Open Access and Data

Prof. Rūta Petrauskaitė
ruta.petrauskaite @vdu.lt

VYTAUTAS
MAGNUS
UNIVERSITY
MCMXXII

Education 360°

Areas of Open Science, types of data

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

- Open Access to Publication, Research Infrastructures and Data as different approaches to openness
- Concept of data is broad and diverse, comprises many types
  - Based on time
    - Post-research/output data (most popular/default use)
    - Pre-research/input data (collected for different purposes: public administrative, clinical, internet, previous research, etc.)
  - Based on place
    - Institutional repositoria
    - National and/or disciplinary data archives
    - Data-based research infrastructures
  - Based on type (numerical, textual, visual, etc.)

  - "*Research data is understood in this policy to mean the registration/recording/reporting of numerical scores, textual records, images and sounds that are generated by or arise during research projects*"
    - Open Access to research data. Policy for the Research Council of Norway

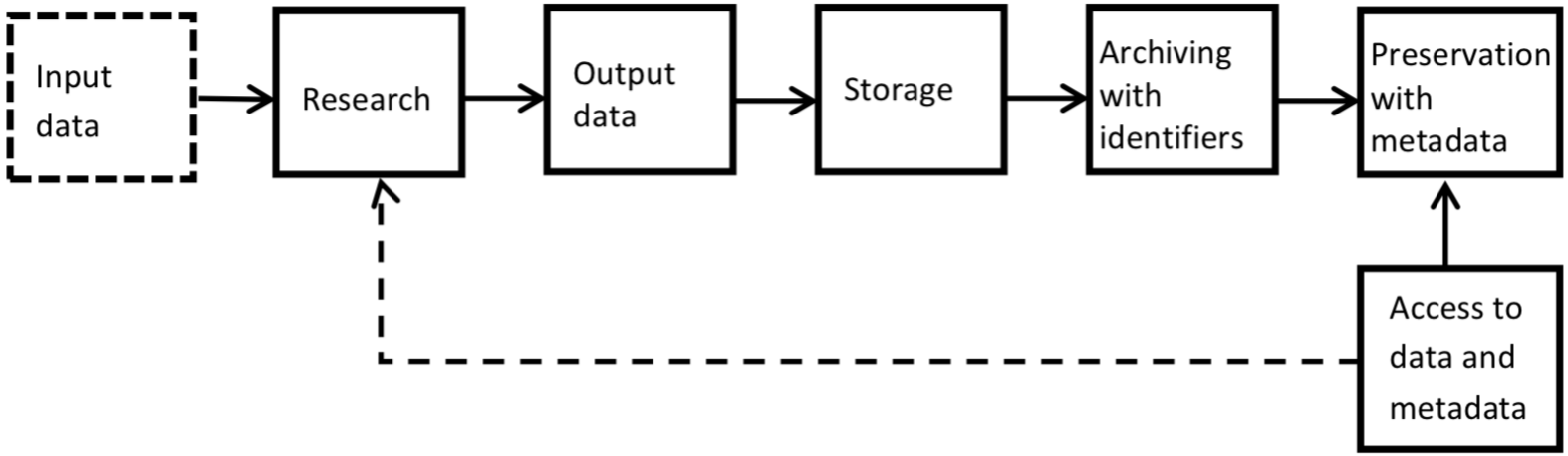
OA to data. Unified funders' approach to the results of research

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

- UK 2013 applies the same rules to data as to publications they are attached
- Canada 2015 requires either to open data together with publications or to preserve it for 5 years
- ERC 2014 recommends to preserve and to share RD within the limits of copyright, personal data, national security restrictions. Supports OA disciplinary data archives
- Norway 2014 has "an open by default" policy to data other than to publications. Apply only to output data, charge users for its retrieval
- Reasons to exempt data from the general principle: security concerns, sensitive personal data, commercial, legal factors
- https://roarmap.eprints.org/

Life cycle of data

ÅOpen Access to research data. Policy for the Research Council of Norway



VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

Status quo and challenges

Data coming from publicly funded research is not always open
Lack of interoperability, closed and silo-based approaches to research
Fragmentation that hampers data-driven science
A bad need for High Performing Computer infrastructure to process data
Revisable EU Copyright legislation for Text and Data mining

https://ec.europa.eu/digital-single-market/en/news/communication-european-cloud-initiative-building-competitive-data-and-knowledge-economy-europe

The majority of the challenges (…) are social rather than technical
The EOSC aims to accelerate the transition to more effective Open Science and Open Innovation in a Digital Single Market by removing the technical, legislative and human barriers to the re-use of research data and tools, and by supporting access to services, systems and the flow of data across disciplinary, social and geographical borders

https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

What is EOSC made of?

„The Cloud can be understood as the combination of three interdependent elements:
the data infrastructures which store and manage data;
the high-bandwidth networks which transport data;
and the ever more powerful computers which can be used to process the data.

