Open research data
importance and benefits of openness

Store an Open your Data!
Seminar for researchers, doctoral candidates and ÅAU personnel

Turku, 23 October 2017
Åbo Akademi University

Pauli Assinen
University of Helsinki / OpenAIRE NOAD
As open as possible as closed as necessary
Open Scholarship
OPENAIRE
Open Science

Promote & align OA policies
Participatory Infrastructure
Link to national infrastructures
Link to domain discipline data infrastructures
Link to open gov data initiatives
Link to open education & learning environments

Put research in context
OpenAIRE - Openness

- Europe’s diverse landscape requires local support
- Different practices, different mentalities

People support network
- 33 expert nodes in all Europe
  - (OA) Policy aligning
  - Technical assistance
  - Training
- Outreach to international community via COAR
How openness might influence?

What motivates you in research?

What motivates EU to fund research?

What motivates EU to enhance open science?

EU policies and H2020 Guidelines

Who gets benefits?

OpenAIRE
How openness might influence?

Goals

Motivation

Research and Development activities

Outcomes & Benefits
**What motivates you in research?**

**Motivation in research**

- Solve problems
- Research degree
- Get respectability
- Service to society
- Intellectual joy of creative work

+ directives of government, employment conditions, curiosity about new things, desire to understand causal relationships, social thinking and awakening

**Sources:**
What motivates to make research?

Motivation in research

1. Desire to get a research degree along with its consequential benefits;
2. Desire to face the challenge in solving the unsolved problems, i.e., concern over practical problems initiates research;
3. Desire to get intellectual joy of doing some creative work;
4. Desire to be of service to society;
5. Desire to get respectability

However, this is not an exhaustive list of factors motivating people to undertake research studies. Many more factors such as directives of government, employment conditions, curiosity about new things, desire to understand causal relationships, social thinking and awakening, and the like may as well motivate (or at times compel) people to perform research operations.

What motivates EU to fund research?

EU Commission

European Research and Development activities

More competitive EU
What motivates EU to enhance open science? EU Guidelines

Broader access to scientific publications and data therefore helps to:

- **build on previous research results** (improved quality of results)
- **encourage collaboration** and avoid duplication of effort (greater efficiency)
- **speed up innovation** (faster progress to market means faster growth)
- **involve citizens and society** (improved transparency of the scientific process).

Source: H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 Version 3.2
21 March 2017
H2020 Guidelines

Dissemination plan

Research results

Data management plan

Decision to disseminate/share

Decision to exploit/protect

Patenting (or other form of protection)

Depositing research data

Publications

And/or

Gold OA

Green OA

Access and use free of charge

Restricted access and/or use

Source: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2 21 March 2017

Store and Open your Data. Turku, Finland, 23rd October 2017

Open research data – Pauli Assinen
Who gets benefits?

EU Commission Strategy
EU Infra & Services
Horizon 2020

EU funded Research and Development activities
Results
Goal: More competitive EU

Other funded Research and Development activities
Results
Goal: Enhance scientific knowledge, earn credit…?

Other funded Research and Development activities
Results
Goal: Scientific contribution…?

National policies
Institutional policies
Funders’ policies
Publishers’ policies
Evaluators’ policies
Open Science activists
Service providers
Field of science traditions
Who gets benefits? Benefits of Open Access
What is research data?  
H2020 Guidelines

Research Data

- Refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation.
- In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.
- Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge.

Source: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2 21 March 2017
Research, Development and Innovation in EU

EU Commission
Open Science, Open Innovation, Open to the world – a vision for Europe

Strategy / Targets
- "R&D/innovation and more efficient energy use makes us more competitive and creates jobs"

DG for Research and Innovation (DG RTD)
European Research and Development activities

DG for Communications Networks, Content & Technology (DG CONNECT)
Digital Single Market
European Cloud Initiative
- European Open Science Cloud (EOSC)
- European Data Infrastructure (EDI)
- High Performance Computing (HPC)

Horizon 2020 Framework Programme
- Work Programmes
- Annotated Model Grant Agreement
- Guidelines

More competitive EU
H2020 mandate from EU parliament

(28) To increase the circulation and exploitation of knowledge, open access to scientific publications should be ensured. Furthermore, open access to research data resulting from publicly funded research under Horizon 2020 should be promoted, taking into account constraints pertaining to privacy, national security and intellectual property rights.

