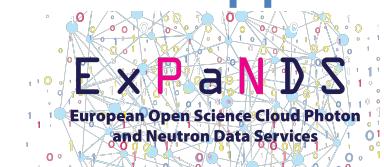


These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641

## ExPaNDS WP2: Enabling FAIR data PaNOSC WP2: Data Policy and Stewardship

- Review and recommend the policies, practises, standards and tools which would develop best practise for FAIR data generation and use in the National Photon and Neutron RIs.
  - In the policies of the RIs
  - In the data-generation, collection and analysis process
  - In Data Management Planning
- Raising awareness and competence in FAIR data of our scientific communities.

To guide services to support FAIRness



# Why is data stewardship important ?

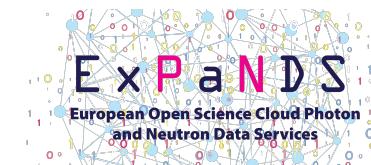
To support Open Science by making data FAIR

to do this

We need to manage our core product → DATA

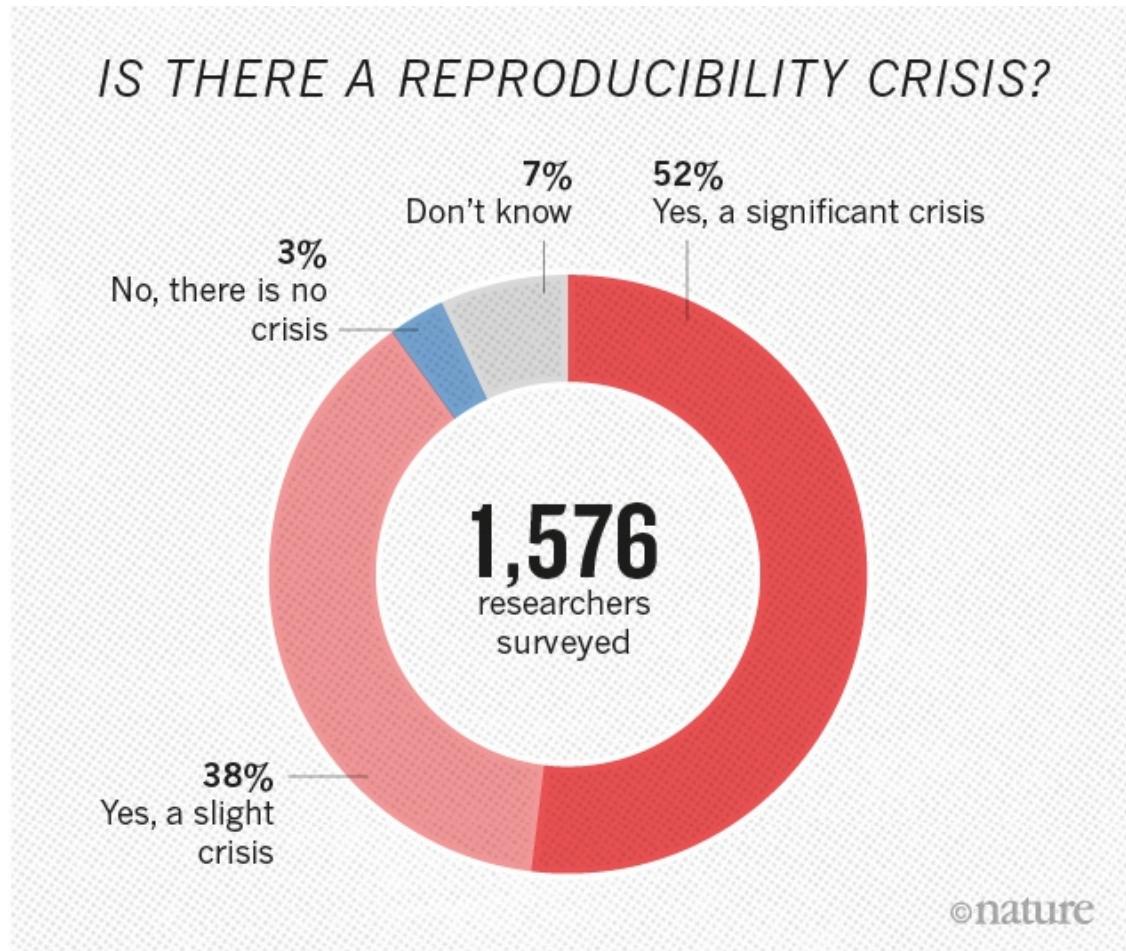


These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641





# To address this ...



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N

NATURE |

1,500

Survey shows

Monya Baker

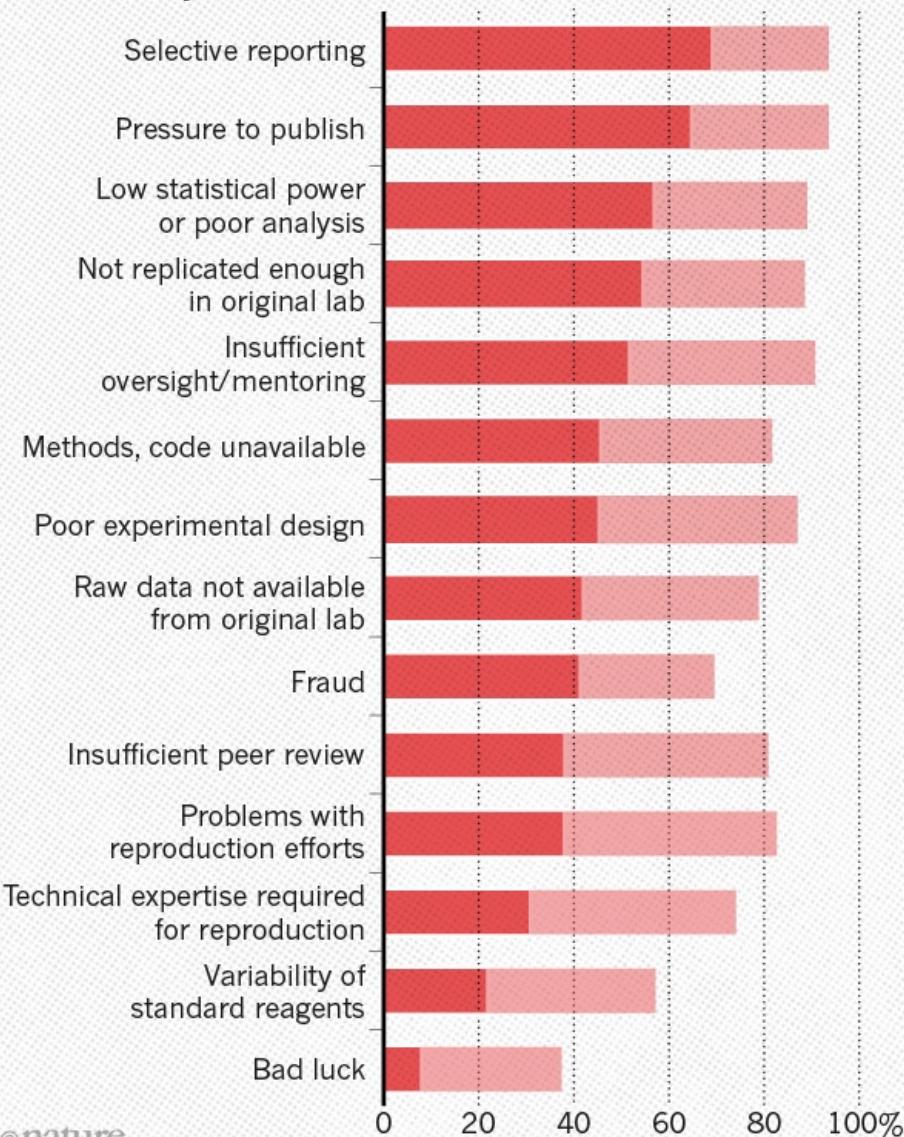
