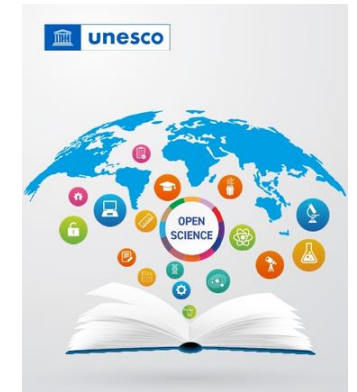
- ✓ In the **European Commission Open Science policy**, FAIR and open data sharing is one of the eight pillars of Open Science
- ✓ Similarly, in the **UNESCO Recommendation on Open Science**, FAIRness has become **an essential feature of what has been called “open science culture”** (multi-dimensional / complex concept)
- ✓ In **EOSC** - the FAIRsFAIR project addresses “*the development and concrete realization of an overall knowledge infrastructure on academic quality data management, procedures, standards, metrics and related matters, based on the FAIR principles*”, as a kind of general reference/guide to best practices in Higher Education and Research.

***At the moment EOSC have begun their own research to define and introduce guidelines to application of FAIR principles to digital objects [not necessarily limited to the research domain], thereby expanding the scope to the entire digital environment and all data-based objects, including research objects such as scientific articles and software, available on the Internet.



EUROPEAN COMMISSION: OPEN SCIENCE POLICY PLATFORM 8 PILLARS OF OPEN SCIENCE

Future of Scholarly Communication	Research Integrity
European Open Science Cloud (EOSC)	Rewards
FAIR data	Altmetrics
Skills	Citizen Science



**EUROPEAN OPEN
SCIENCE CLOUD**

METHOD

Exploratory study:

- (1) an overview of the challenges associated with research research information management,
- (2) analysis of each level based on a review of relevant studies,
- (3) definition of perspectives for further research, thereby raising an awareness of this topic and making **a call for other researchers to refer to it.**

Systematic Literature Review of euro-CRIS repository (!) - all relevant studies available in euroCRIS repository were selected → 3 studies were found to be relevant to illustrate the approach

The scope was extended by referring to projects and initiatives around the world identified using a **snowballing approach**, i.e., referring to the projects covered in the selected studies, OR based on our own experience (at both regional & (inter)national levels, representing different communities).

!!!The low number of relevant studies points out the limited body of knowledge on this topic, thereby making this study unique and constituting a call for action!!!



THE INTERNATIONAL ORGANISATION
FOR RESEARCH INFORMATION

euroCRIS was founded in 2002 as an international not-for-profit association that brings together experts on research information in general & RIS in particular.

One of the things that euroCRIS does is maintain a metadata standard known as Common European Research Information Format (CERIF).



National Library of Sweden

Sweden: FAIRness of research outputs

- ✓ in 2017, the government of Sweden gave the Swedish Research Council and the National Library of Sweden parallel assignments to propose **criteria and a method for assessing how well research data and scholarly publications** produced by Swedish organizations **comply with the FAIR principles**.
- ✓ Suggested criteria aimed at providing an “overall picture of FAIRness” of national research results, through the collected metadata:
 - (1) metadata quality (richness)
 - (2) licensing and persistent identifiers
 - (3) openness
 - (4) accessibility
 - (5) standard vocabularies.

Austria: FAIRness of research outputs



- ✓ Reflection on **how the implementation of a commercial CRIS** at the University of Vienna and the creation of a national network of CRIS managers from all Austrian universities (FIS/CRIS Austria) contributed to the **visibility, findability, accessibility** and **interoperability of research information**, through the development of standards (including identifiers and data models) and shared strategies.
- ✓ More recently, this network developed **a tool that enables tracking and monitoring of the transition to open access** based on data stored in local CRIS which is interoperable and connected with OpenAIRE.



Belgium: metadata-driven assessment of FAIRness and compliance with OS policy

FRIS

Flanders Research Information Space

Virtual space collecting all information about **publicly-funded scientific research** in Flanders.

- ✓ An “application profile for research data” based on experience with the Flemish research information system (FRIS), including various aspects of metadata such as *description, discovery, contextualization, coupling users, software and computing resources to data, research proposal, funding, project information, research outputs, outcomes, impact* etc. to assess FAIRness and compliance with open science policy
- ✓ a common metadata model and interoperation/ interoperability across multiple metadata models

To sum up...



- ✓ What these initiatives have in common is that **CRIS is used to assess different levels of FAIR-ness and FAIRness of the object under assessment** (as part of the open science assessment). This assessment is **primarily based on the collecting and processing metadata**
- ✓ Confirm that **CRIS has the potential and capacity at the institutional, regional or national level to contribute to the monitoring of open science policies and, in particular, to the follow-up of projects aimed at improving the FAIRness of research data, research repositories and other related research infrastructures**
- ✓ In addition, it has been clearly recognized that **CRIS has the potential to support and facilitate more responsible research assessment systems to reward and incentivize researchers for open science practices, including open and FAIR data*****

