



| The European Synchrotron



Extending FAIR data management with processed data integration in the ESRF data portal

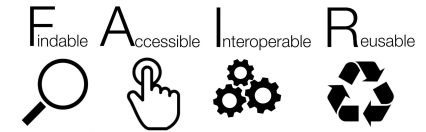
Marjolaine Bodin

Alex de Maria, Maël Gaonach, Guillaume Gaisné

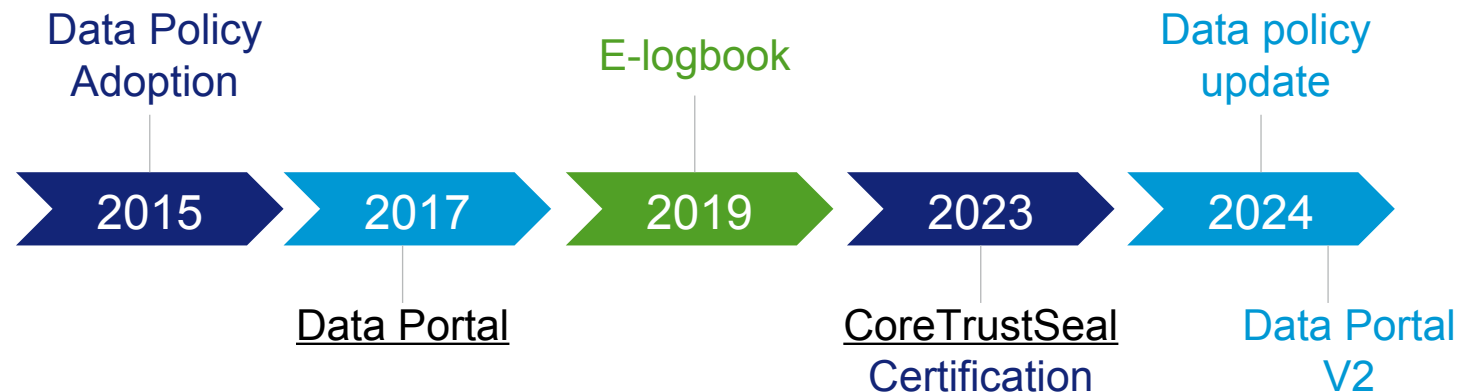
(ESRF)

CONTEXT AND REQUIREMENTS

- Data serves as the foundation of science and is our primary output.
- Data held in archives must remain useful and meaningful into the future.
- The **ESRF** achieves this goal by implementing the **FAIR** principles:



Findable, Accessible, Interoperable, Reusable.



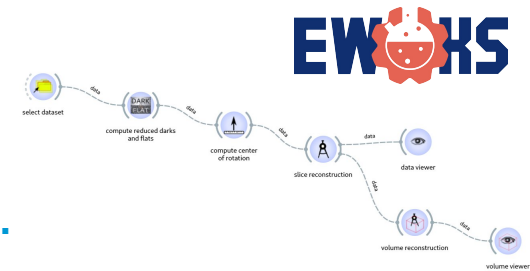
DOI [10.15151/ESRF-DC-1534175008](https://doi.org/10.15151/ESRF-DC-1534175008)

One of the main changes of the **Data policy** update is the inclusion of both processed data and results. To better handle processed data in the data portal, eg. offering different displays depending on the technique/beamline, we changed the portal's architecture to implement a more flexible application:

<https://data2.esrf.fr>

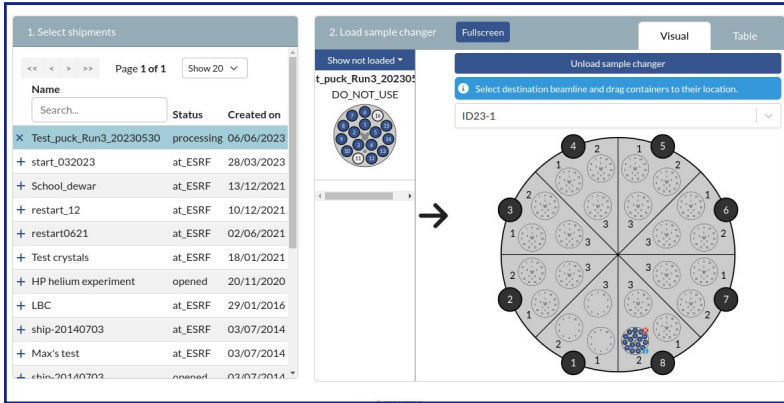
MOTIVATIONS FOR MANAGING PROCESSED DATA

- Processed data has been added to the catalog based on demand, especially to reference data in a publication with a DOI.
- Processed data helps to assess the quality of the data and can help to the fine-tuning of beamtime experiments, leading to more efficient use of beamtime.
- The demand is increasing because:
 - Large amount of raw data
 - Increasing number of samples
 - Development of automatic data analysis
 - Online: **EWOKS*** supports uploading the processed results to the data portal.
 - Offline:
 - Users may want to associate their manually analyzed data with their raw data for publication.
 - Users may want to reprocess data with different parameters.