25 May 2015



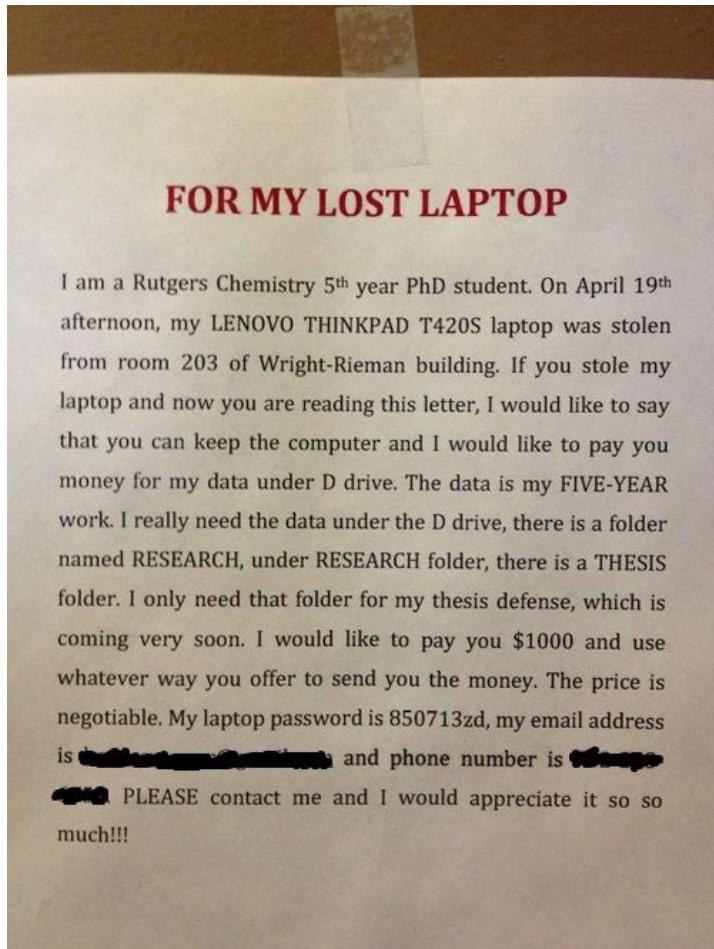
## WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.

- Always/often contribute
- Sometimes contribute



# And help avoid this ...



**CASH REWARD**  
for returning my lost backpack



365 Adventure.com

- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is **VERY** important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on: 07804430054 (ar456@cam.ac.uk)

Thank you!!



# PaNOSC WP2 – FAIR data policy published

The screenshot shows a Zenodo page for the PaNOSC FAIR Research Data Policy framework. At the top, there is a navigation bar with 'zenodo' logo, search bar, upload button, communities dropdown, and an email input for publications@panosc.eu. Below the header, it says 'May 20, 2020' and has buttons for 'Project deliverable' and 'Open Access'. The main title is 'PaNOSC FAIR Research Data Policy framework'. Below the title, it lists authors: Gotz, Andy; Perrin, Jean-Francois; Fangohr, Hans; Salvat, Daniel; Gliksohn, Florian; Markvardsen, Anders; McBirnie, Abigail; Gonzalez-Beltran, Alejandra; Taylor, Jonathan; Matthews, Brian. A text block describes the paper as presenting a new photon and neutron research data policy framework based on the previous PaNDa policy, aligned with FAIR principles, and prepared by the EOSC projects PaNOSC and ExPaNDS. To the left, there is a preview of the PDF document, which includes the PaNOSC logo, the title 'PaNOSC Photon and Neutron Open Science Cloud H2020-INFRAEOSC-04-2018 Grant Agreement Number: 823852', and a table of contents with one entry: 'PaNOSC-D2.1.pdf' (1.1 MB). To the right, there are sections for 'Communities' (Photon and Neutron Open Science Cloud), statistics (586 views, 363 downloads), and indexing (OpenAIRE). Below these are details about the publication date (May 20, 2020), DOI (10.5281/zenodo.3862701), keywords (research data policy, FAIR, PaNOSC, ExPaNDS), grants (European Commission: PaNOSC - Photon and Neutron Open Science Cloud (823852)), communities (Photon and Neutron Open Science Cloud), and license (Creative Commons Attribution 4.0 International).



