3rd PaN EOSC Symposium @ICRI2022

An Overview

18 October 2022, Brno & online

PaNOSC and ExPaNDS organised the 3rd Photon and Neutron (PaN) EOSC Symposium on sustainable data from PaN facilities as a satellite event to the International Conference of Research Infrastructures (ICRI), held in Brno (CZ) on 18th October 2022.
The symposium, which hosted 12 participants onsite and 32 online attendees, aimed to share the major results achieved in making FAIR data a reality at PaN facilities across Europe, and to explore how a "PaN Data Commons" can be integrated into the EOSC, in collaboration with the other ESFRI cluster projects.

AGENDA

Introduction and presentation of PaNOSC and ExPaNDS results (13:30-14:00)

- Andy Götz, Software Group Leader at ESRF - PaNOSC Project Coordinator
- Patrick Fuhrmann, Head of DESY IT R&I group - ExPaNDS Project Coordinator

Sustainable FAIR Data at High Power Laser Facilities (14:00-14:20)

- Teodor Ivanoaica, Scientific Computation and Data Manager at ELI ERIC

Panel discussion on EOSC Data Federation Sustainability (14:40-15:40)

- Dale Robertson, EGI EOSC Liaison Manager
- Franciska M. G. de Jong, CLARIN ERIC Executive Director - SSHOC
- Andreas Petzold, ENVRI-FAIR Project Coordinator
- Andrew Götz, ESRF - PaNOSC Project Coordinator
- Niklas Blomberg, ELIXIR Project Coordinator - EOSC Life
- Giovanni Lamanna, Director of the "Laboratoire d'Annecy de Physique des Particules"- ESCAPE Project Coordinator

Panel discussion on sustainability models for the PaN Data Commons (16:00-17:00)

- Robert McGreevy, LENS Chair
- Mirjam van Daalen, LEAPS Representative
- Allen Weeks, ELI ERIC Director General
- Florian Gliksohn, ELI ERIC Executive Director
- Andy Götz (ESRF)
- Patrick Fuhrmann (DESY)

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

Overview

The two projects' coordinators, **Andy Götz** (PaNOSC) and **Patrick Fuhrmann** (ExPaNDS) opened the event by introducing to the audience to PaN facilities and the role of the two projects in preparing the path towards FAIR data for both PaN research infrastructures (RIs) managers and staff scientists, as well as current and future users.

The coordinators explained how, to reach the goal of making FAIR data a reality at PaN facilities in Europe, consultations took place before updating and publishing a FAIR data policy framework for all PaN facilities to update their policies accordingly. Such a framework implies open metadata and data, the adoption of global unique persistent identifiers (PIDs) for data and instruments, a common ontology for PaN techniques for the use in catalogues and datasets, open access protocols, human and machine-readable access to data and metadata, community standards for contextual metadata and standard file formats. The projects have jointly developed active data management plans (DMPs) for PaN RIs and their users to effectively manage data throughout the whole data lifecycle. The process towards the adoption of a FAIR data policy at PaN facilities was showcased. This included a previous FAIR assessment on the basis of a Guidebook provided to each facility, and specifically to the people in charge of each instrument, who have been properly trained on how to assess FAIRness of their own beamlines' data-taking process.

The services developed throughout the period of implementation of the two projects were further presented, spanning the e-learning platform and training catalogue, a single AAI (Umbrella ID) enabling users to login to multiple applications and websites with one single set of credentials, the federated search API for PaN data catalogues, an Open Data Portal for searching and downloading data, a common protocol for harvesting data and metadata to make public datasets available to third-party EOSC cross-discipline repositories, the Virtual Infrastructure for Scientific Analysis – VISA, allowing access to data, software and services for data analysis and simulation using remote desktops and/or Jupyter notebooks.

All this has been pursued having in mind the projects' ultimate goal: Open Science and FAIR data, which accelerates scientific findings, as it allows verifiable results, allows to easily find relevant data, reduces the download-data barrier, allow to re-use data from other groups, or to find data not yet analysed or published by the original authors.

The coordinators concluded that the two projects have paved the way forward to making PaN FAIR data a reality and have helped spread the FAIR principles to the PaN community, which will keep on contributing FAIR open data and data services to the EOSC.