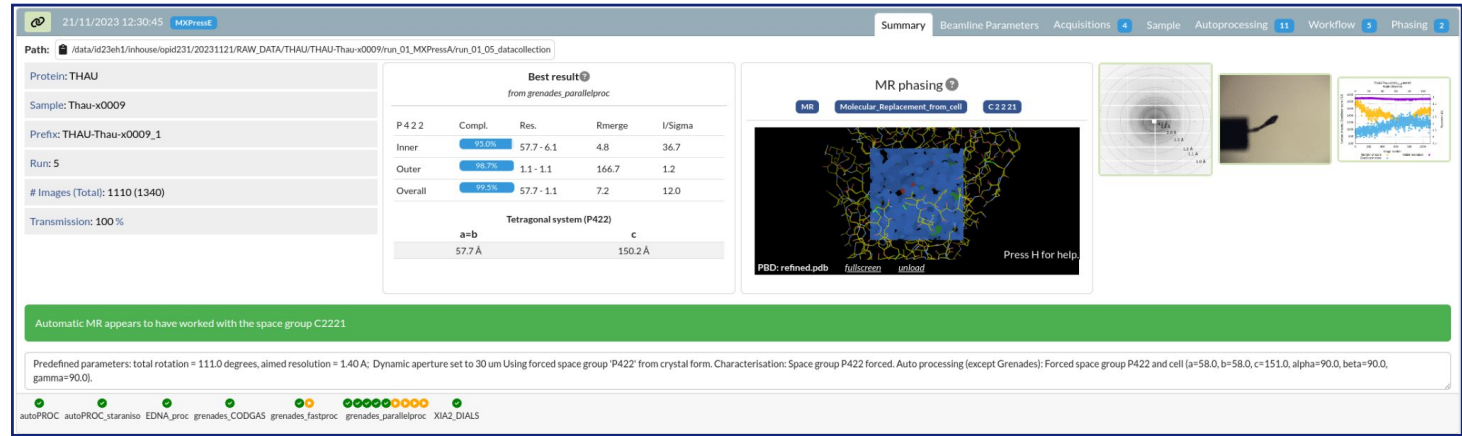


* 16h40 Loic Huder: Automated data processing with EWOKS for ESRF beamlines and users

ISPyB provides users of the Structural Biology (SB) community with a LIMS to manage their data efficiently.

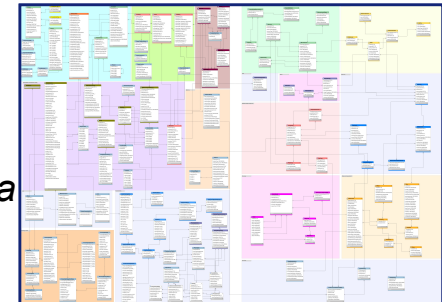


Experiment preparation



Specialized data visualization

ISPyB database schema



It is difficult to extend to other techniques (not flexible architecture).

Meanwhile, the raw data is also available in the Data Portal, without a specific display, but managed according to the data policy and offering other features like e-logbook, DOI minting, etc.

Therefore, managing processed data for all techniques within the Data Portal seems to be the optimal approach.