(28) To increase the circulation and exploitation of knowledge, open access to scientific publications should be ensured. Furthermore, open access to research data resulting from publicly funded research under Horizon 2020 should be promoted, taking into account constraints pertaining to privacy, national security and intellectual property rights.

As open as possible
as closed as necessary
H2020 Framework Programme

- Horizon 2020 will **help to achieve smart, sustainable and inclusive economic growth.** The goal is to ensure Europe produces world-class science and technology, **removes barriers to innovation** and makes it easier for the public and private sectors to work together in delivering solutions to big challenges facing our society.

- Horizon 2020 is the biggest EU Research and Innovation programme ever with nearly **€80 billion of funding available over 7 years (2014 to 2020)** – in addition to the private investment that this money will attract. It promises more breakthroughs, discoveries and world-firsts by taking great ideas from the lab to the market.

(HORIZON 2020 in brief. The EU Framework Programme for Research & Innovation
H2020 Guidelines
EC’s Guidelines on Data Management in Horizon 2020

FAIR principles for research data
1. Findable
2. Accessible
3. Interoperable
4. Re-usable


Store and Open your Data, Turku, Finland, 23rd October 2017
Open research data – Pauli Assinen
H2020 Guidelines
EC’s Guidelines on Data Management in Horizon 2020

FAIR principles for research data - FINDABLE

F1. (meta)data are assigned a **globally unique and eternally persistent identifier**.

F2. data are described with **rich metadata**.

F3. (meta)data are **registered or indexed in a searchable resource**.

F4. metadata **specify** the data identifier.

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

Store and Open your Data, Turku, Finland, 23rd October 2017
Open research data – Pauli Assinen
H2020 Guidelines
EC’s Guidelines on Data Management in Horizon 2020

FAIR principles for research data - 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.

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

Store and Open your Data, Turku, Finland, 23rd October 2017
Open research data – Pauli Assinen
H2020 Guidelines
EC’s Guidelines on Data Management in Horizon 2020

FAIR principles for research data - 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.

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

Store and Open your Data, Turku, Finland, 23rd October 2017
Open research data – Pauli Assinen
H2020 Guidelines
EC’s Guidelines on Data Management in Horizon 2020

FAIR principles for research data – **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.**

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

*Store and Open your Data*, Turku, Finland, 23rd October 2017
Open research data – Pauli Assinen
The Open Science agenda will be supported, notably through **dedicated data driven actions**, the embedding of approaches and the **mainstreaming/promotion of Open Science principles**.
• Open Access and Open Data defaults
• Opt-outs possible, must have reasons out written
• Openness not evaluation criterion, but must be mentioned
  how data is handled or also if there will not be any data

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

Store and Open your Data, Turku, Finland, 23rd October 2017
Open research data – Pauli Assinen
ARTICLE 29 – DISSEMINATION OF RESULTS — OPEN ACCESS — VISIBILITY OF EU FUNDING

• 29.2 Open access to scientific publications
• 29.3 Open access to research data
• 29.4 Information on EU funding — Obligation and right to use the EU emblem
• 29.5 Disclaimer excluding [Commission][Agency] responsibility
• 29.6 Consequences of non-compliance
Actions participating in the pilot must draw up a **data management plan (DMP)** within the first 6 months of the project implementation.

The data management plan must support the management life-cycle for all data that will be collected, processed or generated by the action. It must cover how to make data findable, accessible, interoperable and re-usable (FAIR), including:

- the handling of data during and after the project
- what data will be collected, processed or generated
- what methodology and standards will be applied
- whether data will be shared / made open access (and how) and, if any, what data will not be shared / made open access (and why)
- how data will be curated and preserved.

The data management plan should be updated (and become more precise) as the project evolves. New versions should be created whenever important changes to the project occur (e.g. new data sets, changes in consortium policies, etc), at least as part of the mid-term review (if any) and at the end of the project.