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



# PaNOSC WP2 – FAIR data policy framework

- FAIR Data Policy framework together with colleagues from ExPaNDS (published on Zenodo)
- **NEXT STEPs**
  - Each PaNOSC site to adapt it to their needs:
    - Case #1 – Institute does not have a DP yet
    - Case #2 – Institute does have a DP
  - Submit changes to Management for adoption directly or submission to governing bodies
- **Question : how far is each institute in this process?**



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



# PaNOSC WP2 – FAIR data policy framework

RDA FDMM was used to analyse how FAIR the data managed by the Data Policy framework is

The PaNOSC Data Policy Framework have been analysed according to the FAIR indicators in the RDA FAIR Data Maturity Model<sup>15</sup>. The results are presented in the table and in the plots below.

FAIR INDICATOR_ID	INDICATORS	PRIORITY	DP METRIC	DP ITEM
F1 RDA-F1-01M	Metadata is identified by a persistent identifier	Essential	4	3.3.1
F1 RDA-F1-01D	Data is identified by a persistent identifier	Essential	4	3.3.1
F1 RDA-F1-02M	Metadata is identified by a universally unique identifier	Essential	4	3.3.1
F1 RDA-F1-02D	Data is identified by a universally unique identifier	Essential	4	3.3.1
F2 RDA-F2-01M	Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard	Important	4	3.2.5
F3 RDA-F3-01M	Metadata includes the identifier for the data	Important	3	IN-9
F4 RDA-F4-01M	Metadata is offered/published/exposed in such a way that it can be harvested and indexed	Essential	4	3.2.16



These projects



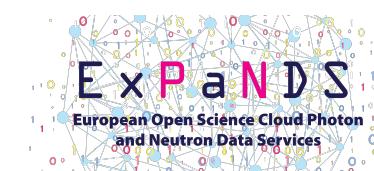
DOI: 10.15497/rda00050

# PaNOSC WP2 – Guidelines + Best practices

- **Guidelines and best practices from sites who already have implemented a Data Policy were gathered and published on GitHub**
- **NEXT STEPs**
  - Convert to a deliverable which could help other sites still in the process of adopting or implementing a Data Policy
- **Question : what are the major challenges being faced by sites who are in this process?**



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



# PaNOSC WP2 – Guidelines + Best practices

- Topics to be covered

- Discussing with scientists + management
- Data catalogue solutions
- Data portal needs
- Long term archiving
- Metadata standards
- Standard techniques
- Electronic logbooks
- Digital Object Identifiers
- ORCID user IDs
- Processed data
- ...



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641

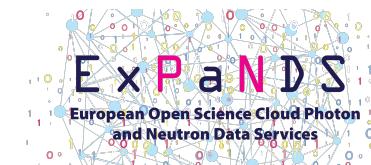


# PaNOSC WP2 – Data Management Plans

- **Data Management Plans** are in the process of gathering needs of each site; closely working with **ExPaNDS** on implementation
- **NEXT STEPs**
  - Finalise requirements for DMPs for all partners (by December 2020)
  - Choose a tool for implementing a prototype
  - Integrate into local proposal systems workflows
- **Participants:** ESS (lead), HZB, ESRF, EuXFEL, CERIC, ELI
- **Question :** which other institutes are interested?



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



# PaNOSC WP2 – Data Management Plans

**Goal:** pre-fill DMP and update it automatically

## DATA DESCRIPTION

- Your beamtime will produce the following types of data (**prefilled**). Can you deal with these?
- Are there any existing data from the ESRF (or elsewhere) that you can reuse?
- Are there any restrictions on the reuse (and subsequent sharing) of third-party data?