Supported use cases

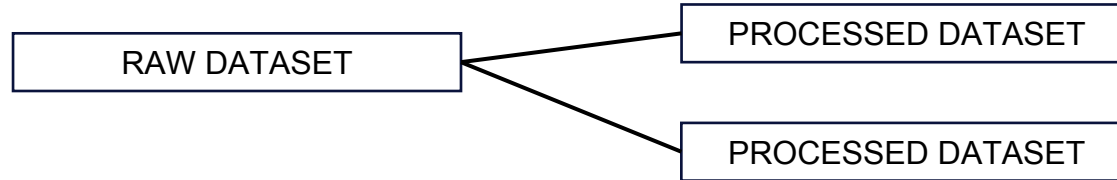
1:1



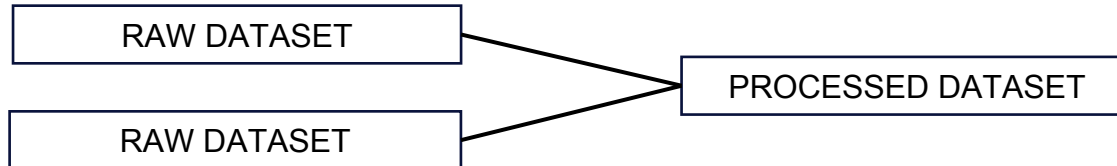
1:1:1



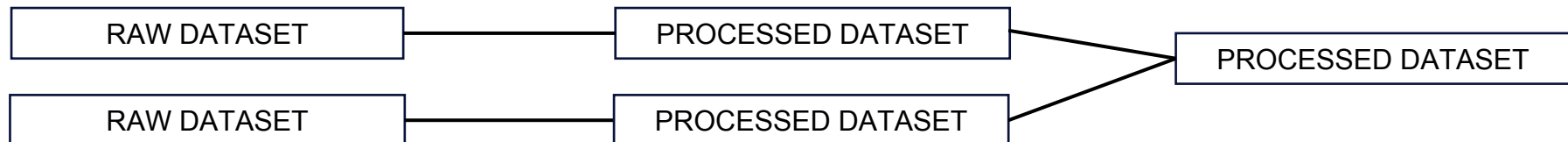
1:N



N:1



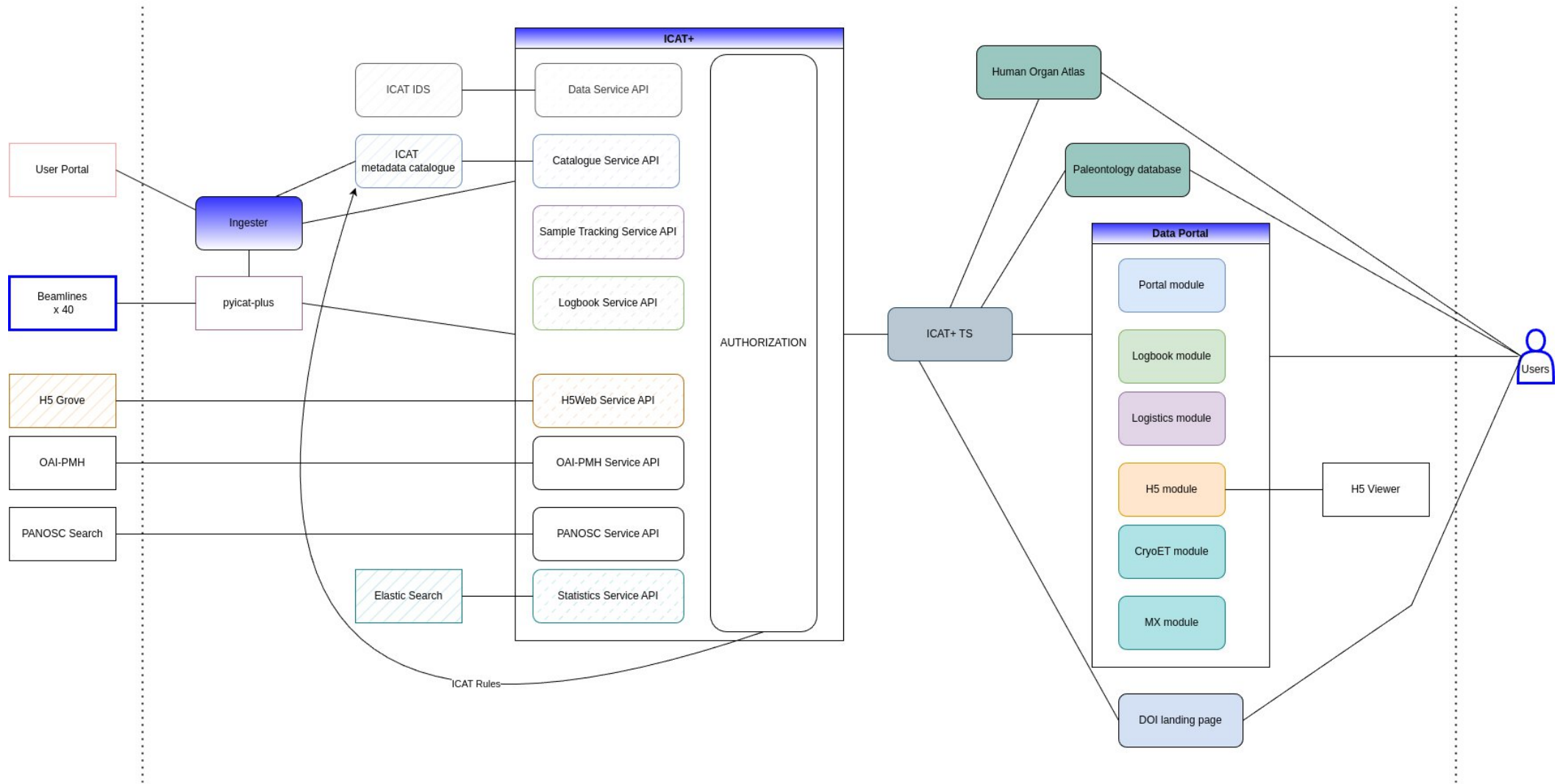
N:N:1



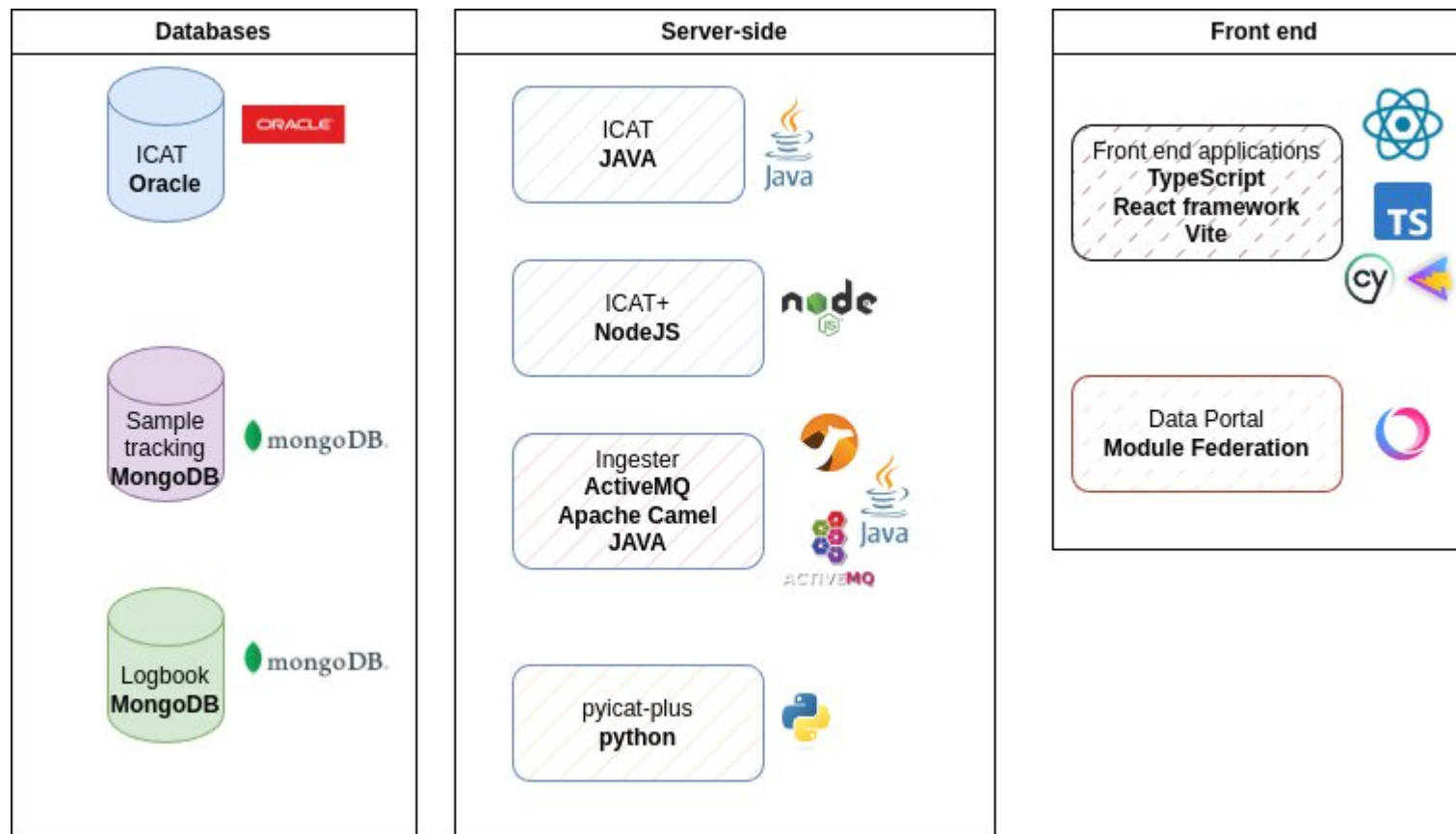
Data Repository for Advancing open sScience

<https://gitlab.esrf.fr/icat/drac>

<https://repo.icatproject.org/site/icat/server/5.0.0/>



TECHNOLOGY STACK



- Modular codebases
- Incremental upgrades
- Independent deployment

Main technology stack used in DRAC

PROCESSED DATA DISPLAY

Data display can be configurable for each beamline and/or technique and/or date.

Summary Files 3 Metadata

Sample ICV_C19B

Description

Technique UNKNOWN

/data/visitor/ma5852/id15a/20231107/ICV_C19B/ICV_C19B_0001 Explore

24/11/2023 15:13:11 Processed ICV_C19B_0001

Sample ICV_C19B

Processed on 24/11/2023 15:13:11

Elapsed time unknown

/data/visitor/ma5852/id15a/20231107/processed/ICV_C19B_0001 Explore Download

Generic approach

ESRF Data Portal Experiments Samples Publications Logistics Beamlines My Jobs 13 Manager

Home / mx2573 CM01 09/10/2023 - 11/10/2023 / Datasets

Sample: mx2573_microtubules_grid1_TOMO Position_56_3

Back to grid

Constant acquisition parameters

grid name : mx2573-grid1 protein acronym : microtubules voltage : 300000 spherical aberration : 2.7 amplitude contrast : 0.1

Show all Filter by angle

Grid

mx2573_microtubules_grid1_TOMO

Datasets

View as: List Summary

12/10/2023 10:38:01 60° mx2573_microtubules_grid1_TOMO/Position_56_3

Raw data

Motion correction

CTF

Dose initial 0.0

Dose per frame 0.48

Images count 6

Magnification 33000

Sampling rate 1.35

Tilt angle 60.0

Average motion 10.3

Frame dose -1.0

Frame range 5

Total dose -1.0

Total motion 51.4

Angle 28.24

Correlation 0.002330

Defocus u 66681.25

Defocus v 66508.30

Estimated b factor -14.66

Resolution limit 14.407

/data/visitor/mx2573/cm01/20231009/RAW_DATA/mx2573_microtubules_grid1_TOMO/Position_56_3/041 Explore Download