Source: H2020 Programme / AGA – Annotated Model Grant Agreement, Version 4.0.1, 20 June 2017

Store and Open your Data, Turku, Finland, 23rd October 2017

Open research data – Pauli Assinen
29.6 Consequences of non-compliance

- If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).
- Such a breach may also lead to any of the other measures described in Chapter 6.
H2020 regulations and guidelines

Europe 2020 strategy and mandate from EU parliament
- "R&D/innovation and more efficient energy use makes us more competitive and creates jobs"

Horizon 2020 Framework
"Central role of knowledge and innovation in generating growth"

Work Programmes 2018-2020

Calls

Proposals and evaluation

Annotated Model Grant Agreement and other agreements

Data Management Plan

Open research data – Pauli Assinen
Researcher’s context

- Research projects in different phases
- Teaching and supervising students and postgraduates
- Development of teaching in her faculty
- Research subjects
- Research instruments and other infrastructure
- Collaboration existing and potential partners
- Requirements from university administration and funders

Adapted from Mari Elisa Kuusniemi: https://www.slideshare.net/MariKuusniemi/unica-data-librarian

Store and Open your Data, Turku, Finland, 23rd October 2017
Open research data – Pauli Assinen

Photo: © Joel Grandell / University of Helsinki
H2020 Impact and Future

- H2020 Interim evaluation results

- FP9 planning
  - LAB – FAB – APP Investing in the European future we want - Report of the independent High Level Group on maximising the impact of EU Research & Innovation Programmes
    - European Commission Directorate-General for Research and Innovation Directorate A – Policy Development and Coordination Unit A.5 – Better Regulation
LAB – FAB – APP recommendations

• Recommendations for designing a post-2020 EU programme for research and innovation
• “We need research (“Labs”), innovation (competitive fabrication (“Fabs”) and applications for the benefit of all (“Apps”).
• Recommendation 8. Mobilise and involve citizens
  • “Whenever possible, citizen science should be encouraged, where citizens become providers and users of data. This will enforce and give meaning to the policy of open access to publications and data; this openness should enable citizens and citizen groups to participate in evidence-based policy and decision making.”

Source: LAB – FAB – APP Investing in the European future we want - Report of the independent High Level Group on maximising the impact of EU Research & Innovation Programmes
Recommendation 10. Make international R&I cooperation a trademark of EU research and innovation

“… open the EU R&I programme to association by trading partners of a similar level of excellence, such as Canada and Australia. Association of non-EU countries to future EU R&I programmes should be governed by excellence in R&I, not confined to a particular part of the world. This will make the EU programme the potential nucleus of a global programme for open science and open innovation, exploiting good regulatory and improving international trading and investment conditions.”

Source: LAB – FAB – APP Investing in the European future we want - Report of the independent High Level Group on maximising the impact of EU Research & Innovation Programmes
EU links

- EC’s Guide on Open Access to Scientific Publications and Research Data in Horizon 2020:
- EC’s Guidelines on Data Management in Horizon 2020:
- EC’s Agenda on Open Science:
- DCC’s DMPonline tool: https://dmponline.dcc.ac.uk
Who we are

- In 24x7 operation since Dec 2010
  - OpenAIRE
  - OpenAIREplus
  - OpenAIRE2020
  - A legal entity in 2017?

- Consortium of 50 partners

Open Access experts
- Institutional, national and international perspectives on OA policies & e-Infrastructures

Information & Computer Science experts
- Building efficient e-Infra technologies
- State of the art technologies (big data, linked data)

Legal experts
- Legal & policy recommendations

Data communities
- Best practices for data
- Linking to data infrastructures
WHAT WE DO
Result: Integrated Scientific Information System

11.5 mi unique publications
7 mi authors
590+ data providers
130K publications linked to projects from 3 funders
5.5 K datasets linked to publications
33K organizations
The recipe

1. Open data
   ...for sharing, re-using, validating
   ...at the source

2. Make data providers interoperable
   ...for clean, enriched, linked data

3. Provide robust services
   ...for building trust
   ...for uptake from 3rd party providers

Transparent & validated decision making processes
2. Interoperability

- Guidelines and standards
  - OpenAIRE guidelines for literature, data, CRIS
  - **Global alignment and adoption** (RDA, WDS, W3C, ...)

- Identifier schemes
  - Synergies with ongoing e-Infras

- Uniform vocabularies
  - Interdisciplinary classification
  - Multilinguality (e.g., EUROVOC)

- Links to other domains
  - Mapping of data models (DCAT, LOM, ...)
  - LOD
3. Service Oriented Infrastructure

- Support & Training
- Sharing scientific results
  - OA Publishing & Depositing
  - Validating and Registering
- Discovery & Crowdsourcing
- Reporting & Monitoring
- Analytics
  - ...