*****NB: FAIR data are not necessarily open, although open FAIR data is a target to go!**

CONTRIBUTING TO THE FAIRNESS OF RESEARCH INFRASTRUCTURES

- ✓ The FAIR data principles provide **a comprehensive framework and guidance on the criteria that well-preserved data must meet** & on the **standardization of data schemes**
- ✓ Several tools have been developed to assess the FAIRness of research data and / or data repositories: Australian Research Data Commons (ARDC) FAIR Data self-assessment tool, the Dutch DANS FAIRdat tool or the EUDAT Fair Data Checklist***
*****The results obtained using different tools tend to differ significantly with a very vague understanding on what should be done to improve the result if another tool assessed the level of FAIRness as sufficient → create new information silos and in most cases are not linked to professional assessment systems such as CRIS.**



Australian Research Data Commons

Data Archiving and Networked Services



FAIR Evaluation Services



IMPROVING THE FAIRNESS OF CRIS

FAIR is typically discussed at three levels:

- (1) **digital object** (e.g., dataset, video, journal, book etc.)
- (2) **metadata about this object on elementary level**, including title, creator, identifier, date etc.,
- (3) **metadata records with the reference to the body of metadata element on the object in a specific database**

also in line with the Science Europe Position Statement on Research Information Systems that suggests that “research information systems should foster the findability, accessibility, interoperability, and reusability of the data that they store by implementing the FAIR Guiding Principles for research activity data»

!!!we suggest to apply it not only to data and information, BUT also to the upper level of the data or information management systems, i.e., CRIS

CRIS itself can be improved by following and / or applying FAIR principles on it

CRIS is not only improving the FAIRness of research data management,

BUT

the FAIR principles are also beneficial for the further development of sustainable and FAIR CRIS*



FAIRNESS OF CRIS

Two levels can be distinguished:

- ✓ the need for standard data and metadata, especially persistent identifiers, requires **tools capable of producing and processing, handling them**, and this is **a strong argument in favor of CRIS as the central system (middleware) in the research infrastructure ecosystem**
- ✓ this need also calls for **more standardization of CRIS, improved data models and formats especially the long tail of less standardized research information systems** (cf. the large diversity and heterogeneity of CRIS)
- ✓ The standard format can improve CRIS interoperability BUT it is not enough – **CRIS should (also) prefer open identifier systems “to make things findable” and link information on source data and rights information to support access and facilitate reuse**
- ✓ The interconnection of infrastructures based on the FAIR principles is another example of an improvement in CRIS, which must fulfill certain technical requirements based on the FAIR principles.

Example: European OpenAIRE community accepts only CRIS meeting their FAIR requirements



!!! However, the FAIRness of research information management infrastructures has its specific «*limitations*» due to the nature of the data and the potential impact of their reuse - some of the data can be personal data and protected by privacy laws such GDPR, other data may be confidential being highly financial and of interest for competitors etc.




for ethical and legal reasons, the accessibility of CRIS data must be controlled and respect the above, i.e., it cannot be a guideline and require an openness of all data, following the H2020 Program Guidelines on FAIR Data of “*as open as possible, as restricted as necessary*”



TO SUM UP...

- ✓ It is crucial that the research information is available in such a way that it can be found, accessed, linked and reused as easily as possible (for authorized users) thereby being **as FAIR as possible**
- ✓ **Research information is not just research data & research information management systems such as CRIS are not just repositories for research data → they are much more complex, alive, dynamic, interactive and multi-stakeholder objects**
- ✓ CRIS are part of the research infrastructure ecosystem and are linked to data repositories, where the idea of CRIS partly overlaps with the main goal of FAIR principles
- ✓ **CRIS can (already does) improve the FAIRness of research infrastructures and data through the evaluation (monitoring) & standardization of data & metadata**

The improvement of FAIRness is a dual or bidirectional process ⇒ CRIS are beneficial for FAIR, and FAIR is beneficial for CRIS



CRIS promotes and contributes to the FAIRness of data and infrastructures

FAIR principles push for further improvement in the underlying CRIS data model and format

TO SUM UP...

- But nevertheless, the impact of CRIS on FAIRness is mainly focused on the:
 - (1) findability through the use of persistent identifiers,
 - (2) interoperability through standard metadata,while the impact on the other two principles, namely accessibility and reusability seems to be more indirect, related to and conditioned by metadata on licensing and access
- Paraphrasing “*FAIRness is necessary, but not sufficient for ‘open’*” → “**CRIS are necessary but not sufficient for FAIRness**”
- Rewards and incentives as recommended by Science Europe, is critical to ensure the “independence and transparency of the data, infrastructure and criteria necessary for research assessment and for determining research impacts”

CALL FOR ACTION!

- ✓ **more case studies are needed to explore the potential of research information management to monitor FAIR projects and infrastructures at the local, regional, national and international levels**
- ✓ **more empirical evidence needs to be presented on the real and specific impact of CRIS on the development of FAIR data repositories and other research infrastructures, with a particular focus on standardization**
- ✓ **further development of CRIS data models and formats should focus on FAIR principles, especially findability and interoperability, in an explicit way.**
- ✓ **ethical and legal aspects of accessibility of CRIS data require further investigation to get a full picture of what it really means to apply the FAIR principles to research information management*****

******this is a research currently conducted by the EOSC Task Force by means of both surveys, interviews, case studies and other activities, it can and should be supplemented with other independent and use-case based studies***



THANK YOU FOR ATTENTION! QUESTIONS?

For more information, see

[ResearchGate](#), anastasijanikiforova.com

For questions or any other queries,

contact us via email -

Nikiforova.Anastasija@gmail.com,

azeroual@dzhw.eu