12/10/2023 10:37:59 -60° mx2573_microtubules_grid1_TOMO/Position_56_3

Raw data

Motion correction

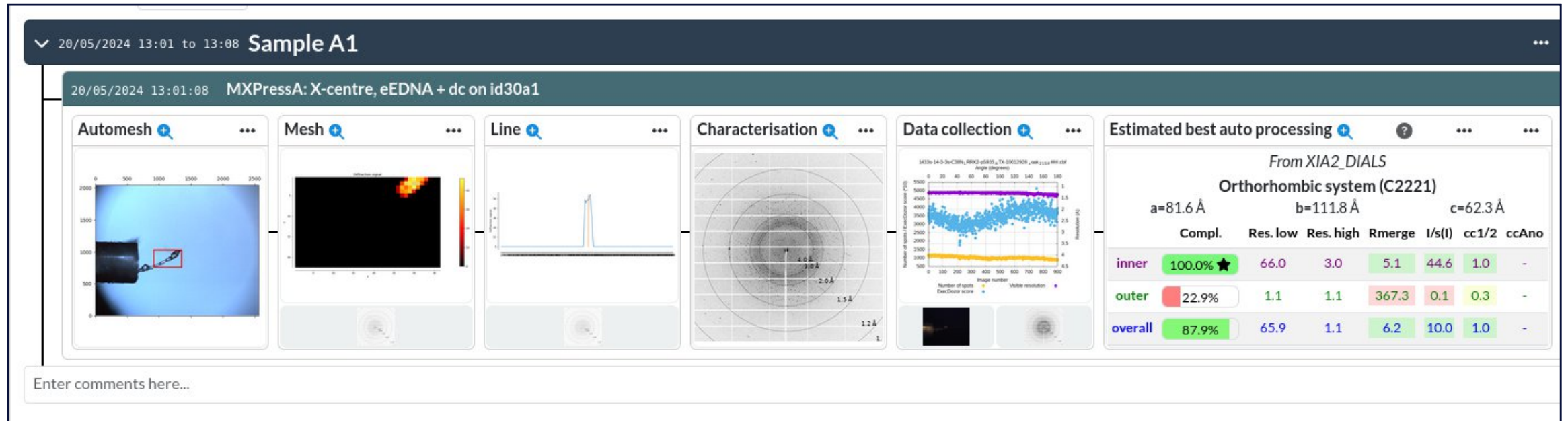
Dose initial 0.0

Average motion 11.8

Cryo-electron tomography (Cryo-ET) example on CM01

MX USE CASE

- Raw and processed datasets displayed together as actions.
- Ranking of best result and display of the best estimation (configurable by the user).

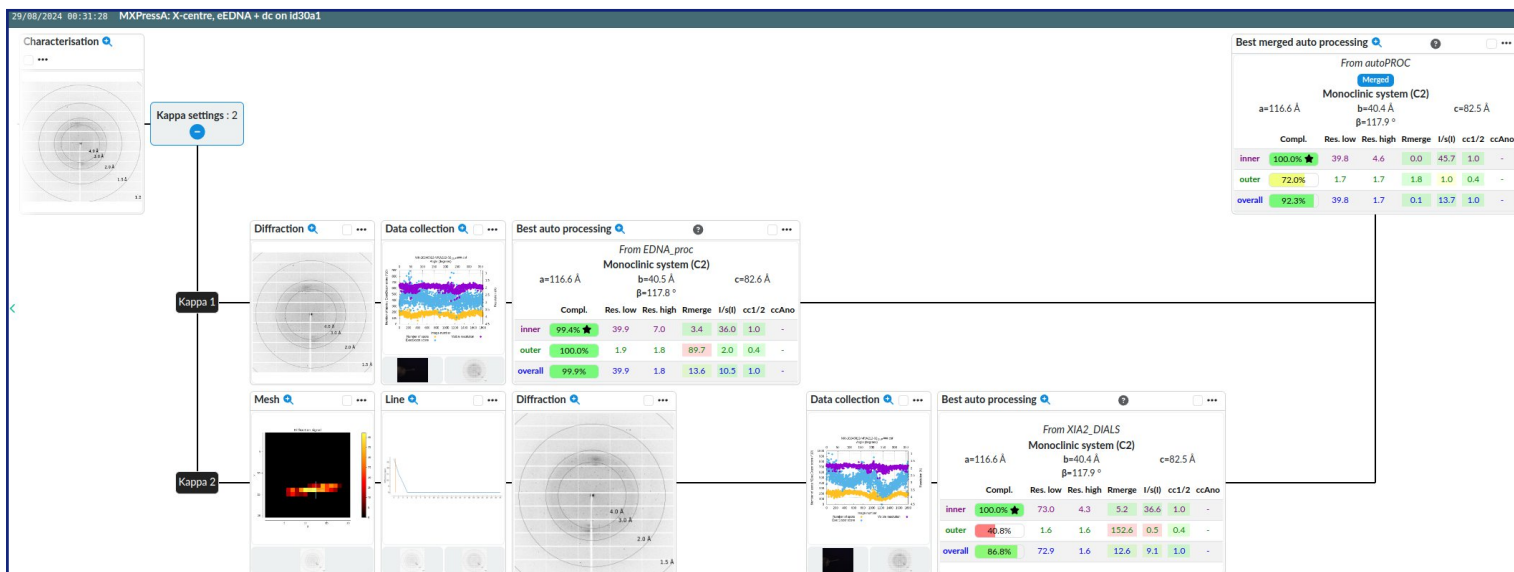


User or workflow “actions”