Download the presentation by Patrick Fuhrmann
https://bit.ly/3rd-PaN-EOSC-symposium-P-Fuhrmann

Download the presentation by Andy Götz
https://bit.ly/3rd-PaN-EOSC-symposium-A-Gotz

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

FAIR at ELI-ERIC

As a case study of FAIRification of a RI which has very recently started its operations, the Senior Coordinator for Scientific Computation and Data Management - **Teodor Ivanoaica** presented ELI ERIC, its access modes to data and infrastructure, and the work carried out in the framework of PaNOSC, towards the adoption of FAIR data standards and approach to data management.
ELI ERIC, which launched its first open call for users earlier this summer, has adopted the PaN FAIR data policy and active DMPs, and is in the process of implementation of the FAIR assessment and of a PID framework, and engagement with scientists is ongoing towards standardisation of metadata with Nexus/HDF5. The federated search API will have the first data available before the end of the project, and the Open Data portal is expected to be integrated shortly.
ELI ERIC allows access to its services via the AAI Umbrella ID, as well as via ORCID. JupyterLab, Jupyter notebooks and Nexus/HDF5 files visualisation are still not yet in place, but ELI staff are working on having these services implemented by the end of the project.
Last refinements are being introduced to allow remote data analysis with VISA.
ELI ERIC has actively contributed in providing material for the e-learning platform and training catalogue.
The next steps then imply further coordination to integrate tools and services, and designing the unified missing layers for the data management and operations.
Teodor also presented the main steps to be taken towards FAIR data sustainability, which can be summarised as follows:

- FAIR methodology must be backed by the right processes and infrastructure, deployed wherever an organisation's data resides;
- Deploy and integrate a data catalogue;
- Ensure observability and improve standards, to know and understand how your data is used;
- Implement security, governance and lineage, enforcing proper access controls in a consistent, repeatable and explainable fashion;
- Promote preservation of data along with all relevant metadata.

Teodor highlighted the importance of encouraging scientists to promote FAIR data standards and knowledge in data management, to support the RI in curating the data through DMPs, and to integrate with other scientific communities to get more analytic insights.

Finally, for sustainability, he concluded that sharing data is key for science advancements, and that the laser community is key to enabling FAIR data for laser science. Users should be engaged to share their know-how about data management and software development, and services should be common and consistent across the entire community.

Download the presentation by Teodor Ivanoaica
https://bit.ly/3rd-PaN-EOSC-symposium-T-Ivanoaica

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

EOSC Data Federation

The event continued with the first panel discussion on EOSC Data Federation (EDF) sustainability, in which representatives from the EOSC Association and the ESFRI cluster projects discussed the various possibilities to make the services developed throughout the projects financially sustainable in the long-term, also taking into account the outcomes of the discussion on the topic by the members of the EOSC task force on financial sustainability.

EOSC Task Force - Dale Robertson (EGI)

Dale Robertson, co-chair of the task force (TF), moderated the session, and introduced the discussions held within the TF on the risks of duplication of efforts, on who is providing data and services to EOSC, who is responsible for data curation and maintenance, legal and ethical aspects, and what the architecture of the Data Federation could be. She then presented the possible architecture models for the EDF, which have been identified:

- **Overlay**

Intermediate (software) layer, centrally managed and maintained by EOSC, ensuring data interoperability to the highest degree possible, and built "on top of" (i.e., in addition to) any data federation layer which already exists, e.g., in a specific scientific domain or geographical area.

- **Metadata catalogue**

A system to track metadata organised into a catalogue, achieving data interoperability by having a coherent and consistent metadata classification scheme for diverse types of data.

- **Catalogue of data providers and platforms**

A list of all data providers and the platforms they use, indicating how to access single or combined data sources depending on their scientific discipline or geographical area.

- **Continuing with the existing landscape, towards natural evolution**

Existing data infrastructures remain as the go-to sources for researchers. Convergence may be slow and uncoordinated, and a siloed data landscape persists.

