Study on the cost of not having FAIR research data

A PwC study for DG RTD of the European Commission delivered under SC519DI07171
Agenda

Objectives of the study

Cost of not having FAIR data

Methodology

Indicators

Cost

Preliminary findings

Cost-benefit measuring mechanism

Conclusion
Objectives of the study
Objectives of the study

01 To determine the cost of not having FAIR research data

02 To estimate the costs and benefits of FAIR research data

03 To provide recommendations on the next steps for making research data FAIR
**Study timeline**

**Current work**

- **Report**
  - Cost of not having FAIR research data

**Seeking input**

- **Cases studies to apply to the cost benefit measuring mechanism**

- Interviews with experts

- **Report**
  - FAIR cost benefit measuring mechanism

- **RDA 11th Plenary Meeting**

- **Report**
  - Policy recommendations

- **Report**
  - Report on the application of the FAIR cost benefit measuring mechanism
Estimating the cost of not having FAIR research data
What does not having FAIR mean and how much does it cost?

Why?

Making research data FAIR has a cost, but not doing it has also a cost.

This study:

• assesses the cost of not having FAIR research data in Europe; and
• creates a quantitative methodology for assessing the cost of not having FAIR research data, which would be easily applicable and repeatable in the future.

Who is bearing the costs?

• Research funders
• Research (data) infrastructure
• Research performing organisations

What?

Research data

Research Data is defined as information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation

- Underlying data and the associated metadata
- Any other data and the associated metadata

Open data

The concept of FAIR is not complete without open data. We consider that 80% of the research data could be made open as a consequence of implementing the FAIR principles

Limitations

Does not take into account the cost of implementing FAIR
**Cost estimation approach**

### Indicators identification

- Desk research
- Interviews with subject matter experts
  - Simon Hodson
  - Peter Wittenburg
  - Sarah Jones
  - Barend Mons
  - Ross Wilkinson
  - Peter Doorn
  - Sarah Kildahl Nielsen
  - Esben Aton Schultz

### Indicators Categorisation

#### FAIR Principles
- Indicator #1: Time spent
- Indicator #2: Cost of storage
- Indicator #3: Research retraction
- Indicator #4: Research duplication
- Indicator #5: Economic growth
- Indicator #6: Data license cost

#### Impact on research activities
- Impact on cross-fertilisation
- Impact on innovation

### Cost of not having FAIR research data

#### Cost calculation
- Estimation of the **cost of not having FAIR research data** per indicator.
- The actual cost is likely to be much higher due to unquantifiable elements such as the value of improved research quality and other indirect positive spill-over effects of FAIR research data

#### Indicators quantification

- Data from micro-economic scale
  - Domain specificity
  - Geographical coverage
  - Reusability
- Extrapolation to EU
  - # researchers + average salary
  - Margin of error
  - Conservative estimates
  - Spill-over effects

---

March 2018
Main types of sources

**Official sources (EU, Member States)**
- EUROSTAT
- DG RTD

**Open data**
- European Data Portal
- Open journals

**Data from surveys or research data**
- DAF survey in the UK
- Specific published articles

**Data from cost analysis**
- European Bioinformatics Institute
- Australian National Data Service
**Indicators quantification**

**Indicator #1 : Time spent**

Wages – Time Research/Activity – Inefficiency rate

**Indicator #2 : Cost of storage**

Cost of storage – Data Volume/Researcher – Storage Reduction

**Indicator #3 : Research retraction**

Wages – Time Research/Activity – Retraction Inefficiency rate

**Indicator #4 : Research duplication**

Research funds – #Researches– Overlap proportion

**Indicator #5 : Potential economic growth**

Spinoff Network – Spinoff missed opportunities

**NOTE:** Due to insufficient data we were not able to quantify indicator #5
## Cost calculation

<table>
<thead>
<tr>
<th>Indicator #1: Time spent</th>
<th>Indicator #2: Cost of storage</th>
<th>Indicator #3: Research retraction</th>
<th>Indicator #4: Research duplication</th>
<th>Indicator #5 Potential economic growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Split academics &amp; non-academics</td>
<td>• Data volume per researcher/year</td>
<td>• Split academics &amp; non-academics</td>
<td>• Funds allocated to research grant in EU28</td>
<td>• Positive impact on the number of citations</td>
</tr>
<tr>
<td>• EU28 Average researcher salary</td>
<td>• Cost per TB</td>
<td>• EU28 Average researcher salary</td>
<td>• Number of research grant in EU28</td>
<td>• Societal value of research increased</td>
</tr>
<tr>
<td>• Number of researchers</td>
<td>• Number of repositories where data is stored</td>
<td>• Number of researchers</td>
<td>• Number of suspicious overlaps</td>
<td>• Improve the availability of research data</td>
</tr>
<tr>
<td>• Time dedicated to research</td>
<td>• Data retention period</td>
<td>• Time dedicated to research</td>
<td>• Time wasted due to retraction</td>
<td>• Benefits for making research more accessible</td>
</tr>
<tr>
<td>• Time wasted</td>
<td>• Reduction of the number of repositories where data is stored (FAIR)</td>
<td>• Time wasted due to retraction</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \#1 \text{ in } \€ + \#2 \text{ in } \€ + \#3 \text{ in } \€ + \#4 \text{ in } \€ = \text{Total } \€ \]
Preliminary estimations of the cost of not having FAIR data in EU-28

Disclaimer: This study is still ongoing. The figures presented have not been validated and are subject to change.
Interpreting the overall cost of not having FAIR research data as a single value overlooks many non-quantifiable benefits of FAIR.

Nonetheless, at the EU-28 level, the measurable cost of not having FAIR research data makes an overwhelming case in favour of the implementation of the FAIR principles.
Cost-benefit measuring mechanism for FAIR data: creating a solid business case
Cost-benefit measuring approach

Develop a measuring mechanism for cost and benefits deriving from FAIR research data.

I. Input sheets
II. Assumptions sheet
III. Calculation sheet
IV. Result sheet
V. Dashboard

Application in 2 case studies

Are you a research organisation (e.g. university or research lab) with one of the following characteristics:

- European region/country
- Discipline-specific

Contact us and be one of the two featured cases!
Conclusion
Thank you for your attention

Stay connected
nikolaos.loutas@pwc.com
barthelemy.florian@pwc.com

Get news from the FAIR research data study
Connect with DG RTD
Responsible officer:
athanasios.karalopoulos@ec.europa.eu

Nikolaos Loutas, PhD
Director Analytics & AI

Florian Barthélemy
Consultant Data management