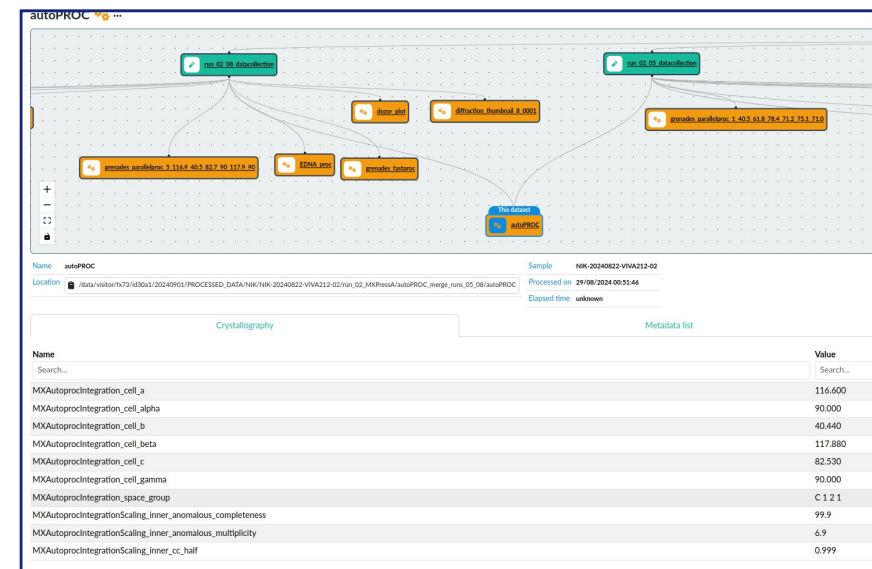
Best estimation

MX USE CASE – MERGE MULTIPLE STRATEGIES INTO A SINGLE RESULT

Flexible display of the results based on the datasets metadata.

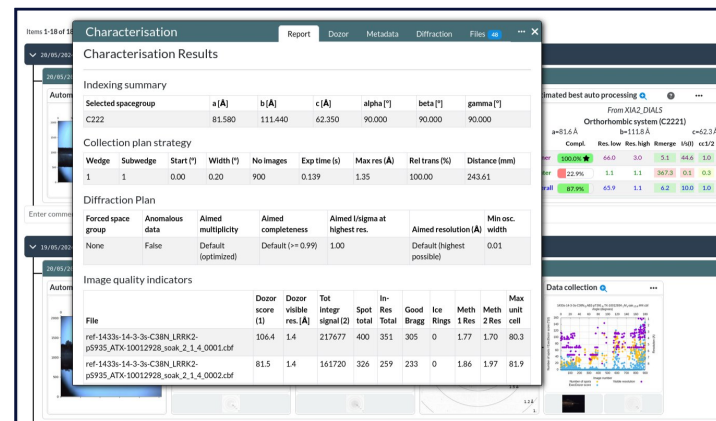


Example of multiple pipelines with merging



Metadata associated to each dataset:

- Key-values in ICAT database
- Files (stored on disk)



Example of metadata for a characterisation dataset

MX USE CASE - SUMMARY

Results available in the Data Portal showing full quality assessment of crystals

MX data list view and summary view

- **EWOKS** workflows can be made available via the data portal.
- Workflows can be launched by the users seamlessly.
- Results appear in the data portal (as a new dataset).
- Currently available only for MX, but it can be generalized.

Open dataset page

✓ Add to selection

Reprocess

MXDataReductionPipelines (ID30A-1)

Download

Best auto processing

From XIA2_DIALS

Orthorhombic system (P212121)

a=76.5 Å b=119.1 Å c=143.0 Å

	Compl.	Res. low	Res. high	Rmerge	I/s(I)	cc1/2	ccAno
inner	100.0% ★	143.2	8.5	5.5	45.4	1.0	-
outer	97.1%	3.2	3.1	354.1	0.2	0.3	-
overall	100.0%	143.0	3.1	15.0	7.9	1.0	-

Reprocess from the data portal

Pipeline

EDNA_proc

autoPROC

XIA2_DIALS

grenades_fastproc

Exclude data range

926-967

Supported format: "50-50, 100-200, 300-400".
You can edit the range in text or select it on the graph below. Hold SHIFT for multi selection.

Excluded images will not be processed by XDS.

