



PaNOSC Closing Event

Paving the way towards the PaN FAIR Data Commons

29-30 November 2022

Grenoble - France

Title

Crystallography

A user perspective. Vision on FAIR data for PaN science: Users' presentations on the FAIR data experiences

Author:

John R Helliwell

Affiliation:

International Union of Crystallography (IUCr)

Chairman of the Committee on Data of the IUCr & IUCr Representative to CODATA



DATE:

29th November 2022



PaNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 823852



The University of Manchester

Talk Contents

- What's been achieved?
- What is still missing?

- Hints on how to steer a change towards a more FAIR culture among the PaN user community

- **Users' vision on FAIR data compared with Research Infrastructures' achievements (in Europe)**

Firstly, the respective visions:

There is a huge overlap between PaNOSC's Vision and IUCr's

PaNOSC

- Trust in science
- Opportunities linking raw diffraction data to publications
- **Consulting** European PaN Facilities
- Unpublished data
- Modern data rates are a challenge

IUCr

- Trust in science
- Opportunities linking raw diffraction data to our publications
- **Consulting** the IUCr Commissions
- Unpublished data
- Modern data rates are a challenge

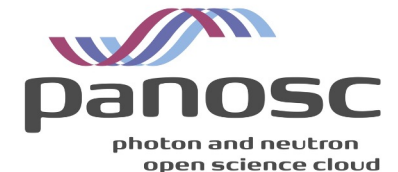
*What's been achieved within the IUCr Global Community?:
The Crystallographic Information Framework#
facilitates trust in crystal structures*

Trust is needed in:

- Data transmission/exchange
 - Crystallographic Information File (1991)
- Data consistency
 - checkCIF for derived (coordinate) data (1998)
 - checkCIF including structure factors (2007)

IUCr COMMITTEE FOR THE MAINTENANCE OF THE CIF STANDARD (COMCIFS)

In 2003 wwPDB Validation started, which IUCr keenly supported;
“Validation Report” as a term was 2010 onwards



What is still missing? or Should we say work still underway:

New opportunities and initiatives stemming from being able to store large quantities of raw data

- *Better understanding of what we do experimentally*
- *Harnessing new methods and software*
- *Enabling new science*
- *Understand the subjective choices made in raw diffraction data processing*

- *The above are in addition to the decades long benefits of:-*
 - *archived coordinates (structure and bonding trends and snapshots of conformational dynamics)*
 - *then processed diffraction data (re-use/re-refinement of a structure based on authors' processed structure factors)*

IUCr Journals has launched IUCrData's Raw Data Letters

raw data letters



IUCrData

ISSN 2414-3146

Received 20 April 2021
Accepted 1 May 2021

Keywords: twinning; diffuse scattering; tetraspanin CD9_{EC2}.



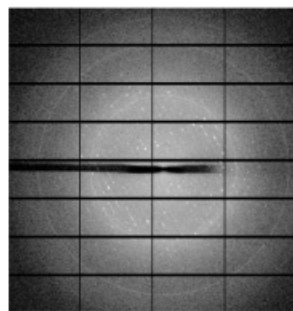
Second extracellular domain of human tetraspanin CD9: twinning and diffuse scattering

Viviana Neviani, Martin Lutz, Wout Oosterheert, Piet Gros and Loes Kroon-Batenburg*

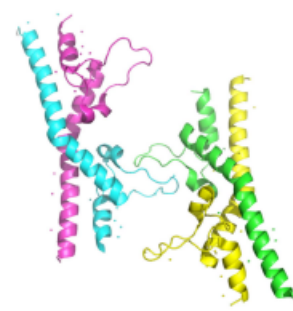
Department of Chemistry, Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, Utrecht, The Netherlands. *Correspondence e-mail: l.m.j.kroon-batenburg@uu.nl

Remarkable features are reported in the diffraction pattern produced by a crystal of tetraspanin CD9_{EC2}, the structure of which was described previously [Oosterheert *et al.* (2020). *Life Sci. Alliance*, 3, e202000883]. CD9_{EC2} crystallized in space group *P1* and was twinned. Concurrent with the twinning, diffuse streaks were seen in the direction perpendicular to the twinning interface. Preliminary conclusions are made on packing disorder and potential implications for the observed molecular structure. It is envisaged that the raw diffraction images could be very useful for methods developers in trying to remove the diffuse scattering to extract accurate Bragg intensities or by using it to model the effect of packing disorder on the molecular structure.