In terms of costs related to the data federation, the TF identified categories of additional costs arising when an RI attempts to federate data. These relate to:

- Making data FAIR
- Making experiments reproducible
- Ensuring long-term access to data
- Federating data to EOSC

It is clear that different architecture models entail different costs and benefits, and different distributions of these costs. This aspect will be further discussed in the TF. At the moment, the issues raised by data federation include:

- Duplication
- Future implications of EOSC Data Federation for existing data portals
- Interoperability and collaboration

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

Following the introduction by Dale Robertson, invited panellists introduced their clusters and replied to the following questions:

- **What resources do you envisage that your cluster will contribute to the EOSC Data Federation?** (What is required to achieve this and what is your cluster strategy for sustainability?).
- **What do you think the EOSC Data Federation will "look like", i.e., what model should it have?** (e.g., metadata catalogue, catalogue of data providers and platforms, "overlay" additional layer, or left to form by "natural evolution"?).

SSHOC - Prof. Dr. Franciska de Jong (CLARIN ERIC)

Prof. Dr. Franciska de Jong, executive director of CLARIN ERIC, represented the SSHOC (Social Sciences and Humanities Resources Open Cluster) project, which provided support for research based on cultural data, language data, survey data, and other relevant digital objects, with the goal of federating distributed SSH resources, such as data, tools, publications and training material.
The main achievement of the SSHOC project has been the delivery of the SSH Open Marketplace, a common discovery portal included in the EOSC portal.
Among SSHOC contributions to the EOSC Data Federation are a federated service model for distributed data in diverse but interoperable certified repositories. Data span multilingual, multimedia and heterogeneous data.
The service is expected to be sustained with resources at individual RIs, project funding, as well as through an MoU.
From the SHHOC point of view, EDF is expected to facilitate cross-disciplinary initiatives, leveraging existing strengths at the level of disciplines and clusters; to allow for quality ensuring mechanisms at the level of communities, drawing on existing models and requirements for quality; and to guide, but not to shape thematic initiatives.
Prof. de Jong concluded by mentioning that, for SSHOC, EDF is conceived as a catalogue of data providers and discovery services, but with room for adaptive evolution.

The question arose, on whether discussions at the national level have been taking place to ensure that the architecture model foreseen by the clusters is compatible with what already is planned nationally. Andy Götz replied that national institutions are not in favour of storing data for users and facilities outside their boundaries. This is a concrete problem, in particular for European facilities with members and partners from different countries, as well as for national ones offering services for the international scientific community. Prof. de Jong replied that if a country invests in national data infrastructure, then by being part of European initiatives, including the EOSC, they may contribute to leverage funding.

PaNOSC - Andy Götz

Andy Götz, project coordinator of the PaNOSC project, introduced the audience to the PaN cluster, also showcasing examples of applied research in the fields of climate, energy, geosciences, astronomy, life sciences, cultural heritage and more, and highlighted that data curation has become

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

a new requirement for a number of PaN facilities, which produce petabytes of raw and processed data, and metadata.
To sustain the services provided by the project, such as the PaN data portal, FAIR data must be integrated in the strategy of the PaN facilities, so that data curation and metadata catalogues are sustained by the facilities as a service to the users and to Open Science. The PaN cluster is also planning to further consolidate the federated data portal through the upcoming INFRAEOSC-01-01 call, whereas additional funding from EU and other projects will certainly help financing data curation for specific domains. Furthermore, increased impact of metrics and data citation will certainly promote FAIR data and therefore sustainability.
Andy mentioned that domain specific catalogues, such as Protein Data Bank, COVID-19 portal, Genome portal, and more, are the most successful and should stay with the clusters, since there are the people able to understand and curate the data. The EDF should thus be a way for searching a collection of domain-specific catalogues, and the Data Spaces could be the place where data is federated. Finally, the EOSC WG could define standard APIs for searching federated catalogues which go beyond the existing ones like OAH-OMI. He noted that the EOSC Marketplace is currently not yet adapted for user communities sharing and looking for data, who would rather use services such as Google dataset search.