Low resolution limit

70.1

High resolution limit

Forced spacegroup

P21

Anomalous

Run

Goals:

- Refine the automatic data processing with manual parameters
- Publication purposes

Electronic logbook

The screenshot shows the ESRF Data Portal e-logbook interface. The top navigation bar includes 'Data Portal', 'Experiments', 'Samples', and 'Publications'. A search bar is present on the right. The left sidebar contains navigation options: 'Experiment session', 'Statistics', 'Datasets', and 'Logbook'. The 'Logbook' section is active, showing a search bar and filter options. The main content area displays a log of events, including file uploads, dataset creation, and machine adjustments. A graph titled 'align_spec01.dat 28.1_err_001' shows intensity vs. energy with multiple colored curves. The log entries are timestamped and include user names like 'Wout DE NOLF' and 'Giulia VERONESI'.

- Timestamped e-logbook
- Experiment vs Beamline logbook
- Notifications from acquisition software (**BLISS**) and ingester
- Notifications from accelerator
- Manual editing via the data portal
- Searchable
- Export as PDF or plain text

IH-LS-3167 Elemental distributions in microalgae

DOI [10.15151/ESRF-ES-132528770](https://doi.org/10.15151/ESRF-ES-132528770)

DATA PORTAL – ADDITIONAL FEATURES – SAMPLE TRACKING

- Unify sample tracking over all beamlines.
- Introduction of a processing plan.

The screenshot shows the 'Data Portal' interface for a specific parcel. The top navigation bar includes 'Data Portal', 'Experiments', 'Samples', 'Publications', 'Logistics', 'Beamlines', 'My selection', 'My Jobs', and 'Manager'. The breadcrumb trail indicates the current location: 'Home / ID23-1-0000 / ID23-1 / 11/11/2023 - 11/11/2023 / Logistics / Parcel'.

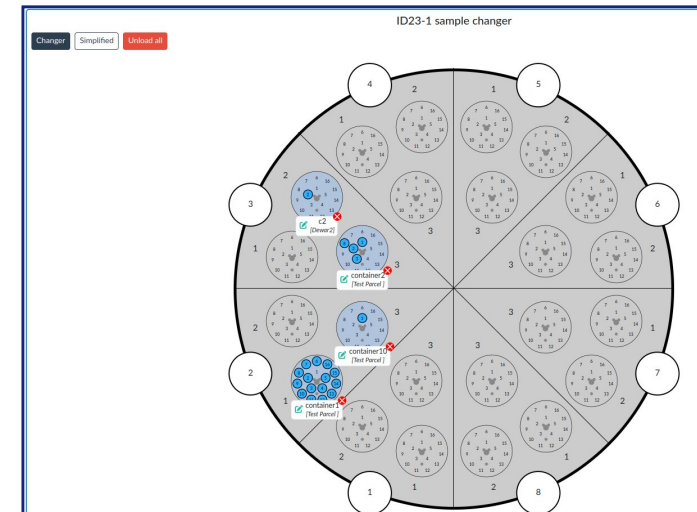
The 'Parcel status' section shows a timeline: CREATED (16/11/2023) → SCHEDULED (04/01/2024) → READY FOR SAFETY CHECK (29/07/2024) → APPROVED (29/07/2024). Below this, there are buttons for 'Mark as SENT', 'STORES', 'BEAMLINE', 'BACK IN STORES', 'SENT BACK TO USER', and 'DESTROYED'.

The 'Content' section is set to 'Containers' and displays a table with the following columns: Container (Type, Name, Sample position), Item (Type, Short name, Name, Description, Comments), Pin barcode, Exp. Type, and Experiment (a, b, c, Alpha, Beta, Gamma, Initial resolution, Radiation-sensitivity, Required complet). The table lists 17 containers, each containing a sample of 'Lysoz' or 'PRP4K'.

Declare samples + experiment plan and processing plan

	Pipeline name	Force S.G	MR	SAD	Reference reflection	Search models	UniProt IDs	Smiles
1	grenades_fastproc	P212121	True					
2	grenades_parallelproc							
3								
4								

Prepare MX Sample Changer



Generate bespoke DOIs for any subset of the data

<https://youtu.be/dPeN855-Mu4>



Clear selection Manage selection


Items 1-1 of 1 Show 50 Bottom

EC-22 (31/01/2009 on ID19)

31/01/2009 00:00:00 Zollikofer et al (2024) 5.06um scans D2854-LR12 Summary Files 3 Metadata

Paleontology database

Related DOI
DOI [10.15151/ESRF-DC-1900353410](https://doi.org/10.15151/ESRF-DC-1900353410)



Sample

Scientific domain	paleoanthropology
Continental region	Europe
Country	Georgia
Localisation	Dmanisi
Formation	-
Era	Cenozoic
Period	Quaternary
Epoch	Pleistocene
Clade 1	Animalia
Clade 2	Chordata
Clade 3	Mammalia
Clade 4	Primates
Clade 5	Hominidae
Clade 6	-
Species	Homo erectus
Material type	Teeth

Additional information

Keyword 1	Paleoanthropology
Keyword 2	Virtual dental paleohistology
Comments	-
Repository Institution	-
Authors	-
Segmentation	-

DOI

Abstract JPEG2000 image stack of 5.06um scan of teeth from D2700/D2735 studied by Zollikofer et al (2024). Sample includes: D2854-LR12

Title D2700/D2735 teeth from Dmanisi, Georgia-5.06um scans

Users Christoph P.E. Zollikofer, Marcia Ponce de Leon

Associated publication

DOI Dental evidence for extended childhood in early Homo from Dmanisi

Ref. Christoph P.E. Zollikofer, Vincent Beyrand, David Lordkipanidze, Paul Tafforeau, Marcia Ponce de Leon

Mint DOI for a subset of datasets

You can only mint a DOI for datasets which are linked to experiment sessions where you are a user.