The ability to analyze and exploit this Big Data is having an impact on the global economy and society, opening up the possibility of major industrial and social innovations. A key part of this impact is the change in the way scientific research is carried out, as we move rapidly towards Open Science“. Other key drivers: Digital Single Market, 4 th Industrial Revolution, European Data Infrastructure, High Performance Computing, etc.

https://ec.europa.eu/digital-single-market/en/news/communication-european-cloud-initiative-building-competitive-data-and-knowledge-economy-europe

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

"What's in a name" of EOSC?

European: research and innovation are global. The EOSC cannot be built exclusively in and for Europe.

Open: not all data and tools can be open. There are exceptions to openness, such as confidentially and privacy. Open is also often confused with 'for free'. Free data and services do not exist1.

Science: the use of the term science explicitly includes the arts and humanities, and in fact no current or future discipline should be excluded from the EOSC. In addition the Science Cloud infrastructure should support not only innovative scientific research but also societal innovation and productivity. The EOSC should also support broad societal participation in Open Innovation and Open Science

Cloud: It can be misinterpreted to indicate that the EOSC is mostly about hard ICT infrastructure and much less about a commons of data, software, standards, expertise and policy related to data-driven science and innovation. The term 'cloud' is a metaphor to help convey the idea of seamlessness. https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

Research data infrastructures

EOSC is envisaged as a federated, globally accessible environment where researchers, innovators, companies and citizens can publish, find and re-use each other's data and tools for research, innovation and educational purposes based on existing research infrastructures

Research data infrastructures besides data include metadata, tools, standards, services, expertise, policies, etc. to insure re-usability of data

Research data, if made FAIR, promote better data stewardship in data-driven and open science that is the key issue for evidence-based decisions

EOSC will be build on existing research infrastructures, both hard and soft, cl. ESFRI ERIC research infrastructures CLARIN ERIC for linguistic resources and DARIAH for digital humanities in particular

Findable Accessible Interoprable and Reusable Data

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

FAIR is a set of principles ensuring that research data is reusable. Data is not either FAIR or unFAIR, it is a spectrum and, a continuum from re-useless to reusable. Variable degree of FAIRness comprise

findable but not accessible (or restricted access)

findable and accessible but not interoperable

findable, accessible and interoperable but not reusable data that allow a great deal of freedom

What FAIRness of data is NOT about

It is NOT a standard, not a prescriptive approach, but permissive
It is NOT meant just for humans, but also for machines as FAIR data have to be machine actionable
It is NOT equal to Open data as far as national security, competitiveness and personal privacy is concerned. FAIRness requires clarity and transparency around the conditions governing access and reuse.

Findable and Accessible Data

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

TO BE FINDABLE:
F1. (meta)data are assigned a globally unique and eternally persistent identifier (PID).
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

TO BE ACCESSIBLE:
A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.
https://www.force11.org/group/fairgroup/fairprinciples

Interoperable and Reusable Data

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

TO BE INTEROPERABLE:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:
R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

https://www.force11.org/group/fairgroup/fairprinciples

Is FAIR data sufficient for the EOSC ? Some additions

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

TO BE REPRODUSIBLE

TO BE REPEATABLE

TO BE REPLAYABLE

TO BE REPURPOSABLE

TO BE REFERENCABLE

TO BE RESPECTFUL

TO BE SUSTAINABLE

Questions concerning access to data FAIR

VYTAUTO DIDŽIOJO UNIVERSITETAS

Responsibility

Who is responsible for a long-term preservation, curation and OA to data – research funding or performing institutions?

Financial support

- **What and how to support: a researcher's data producing project, his institution or research data curating infrastructure?**
- **At what expense: additional donation, research or its dissemination?**
- **For whom OA to data is affordable?**

Closing remarks

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

Data should be as open as possible as closed as necessary

It is more expensive to loose post-research data than to preserve them and it is even more expensive to ignore research data than to collect and to preserve them

The coolest thing to do with your research results (data) will be thought of by someone else

VYTAUTO DIDŽIOJO UNIVERSITETAS
MCMXXII

References

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS European Cloud Initiative - Building a competitive data and knowledge economy in Europe EOSC. Brussels, 19.4.2016 COM(2016) 178 final

Force 11 Group. The Fair Data Principles. https://www.force11.org/group/fairgroup/fairprinciples

Mark D. Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* **3**, 15 March 2016.

Guidelines on FAIR Data Management in Horizon 2020, H2020 Programme. Version 3.0 26 July 2016

Realising the European Open Science Cloud. First report and recommendations on the European Open Science Cloud Realising the European Open Science Cloud of the Commission High Level Expert Group. Luxembourg: Publications Office of the European Union, 2016.

EOSC Declaration. EUROPEAN COMMISSION DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION. Brussels, 10 July 2017. https://www.elixir-europe.org/system/files/eosc-declaration.pdf

Turning FAIR into reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data, European Commission, Brussels, 2018.

Thank you

Education 360°