Raw data



Structure



Raw diffraction data

HDF5 data file, DOI: <https://doi.org/10.5281/zenodo.1234567>

Metadata ImgCIF file, DOI: <https://doi.org/10.1107/S2414314622000384/me6134.cif>

checkImgCIF report [CheckCif for Raw Data]

ImgCIF checker version 2022-07-16

Checking block 5886687 in he4557img.cif

Running checks (no image download)

=====

Testing: Required items: PASS

Testing: Data source: PASS

Testing: Axes defined: PASS

Testing: Our limitations: PASS

Testing: Detector translation: PASS

Testing: Scan range: PASS

Testing: All frames present: PASS

All frames present and correct for SCAN1

Testing: Detector surface axes used properly: PASS

Testing: Pixel size and origin described correctly: PASS

Testing: Check calculated beam centre: PASS

Testing: Check principal axis is aligned with X: PASS

Testing presence of archive:

Testing: All archives are accessible: PASS

Running checks with downloaded images

=====

Testing image 4: Image type and dimensions: PASS

Testing image 4: Overloaded values present: PASS

====End of Checks====

Raw data table generated from the CIF

Raw data	
DOI	https://doi.org/10.5281/zenodo.5886687
Data archive	Zenodo
Data format	HDF5
Data collection	
Beamline	Diamond I04
Detector	
Temperature (K)	
Radiation type	Synchrotron X-ray source
Wavelength (Å)	0.979491
Beam centre (mm)	-166.874, 172.497
Detector axis	-Z
Detector distance (mm)	-287.22
Swing angle (°)	
Pixel size (mm)	0.075 × 0.075
No. of pixels	4148 × 4362
No. of scans	1
Exposure time per frame (s)	

Scan axis	ω , X
Start angle, increment per frame (°)	0.0, 0.1
Scan range (°)	360.0
No. of frames	360

IUCr's Raw Data Letters is very grateful to PaNOSC for its help with the checkcif for raw data.

We imagine this will be useful as well for the PaN Facilities to check the reusability of crystallographic raw diffraction images. [The R of FAIR].

Could this be expanded into the domains of imaging and spectroscopy?

The *checkcif for raw data Project Team*, which underpins the *IUCrData Raw Data Letters* initiative:

Loes Kroon-Batenburg (Main Editor of IUCrData's Raw Data Letters), James Hester (ANSTO and Chair of ComCIFS), *Fabio Dall'Antonia, Julian Hörsch (EuroXFEL) and Andy Gotz (ESRF and PANOSC)* and the staff at the IUCr Editorial Office.

Hints on how to steer a change towards a more FAIR culture among the PaN user community?

Well, our IUCr community consultations continue

- Chemical crystallographers (Amy Sarjeant and Simon Coles) organised a Questionnaire and then a Workshop linked to IUCr Prague <https://www.iucr.org/resources/data/commdat/prague-workshop-cx> to examine the question **When should small molecule crystallographers publish their raw diffraction data? Answer: in special cases**
- X-ray powder diffraction has a “policy discussion paper” in J Appl Cryst in 2018 by Miguel Aranda (ALBA Science Director until recently) **Sharing powder diffraction raw data: challenges and benefits**
- **Already implemented:-**
- IUCr Commission on Biological Macromolecules has effected changes in IUCr Journals Notes for Authors that **data processing methods and new structures papers must have their underpinning raw diffraction data doi cited.**