ENVRI-FAIR - Andreas Petzold

Andreas Petzold is coordinator of the ENVRI-FAIR project, which collects and combines all the infrastructures performing observations in the sub-domains of the earth system (water, earth, air, biodiversity). All RIs in ENVRI-FAIR are running and providing data independently to their user communities. The challenge of the project was to bring these together without disturbing their daily operations, through an integrated access portal to all ENVRI assets (the ENVRI-Hub) to serve as the gateway and collaboration portal for the entire ENVRI community, while maintaining access to the various sub-domain specific portals.
The ENVRI-FAIR catalogue has been designed as the machine-readable gateway to the EOSC, and the Hub has been designed as the human access portal to the assets of the ENVRI community.
ENVRI-FAIR will contribute to the EOSC and EDF via the ENVRI-Hub, including access to data and services via metadata search. Science use cases and virtual research environments provided to the scientists are the main working tool, which is key for interoperability.
To make ENVRI-FAIR's contributions sustainable, FAIR data and RI services are sustained by the ERICs and the associations running the RIs, and by the operational funding by member states. An MoU for collaboration among members of the ENVRI community is also currently in preparation for the cluster's sustainability.
The ENVRI-FAIR Catalogue is designed as a metadata catalogue, and the ENVRI-Hub also contains a catalogue of data providers and platforms. This is how the EDF should look from ENVRI's point of view. An "overlay" additional layer may be difficult to serve all requirements, whereas "natural evolution" is possible but would require very careful monitoring.

EOSC-Life - Niklas Blomberg

Niklas Blomberg, coordinator of the EOSC Life project, which gathers the life sciences community with the goal of: publishing data resources in EOSC; providing the policies, guidelines and process

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

for secure and ethical data reuse; populating an ecosystem of innovative life-sciences tools in EOSC; enabling data-driven research in Europe by connecting life-sciences researchers to EOSC via open calls for participation.
EOSC Life has built a federated catalogue similar to other clusters. The EDF vision may be summarised in what has happened in the frame of the BY-COVID, which links data across a number of different domains, and where there has been an effort to mobilise data and make sure that different data types from different RIs are published and catalogued. A set of standards have been defined to connect data and, through an indexing system, made discoverable at different levels for further analysis by the community.
The EDF should build on RIs and existing assets, which are already services running and well supported by national funders. These services may be re-applied, e.g., taking data catalogues from the underlying RI, to have user-driven portals speaking to the needs of different communities, to connect across diverse data types and country borders.
RIs should thus be long-term stewards of the data, and data catalogues within the domain should be connected through strong standards. EU projects would then be considered as important partners in the process.

ESCAPE - Giovanni Lamanna

The panel discussion ended with the contribution by **Giovanni Lamanna**, coordinator of the ESCAPE project, which represents the cluster for astronomy research. The ESCAPE architecture is based on a Data Lake, which is a scalable federated data infrastructure for open science and the interoperability of large volumes of data. Cross-fertilisation has been important to build flexible science platforms to enable the open data analysis tailored by and for each facility, as well as a global one for transversal workflows. For the implementation of the user interfaces - Virtual Observatory, the established FAIR infrastructure for astronomy was linked to EOSC.
The catalogue of scientific data and software, so called Virtual Repository, has been a major component of the data to be curated in ESCAPE. Finally, a Citizen Science Gateway for citizen science on ESCAPE data archives and ESFRI community was developed to strengthen the links between science and society.
For future collaboration, an international collaboration agreement was signed by the partners of the cluster, to consolidate the EOSC services developed and contribute to the European Research Area (ERA).
Different to the other clusters, ESCAPE view on EDF is to follow natural evolution through a matrix architecture of competences. A vertical approach, oriented towards citizens, society and authorities, would imply a constellation of national open platforms for public data (preserving sovereignty and multilingualism); and a horizontal one, oriented to excellence science, implies domain-based science cluster data platforms (linking existing ones and ESFRI-based data platforms to EOSC).

The closing panel discussion was moderated by **Florian Gliksohn**, Associate Director of ELI ERIC, and focused on sustainability models, which may be implemented for PaN data to be FAIR in the long run. Invited panellists included **Gergely Sipos** from EGI, **Robert McGreevy,** Chair of the LENS initiative, **Mirjam van Daalen**, representing the LEAPS initiative, **Patrick Fuhrmann,** ExPaNDS coordinator, and **Andy Götz,** PaNOSC coordinator.

