Open Data – What Research Communities Think About It

EUDAT working paper

Marie Sandberg, EUDAT (CSC)

EUDAT
www.eudat.eu

EUDAT

Main focus
research communities opinions

- Open access to research data priority in H2020, international organizations and in several countries

- EUDAT strong belief in OA

- As a service provider EUDAT should listen carefully to the research communities

- EUDAT working paper: gather research communities opinions

- Lessons from other studies: RECODE, RDA etc…

EUDAT

Keeping the balance

EUDAT
fundamental
belief in
open access

- research communities in the driving seat
- respect communities different approaches and challenges
- flexible enough for new communities to be integrated

EUDAT Guiding Principles

- Data deposited with the EUDAT CDI will be preserved long-term

- Data are best curated in their own communities.

- Access to data in the EUDAT CDI is free at the point of use

- For an EUDAT community repository to be designated a Trustworthy Digital Repository (TDR), it follows that EUDAT services and infrastructure must be a suitable target for "TDR outsourcing"

- EUDAT will not assert ownership of any data it holds

EUDAT

Open Access Principles

- Two further principles on open access:
  - all data in the CDI should, in time, become full open access. Open access is the norm for CDI data;
  - embargo periods for original producers are fully supported, on condition that such data become openly accessible when the embargo period expires

- These imply:
  - policy harmonisation
  - a common licensing scheme. EUDAT currently recommend CC4.0 and Open Data Commons

EUDAT

Interiews with research communities

- Interviews with fourteen pan-European research communities collaborating with EUDAT

- Fields of life sciences, earth and atmospheric science, astrophysics, climate science, biodiversity, agricultural science, social science and humanities
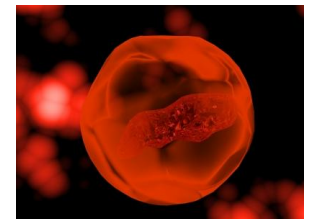
Three main conclusions

- growing awareness of the global open access move, of potential benefits and the necessity to implement open access policies within their communities

- many challenges still to be addressed

- research communities request further guidance and training about all practicalities surrounding open access

EUDAT

Growing awareness

- Several of the communities currently developing policies and recommendations to facilitate open access to their research data

- Working groups with legal experts to assist with and examine challenges in data sharing and reuse

- Take part in initiatives to agree on common principles of data management and sharing across disciplines and Ris. Many referred to RDA WGs/IGs

- Express benefits of open access: new discoveries with cross-disciplinary data, put a stamp of quality on the research by making it accountable

EUDAT

Main challenges –
Not all data can or should be open

- **Personal and sensitive data: life sciences, humanities and social sciences**

  *- Risk with anonomyzed data: if different data sources are connected, it may be possible to re-identify persons.

  - With controlled-behaviour licences data owners and service providers can harm researchers career by denying further access if they break the terms. Works as a governance safeguard for data that risk of becoming pesonal data. Hard to give up this control.

- **Risk of data being reused in a harmful way: example of agri-food scientific communities:**

  - Data on disease outbreaks due to food consumption. Potential harmful use of such data bio-terrorism purposes

EUDAT

Main challenges –
What are the legal implications?

- **lack of knowledge on legal aspects of data sharing and data reuse, in particular IPR, copyrights and licences**

  - results in researchers, research administration staff overly cautious since they are afraid to do mistakes.

  - or realises the importance too late:

    "…Sometimes researchers do their research first and then start asking questions about legal issues – and they realize that they cannot share the results of their work". Thierry Chanier, COMERE project

Main challenges –
"Your data is my data"

- **many researchers significant time and effort spend collecting "hard to get data"**

"Data hoarding is pretty common, people feel a very strong ownership of 'their' data in the medical community. It is understandable, since in this field it takes 10-20 years to assemble good quality data on human subjects as they move through their lives. But from bioinformatics point of view it is a challenge to collaborate due to this. Data sharing policies are sometimes restrictive in the institutes but the researchers themselves also guard the data sets very closely". Ugis Sarkans, diXa

- **Sometimes major financial investments in instruments required to collect the data (astrophysics)**

   -Embargo periods important!

The way forward

- Communities expressed high expectations of initiatives like EUDAT to address these challenges

- They see the need for a cross-disciplinary perspective!