WORKSHOP ON *RAW DIFFRACTION DATA REUSE: THE GOOD, THE BAD AND THE CHALLENGING*

Organized by

Loes Kroon-Batenburg (Netherlands), Selina Storm (Germany), John Helliwell (UK) and Brian McMahon (UK) for the IUCr Committee on Data



TUESDAY AUGUST 22 2023

MELBOURNE, AUSTRALIA

<https://www.iucr.org/resources/data/commdat/melbourne-workshop>

Also at IUCr 2023: Congress Keynote Speaker Andy Gotz ESRF



Acknowledgements

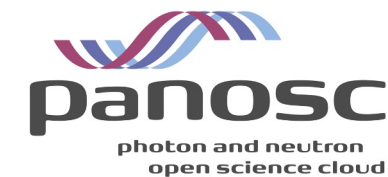
- Members of the IUCr DDDWG 2011-2017 and to the current Members of the IUCr Committee on Data (2017 onwards)
<https://www.iucr.org/iucr/governance/advisory-committees/committee-on-data>
- The CODATA Data Policy Committee
 - <https://codata.org/initiatives/data-policy/international-data-policy-committee/>



The University of Manchester

Thank you

Contact details: john.helliwell@manchester.ac.uk





PaNOSC Closing Event

Paving the way towards the PaN FAIR Data Commons

29-30 November 2022

Grenoble - France

Title

Panel Discussion: Crystallography

A user perspective. Vision on FAIR data for PaN science: Users' presentations on the FAIR data experiences

Author: John R Helliwell

Affiliation: International Union of Crystallography (IUCr)

Chairman of the Committee on Data of the IUCr & IUCr Representative to CODATA

DATE: 29th November 2022



MANCHESTER
1824

The University of Manchester

Treating IUCr as a User then re PaNOSC's question:

What is your users' vision on FAIR data compared with Research Infrastructures' achievements (in Europe)

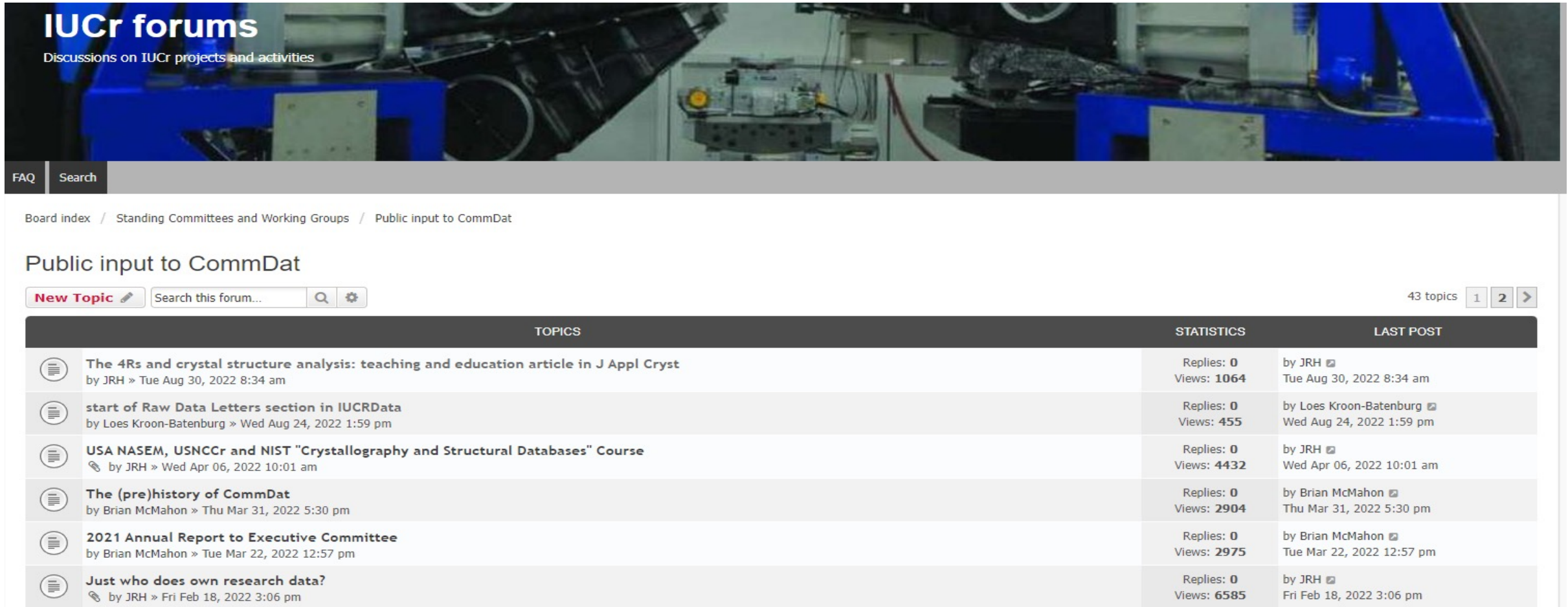
*Our answer would be, **Congratulations to PaNOSC**>>>>>*