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

PaN Data Commons

Data Commons aggregates data from a wide range of sources into a unified database to make it more accessible and useful (https://datacommons.org).
Is it possible to imagine a PaN Data Commons? What should it be and how should it look like?

- A place to find all data from PaN facilities
- A search engine for finding data
- A link to access data easily
- Data for testing algorithms
- Data for students to learn
- Data for machines to learn
- Domain portals for user communities

These were the inputs initially provided by Andy to start the discussion.

Whereas the Google Dataset search is considered too broad and not functional for specific PaN data search purposes, there are other great examples which may be followed, such as the COVID-19 portal, the ESCAPE Virtual Observatory, and more from the other clusters.
An example of domain-specific open data publishing is the Human Organ Atlas. The petabytes of data stored need to be kept forever not to waste the great amount of knowledge produced by the scientists, “a gold mine of data waiting to be reused”.

The vision of the PaN Data Commons, as clearly outlined by Andy, is to create a common space for PaNOSC and ExPaNDS facilities where petabytes of PaN FAIR data, analysis software, notebooks, workflows, and training material can be Found, Accessed (downloaded and/or executed), Re-Used + Improved, i.e., FAIR.

It would allow REMOTE ACCESS, as it will be accessible remotely while being executed locally (close to the data) or via the EOSC (data needs to be moved). It would also enable and encourage remote users and experiments (as was urgently required in the post-COVID-19 phase and to tackle climate change challenges).

To sustain the PaN Data Commons, three possible options are in place:

- Local implementation – all sites implement a local data repository and the PaNOSC API which supports federated searching of Open Data.
- Centralised implementation – all sites contribute data (and money) to one site, which implements a PaN data repository for Open Data + Open Science.
- Hybrid implementation – some sites implement a local data repository and make Open Data available via the PaNOSC search API, sites without a data repository contribute Open Data (and money) to a centralised site.

From a survey distributed to PaN facilities, the majority opted for local implementation while a few opted for a hybrid one. None voted for a centralised model.

PaNOSC and ExPaNDS have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

Business models for the PaN Data Commons imply different possibilities, spanning funding coming from EU / and national funding agencies (for development only), or collaboration contracts between RIs (for operations), or via agreements including other funding schemes (foundations, EU Data Spaces, etc.), or via a new legal entity, such as a new ERIC.

Starting from Andy's introduction, the discussion started to reply to the questions:

- Why are Data Commons important?
- What will the impact of a PaN Data Commons be?
- How can the PaN Data Commons be sustained?
- What do you envisage the challenges to be for the sustainability models presented?
- What role will LEAPS+LENS play in ensuring FAIR data commons?

LEAPS - Mirjam van Daalen (PSI)

Mirjam van Daalen started replying that Data Commons are well ahead of what has been achieved up to now. We are still federating our services and the next step will certainly be the Data Commons, to make our data much more visible to the public at large, which is very important also given that we are serving a wide variety of scientific communities. With reference to sustainability, PaN is certainly a mature community which has been working for years to provide federated services. The national funding has been much on the infrastructure part of these services, whereas the EU part has been dedicated to the networking and federating such services. This is why such funding should remain to continue working together in this direction, as was done during the PaNOSC and ExPaNDS projects.
With reference to the role of LEAPS in sustaining the PaN Data Commons, Mirjam replied that certainly LEAPS and LENS should work together to make the federated data services sustainable and connect them to the EOSC.

LENS - Robert McGreevy (ISIS)

In his contribution to the discussion, **Robert McGreevy** started pointing out that "we cannot make all data FAIR", for a number of reasons. So, we need to focus on what data we want to make FAIR. Secondly, RI managers need to be aware and in favour of FAIRification of data, actively engaging in these issues, scaling activities and dedicating an adequate amount of human resources to the cause.

EGI - Gergely Sipos

In the view of **Gergely Sipos**, if not all data can be made FAIR, probably the datasets with the highest chance of reusability should be selected for FAIRification, or only a subset of facilities should provide FAIR data.
To address sustainability, he suggested having partnerships with national repositories to store the data, or if too large, only selected use cases for their reuse.
Computing is another major issue, for which EGI could further contribute, by providing the users with the additional layer allowing for reuse and analysis of big volumes of data without the need to download them in their local machines.