...
Services

VALIDATING

Data providers
595 OpenAIRE compatible data sources

<table>
<thead>
<tr>
<th>Name</th>
<th>Country</th>
<th>Documents</th>
<th>Institution</th>
<th>Compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>/oops/</td>
<td>Germany</td>
<td>4801 (updated 2015/04/14)</td>
<td>Carl von Ossietzky Universität Oldenburg</td>
<td>OpenAIRE Basic (DRIVER OA)</td>
</tr>
<tr>
<td>A+BE</td>
<td>Netherlands</td>
<td>-</td>
<td>Delft University of Technology</td>
<td>OpenAIRE Basic (DRIVER OA)</td>
</tr>
<tr>
<td>Aaltodoc Publication Archive</td>
<td>Finland</td>
<td>14557 (updated 2015/04/13)</td>
<td>Aalto University</td>
<td>OpenAIRE Basic (DRIVER OA)</td>
</tr>
<tr>
<td>AAPP</td>
<td>Physical, Mathematical, and Natural Sciences</td>
<td>-</td>
<td>-</td>
<td>OpenAIRE 3.0 (OA, funding)</td>
</tr>
<tr>
<td>ABACUS. Repositorio de Producción Cientifica</td>
<td>Spain</td>
<td>3379 (updated 2015/04/14)</td>
<td>Universidad Europea</td>
<td>OpenAIRE Basic (DRIVER OA)</td>
</tr>
</tbody>
</table>
Services

DISCOVERY

Crowdsourcing
Influence of auditory and visual feedback for perceiving walking over bumps and holes in desktop VR
Luca Turchet; Maud Marchal; Cuyer, Anatole L.; Rolf Nordahl; Stefania Serafin (2010)
Projects: NIW (222107)

A preliminary idea for an 8-competitive, log<sub>2</sub>DMAX + log<sub>2</sub>1/U asymptotic-space, interface generation algorithm for two-level hierarchical scheduling of constrained-deadline sporadic tasks on a uniprocessor
Björn Andersson (2011)
Projects: ARTISTDESIGN (214373)

A game theoretic evaluation of rate adaptation strategies for IEEE 802.11 based wireless LANs
Nicola Baldo; Andrea Zanella (2009)
Projects: NEWCOM++ (216715)

3D HapticWebBrowser
Nikolaos Kaklakis; Konstantinos Votsis; Konstantinos Moustakas; Dimitrios Tzovaras (2010)
Projects: AEGIS (224348)

Transforming statistical linked data for use in OLAP systems
Mpgen, Benedikt K.; Andreas Harth (2011)
Projects: PLANETDATA (257641)

TypeChef
Andy Kenner; Stner, Christian K.; Steffen Haase; Thomas Leich (2010)
Projects: SCALPI (203099)
Measurement of the ZZ production cross section using the full CDF II data set


Publisher: American Physical Society

Languages: English

Types: Article

Subjects: High Energy Physics - Experiment, Physics Institute, 530 Physics, hep-ex

Discover through pilot similarity algorithms. Send us your feedback.

Comparative Analysis of Nuclear Transfer Embryo-Derived Mouse Embryonic Stem Cells. Part I: Cellular Characterization

Comparative Analysis of Nuclear Transfer Embryo-Derived Mouse Embryonic Stem Cells. Part II: Gene regulation

Comparative Analysis of Nuclear Transfer Embryo-Derived Mouse Embryonic Stem Cells. Part II: Gene Regulation

2012

2012

2012
MONITORING

Reporting
OA estimate for FP7

Publications distribution per country from institutional repositories

FP7
116.6K publications
Linked to 11.5 K projects
50% OA

Increase of FP7 publications in institutional repos
Other types of monitoring
ANALYTICS to support decision making
What analysis?

- Academia – who is doing what, what are the hot topics/trends
  - Map of authors, relations, collaborations, trends, ...
- Research analytics for decision making
  - Infer clusters, communities, correlations, of programmes or scientific areas
    - Over geographical or thematic regions
    - Over time
  - Identify overlaps or gaps
  - Correlate with other funder programmes
  - ...

From data acquisition to analysis

1. Get the data. Involve as many sources as possible.


   Link to many data sources & external dbs.

2. Classify content based on known classifications.

   Cluster and classify based on statistical methodology – Find the unknown (trends)

3. Work with experts to give meaning to statistical output.

   Visualize: use online tools to see different facets.
Relation of HEALTH programme to other FP7 programmes based on TDM/topic modelling techniques.
Machine Learning, Feature selection, .. is “hot” after 2003
Thank you!

www.openaire.eu
@openaire_eu
facebook.com/groups/openaire
linkedin.com/groups/OpenAIRE-3893548
pauli.assinen@helsinki.fi
twitter: @passinen