We, IUCr, are very impressed with the examples from Europe's facilities of:-

- Diamond Light Source, ESRF and Soleil save all measured data and have a policy committed to release of all raw data after 3 years
- Pioneering from 2018, for 2 years ESRF have generated one DOI per proposal using DataCite (examples: https://search.datacite.org/works?query=10.15151%2F*), users can also create additional DOI per dataset using the ESRF data portal. ESRF asks their users to provide their DOI of the data in their scientific articles.
- In Germany there is the National Forschungsdateninfrastruktur (<https://www.nfdi.de/>) bringing proper data management tools and metadata harvesting to many science areas including the photon and neutron sciences (DAPHNE4NFDI, DATen aus PHotonen und Neutronen Experimenten).

Gathering Hints from the global community

We are very keen on our IUCr Open Discussion Tool:
The IUCr CommDat Forum for Public Inputs on Data

















IUCr forums
Discussions on IUCr projects and activities

FAQ Search

Board index / Standing Committees and Working Groups / Public input to CommDat

Public input to CommDat

[New Topic](#) Search this forum... 43 topics 1 2 >

TOPICS	STATISTICS	LAST POST
 The 4Rs and crystal structure analysis: teaching and education article in J Appl Cryst by JRH » Tue Aug 30, 2022 8:34 am	Replies: 0 Views: 1064	by JRH  Tue Aug 30, 2022 8:34 am
 start of Raw Data Letters section in IUCRData by Loes Kroon-Batenburg » Wed Aug 24, 2022 1:59 pm	Replies: 0 Views: 455	by Loes Kroon-Batenburg  Wed Aug 24, 2022 1:59 pm
 USA NASEM, USNCCr and NIST "Crystallography and Structural Databases" Course  by JRH » Wed Apr 06, 2022 10:01 am	Replies: 0 Views: 4432	by JRH  Wed Apr 06, 2022 10:01 am
 The (pre)history of CommDat by Brian McMahon » Thu Mar 31, 2022 5:30 pm	Replies: 0 Views: 2904	by Brian McMahon  Thu Mar 31, 2022 5:30 pm
 2021 Annual Report to Executive Committee by Brian McMahon » Tue Mar 22, 2022 12:57 pm	Replies: 0 Views: 2975	by Brian McMahon  Tue Mar 22, 2022 12:57 pm
 Just who does own research data?  by JRH » Fri Feb 18, 2022 3:06 pm	Replies: 0 Views: 6585	by JRH  Fri Feb 18, 2022 3:06 pm

Thankyou to Andy Gotz for his PaNOSC's postings at our IUCr CommDat Forum



PaNOSC Closing Event

Paving the way towards the PaN FAIR Data Commons

29-30 November 2022

Grenoble - France

Thank you

Contact details:

john.helliwell@manchester.ac.uk



PaNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 823852