## DATA FORMAT

- Your data will be produced in the following formats (**prefilled**). Do you know how to deal with these formats?
- Your metadata will comply with the following metadata standards (**prefilled**). Do you know how to deal with these metadata?
- Is your software capable of dealing with the metadata and data formats your data is provided in? If not how do you propose to deal with these formats in your software?

## DATA VOLUME

- Your beamtime will generate approximately xxx GB/TB/PB volume of data (**prefilled**). Do you know how to deal with this during and after the experiment?
- Do you have sufficient storage or should you include costs for more?
- Will the scale of the data pose challenges when sharing or transferring data between sites?

## DATA QUALITY

- What standards or methods will you use to create the processed data?

## DATA PROCESSING

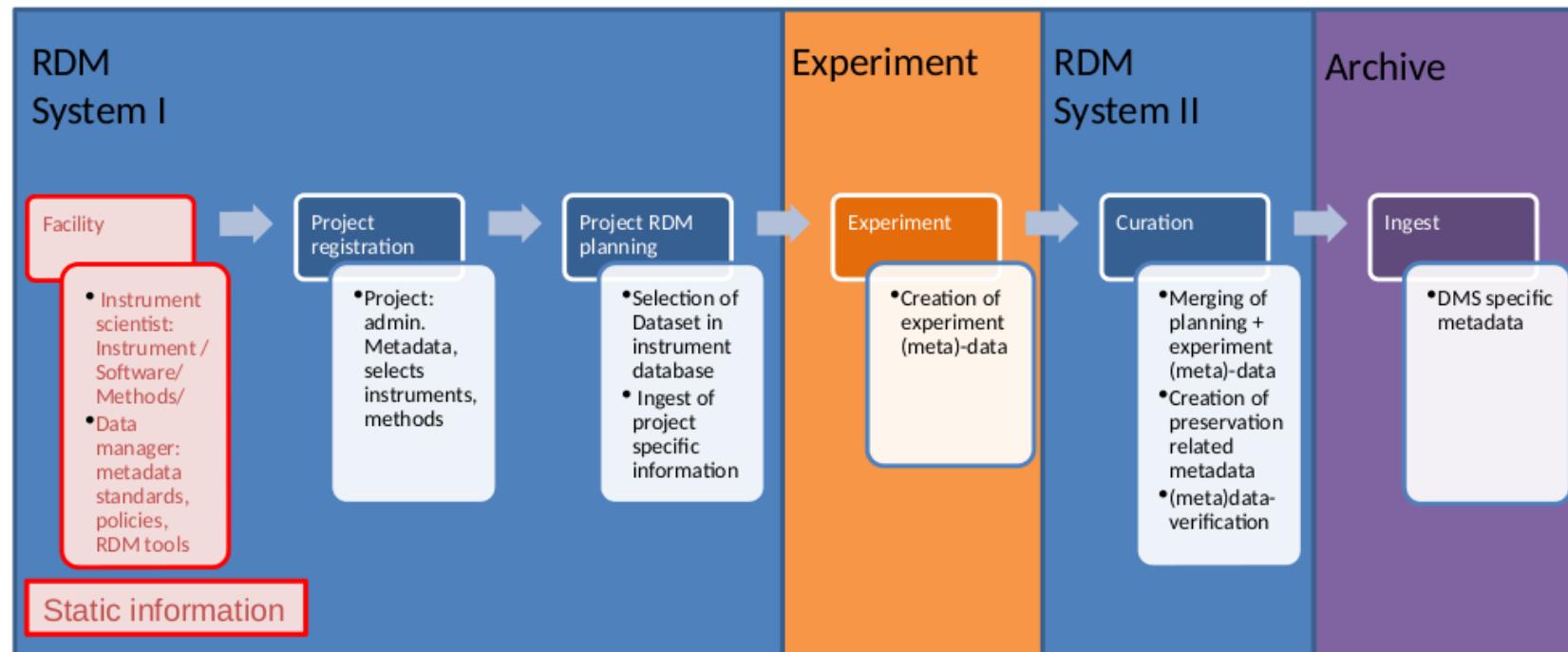


- What are your needs for data processing in terms of CPU/GPU and storage?
- Who will provide these resources?