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

Patrick Fuhrmann replied to the first set of statements by the panellists, pointing out that, in his opinion, all data should be FAIR. It is a matter of the science and the scientists whether or not to keep the data. Considering storing data is not that expensive, a possibility is to start storing as much (meta)data as possible, automatically, to make it as useful as possible for reuse.

From **Andy**'s perspective, all of the statements are true to a certain extent: in fact, it's true that we must automate the collection of (meta)data, but one of the biggest challenges is the sample metadata, because the sample comes to our facilities before users come to the facilities. We don't have standards today to capture metadata, which allow us to them correctly to interpret what that data represents. Still, we should keep on trying, but it's also true that the challenge is huge.

Robert pointed out that the resources needed to do so should not be underestimated.
As stated previously by Mirjam, certainly the complementarity of resources should be considered.

What are then the major priorities to make the vision of the PaN Data Commons a reality?

Patrick replied that the EC should be clear regulations on what has to be done with data, how long it has to be stored, and then leave to the countries or the communities on how to do this, and should ensure that, in the future, funding is granted under certain minimum requirements.
Andy continued that, the first step would need to be that organisations such as LEAPS and LENS make a strong commitment to FAIR open data. Then, the answer to the question on how that money is distributed would come out on the basis of who has resources for doing the implementation and who can best use that money, or host people hired to do that.

In this respect, **Mirjam** mentioned the recently released LEAPS strategy, which already highlights the commitment to FAIR data, to the EOSC and the PaN Data Commons. Though, the facilities of both LEAPS and LENS are not able to sustain this alone and there should therefore be additional funding. One of the goals now is also to show the benefits of our services to the users, as they would then be the one able to have a voice for attracting more resources.

As mentioned by **Florian**, it is important that the culture of the users and of organisations changes, that rewarding mechanisms are in place for successful FAIR data users, and that the value of FAIR data is showcased, to both the user community and society.

To comment on previous statements and as closing remarks, **Andy** said that, "there is no more waiting, there is no more 'we will do this if…'. For a new facility, the open culture is part of the scientific culture now. You cannot say publicly that you are not going to share your data, or you would be perceived as someone who's not a good scientist, who doesn't follow science best practices. The PaN community should stop saying that 'we will do this if we get money'. Managing and storing data is a smaller challenge than building an accelerator, or doing the latest experiments, best optics, etc. We are talking of a challenge, but a small challenge in respect to the rest of what we do. So, it's really the culture which seems to be blocking still". PaNOSC and ExPaNDS are meant to drive this cultural change.

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).

Comment from IUCr - John Helliwell

Before closing the event, an additional contribution on the data DOI publications in journals came from the audience, and specifically from **Prof. John Helliwell** chairman of the committee on data (CommDat) of the IUCr – the International Union of Crystallography, who mentioned that IUCr has various initiatives related to raw data availability in publications. The IUCr Raw Data Letters is certainly a very important one. Moreover, in the biological area, any new structure or new method in IUCr Journals has to have the raw diffraction data DOI available too. In chemical crystallography, their Commission has decided to require preservation of raw data only when a sample is seen to be of considerable challenge, for example where different people's software might produce different results from the raw diffraction data. In powder diffraction, the jury is still out. Miguel Aranda's (scientific director at ALBA) article in J Appl Cryst lays out nicely the pros and costs of those raw data preservation and the powder diffraction commission has been considering this. Overall, then the demand by the communities is varying. Also, as mentioned by Gergely Sipos, irreproducibility is damaging science. If a medical challenge occurs, such irreproducibility will have a massive impact, but versus a fairly abstruse topic interesting to a small group of researchers the societal concern will be much less. As Robert McGreevy said, making all raw data FAIR is too much to plan, and the demand by the users is varying.

The event ended with Andy thanking Prof. Helliwell and IUCr contribution to the work carried out in PaNOSC and ExPaNDS, and with Florian thanking all panellists, wishing to meet them all at the PaNOSC closing event to be held in Grenoble on 29-30 November.

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 823852 (PaNOSC) and No. 857641 (ExPaNDS).