- Suggestions for policy development, guidance, training and coordination efforts, and implementation of suitable technical services and solutions.

training, coordination and policy development

- Want training modules on specific themes such as licensing

- Wish for joint trainings with researchers, legal experts, libraries, research administration etc to get the full picture

    EUDAT to get in touch with different national working groups on legal/ethical issues and open access in order to create synergy between them

EUDAT

training, coordination and policy development

- **Some suggestions:**

  -Much hope put on the new research generation. Get open access mentality in the basic research education. EUDATs role to influence and coordinate how it should be organised on overall national levels

- Gather different data management approaches to design an open model to support heterogeneity of methods for a sustainable common data management over time
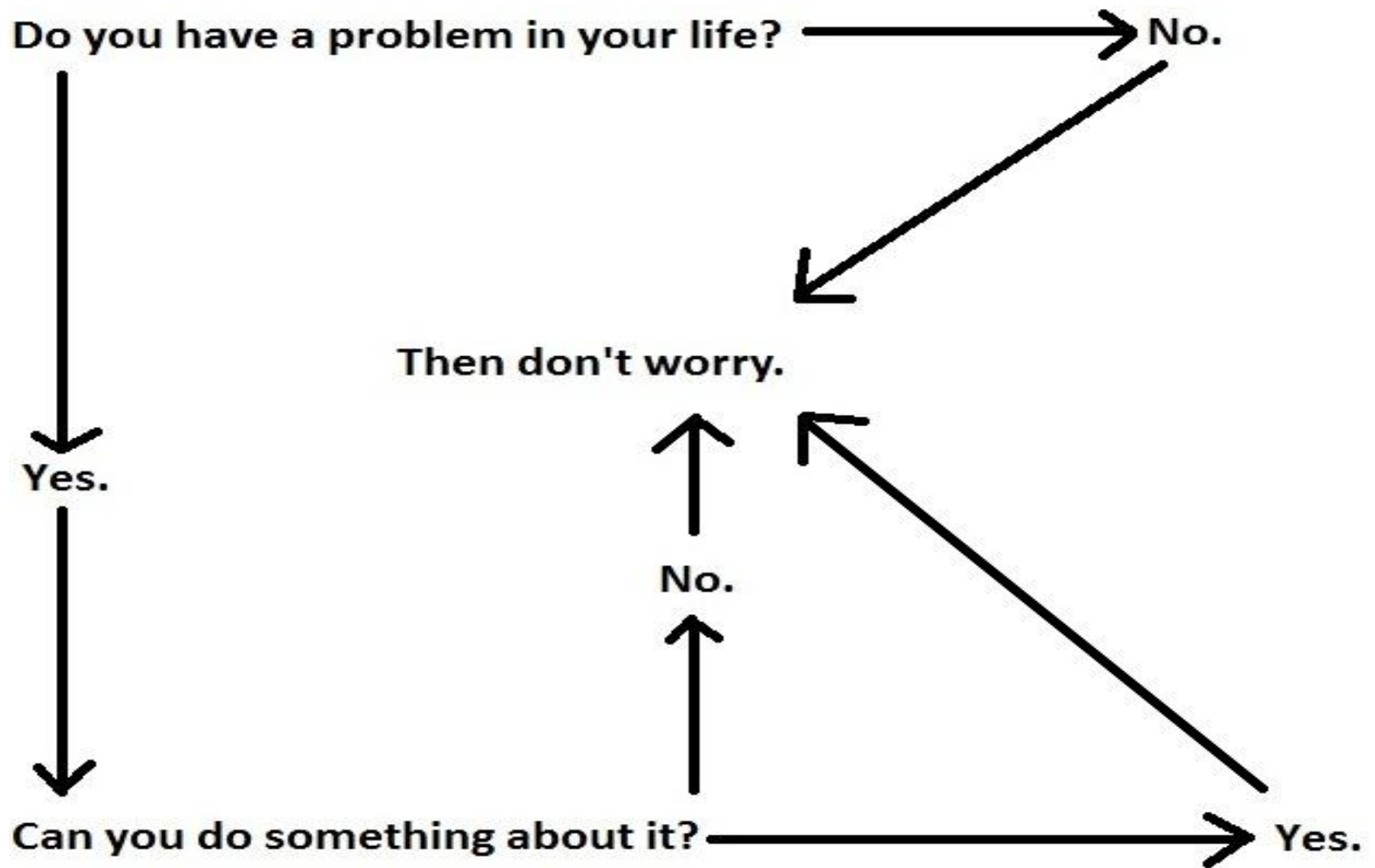
Dont forget the infrastructure!

- **Want the right tools and services to enable policies and, ultimately, the sharing and reuse of data. Sensitive data crucial:**

- To help develop a framework for enabling decision making with respect to sensitive data via recommendations or even tools

- To explore complex licensing scenarios for sensitive data but also metadata – to create some specific scenarios to try various licenses and how they would work under these certain conditions

- To try using PIDs to get access control to sensitive data, such as clinical study data

Some final words

- Open access to research data has only just begun!

- Broad spectra of expectations show that open access to research data needs to underpin the whole system - from strategic planning and polices to the mindset and everyday practice of the individual researcher.

EUDAT

Thank you!
marie.sandberg@csc.fi