# PaNOSC WP2 – Data Management Plans



## RDM WORKFLOW AND KNOWLEDGE SOURCES



Credits: Heike Görzig (HZB)

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



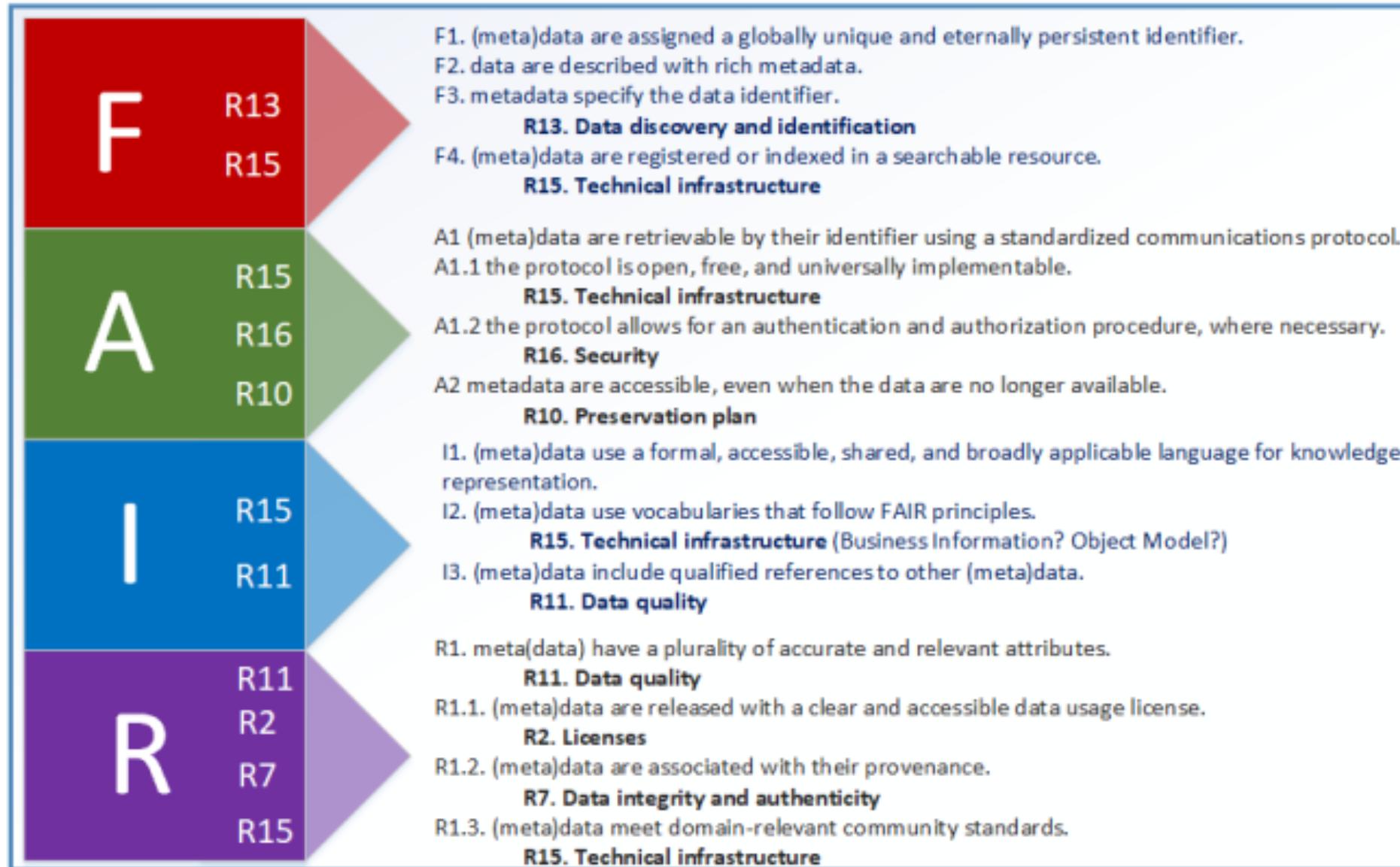


ch to

<https://doi.org/10.5281/zenodo.3828635>



# Data repository certification + FAIR



<https://doi.org/10.5281/zenodo.3734896>



**FAIR\$FAIR**  
Fostering Fair Data Practices in Europe



# PaNOSC WP2+3 - working with standards bodies

- IUCr CommDat committee on data preservation (chaired by J.Helliwell)
- Article on data policy in next IUCr Newsletter
- Need to continue and strengthen our collaboration with the IUCr committees and community standards as **Gold Standard**



ISSN: 2059-7983

Volume 75 | Part 5 | May 2019 | Pages 455-457  
<https://doi.org/10.1107/S2059798319004844>  
OPEN ACCESS  
Cited by 6



Findable Accessible Interoperable Re-usable (FAIR) diffraction data are coming to protein crystallography

These pi

John R. Helliwell,<sup>a</sup> Wladek Minor,<sup>b</sup> Manfred S. Weiss,<sup>c</sup> Elspeth F. Garman,<sup>d</sup> Randy J. Read,<sup>e</sup> Janet Newman,<sup>f</sup> Mark J. van Raaij,<sup>g</sup> Janos Hajdu,<sup>h</sup> and Edward N. Baker<sup>i</sup>



# PaNOSC WP2 KPIs

KPI	Description	Before	30/11/2019
2.1.1	number of instruments on which Data Policy is implemented	53	62
2.1.2	number of techniques / instruments for which metadata are defined	39	41
2.1.25	number of instruments available	135	139
2.1.26	Percentage of techniques for which metadata is defined	113%	114%
2.1.3	number of metadata parameters defined	1310	1,880
2.1.4	number of raw data DOIs	3591	4,612
2.1.5	number of user defined DOIs	5	12
2.1.6	number of downloads	95	194
2.1.7	number of citations of DOIs	242	383
2.1.8	number of datasets	1110	1,584
2.1.9	volume of data archived	5	10
2.1.10	number of datasets cited in publications	237	375

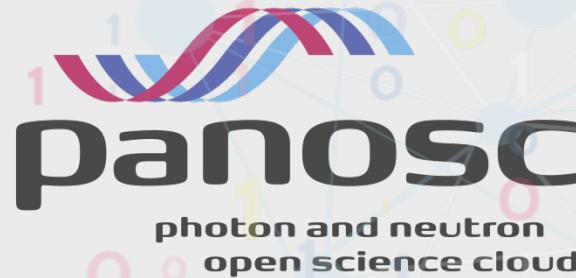
# Questions for discussion

- **What are the major challenges being faced by sites who are implementing a Data Policy?**
- **Who is already using DMPs and/or interested in adopting DMPs for their sites?**
- **What are the major challenges being faced by sites who are in this process?**
- **How can we re-inforce our collaboration with standards bodies and community standards (Gold Standard, CXIDB, PDB, etc.)?**



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641





**E x P a N D S**  
European Open Science Cloud Photon  
and Neutron Data Services

# Thank

Andy Götz & **you** Brian Matthews & WP2 members

[Brian.Matthews@stfc.ac.uk](mailto:Brian.Matthews@stfc.ac.uk) & [andy.gotz@esrf.fr](mailto:andy.gotz@esrf.fr)

<https://expands.eu> & <https://panosc.eu>

Special thanks to:

Jonathan Taylor, Dario Roccella, Rudolf Dimper, Florian Glicksohn, Jean-Francois Perrin, Krzysztof Wrona  
Brian Matthews, Abigail McBirnie, Daniel Salvat, Alejandra Gonzalez-Beltran, Silvie da Graca Ramos



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641