Minting the DOI makes the datasets public immediately.

Title*

Abstract*

Authors*

+ New author Clear all authors Select an author to add...

Name	Surname	ORCID® ID
Paul	Tafforeau	0000-0002-5962-1683
Name*	Surname*	ORCID® ID

Mint DOI

Data Collection Open access

D2700/D2735 teeth from Dmanisi, Georgia-5.06um scans

Christoph P.E. Zollikofer; Marcia Ponce de Leon

JPEG2000 image stack of 5.06um scan of teeth from D2700/D2735 studied by Zollikofer et al (2024). Sample includes: D2854-LR12

Experimental data
Data can be accessed by clicking on the link below
[Access data](#)

Experimental report
No report was found for EC-22. Proposers and session participants may submit it via the [ESRF User Portal](#).

Reference
Researchers must acknowledge the source of the data and cite its unique identifier as well as any publications linked to the same raw data.
Below is the recommended format for citing this work in a research publication.

Zollikofer, C. P. E., & Ponce de Leon, M. (2024). D2700/D2735 teeth from Dmanisi, Georgia-5.06um scans (Version 1) [Dataset]. European Synchrotron Radiation Facility. doi.org/10.15151/ESRF-DC-1900353410

Cited by
No citations were found.

Metadata

Identifier
DOI [10.15151/ESRF-DC-1900353410](https://doi.org/10.15151/ESRF-DC-1900353410)

Proposal EC-22 Beamline [ID19](#)

Public release year 2024

Publisher [European Synchrotron Radiation Facility](#)

License (for files) [Creative Commons Attribution 4.0](#)

Related DOIs
No related DOIs were found.

DOI [10.15151/ESRF-DC-1900353410](https://doi.org/10.15151/ESRF-DC-1900353410)

DOI landing page example

Human Organ Atlas

Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography (**HiP-CT**). Integrate advanced visualization tool like **Neuroglancer**.

Human Organ Atlas EXPLORE SEARCH TUTORIALS HELP 25 RESULTS

PATIENT

Sex

- male
- female

Age

Height (cm)

Weight (kg)

SAMPLE

Organ

- brain
- heart
- lung
- kidney
- spleen

Pathology

covid, pneumopathy, etc.

SCAN PARAMETERS

Pixel size (um)

Detected avg. energy (keV)

Surface dose rate (Gy/s)

VOI integ. dose (kGy)

2.45um_frontal
LADAF-2020-31 brain 19/06/2024
Vertical column in local tomography at 2.45um voxel size performed by HiP-CT on the beamline BM05 of the brain from the body donor LADAF-2020-31 using half-acquisition protocol. This VOI is a cylinder imaging part of the frontal lobe.

2.45um_hippocampus
LADAF-2020-31 brain 19/06/2024
Vertical column in local tomography at 2.45um voxel size performed by HiP-CT on the beamline BM05 of the brain from the body donor LADAF-2020-31 using half-acquisition protocol. This VOI is a cylinder imaging the hippocampus.

6.05um_hippocampus
LADAF-2020-31 brain 19/06/2024
Vertical column in local tomography at 6.05um pixel size performed by HiP-CT on the beamline BM05 of the brain from the body donor LADAF-2020-31 using half-acquisition protocol. This VOI is a cylinder imaging the hippocampus.

2.256um_VOI-13.1
LADAF-2021-17 brain 19/06/2024
Vertical column in local tomography at 2.256 um voxel size performed by HiP-CT on the beamline BM18 of the brain from the body donor LADAF-2021-17 using half-acquisition protocol.

6.54um_ROI-02
LADAF-2021-17 brain 19/06/2024
Vertical column in local tomography at 6.54 um voxel size performed by HiP-CT on the beamline BM18 of the brain from the body donor LADAF-2021-17 using half-acquisition protocol.

Human Organ Atlas EXPLORE SEARCH TUTORIALS HELP ← BACK TO RESULTS

Complete scan at 25.25 um of the brain of the body donor LADAF-2021-17

Description
Complete scan at 25.25 um performed by HiP-CT on the beamline BM05 of the brain from the body donor LADAF-2021-17 using quarter-acquisition protocol.

DOI [10.15151/ESRF-DC-1773964937](https://doi.org/10.15151/ESRF-DC-1773964937)

Users
C. L. Walsh, J. Brunet, C. Berruyer, E. O'Leary, S. Rahmani, A. Bellier, Y. Boukourt, A. de Maria, M. Bodin, K. Engel, K. Dollman, J. Jacobs, C. Werlein, D. Jonigk, M. Ackermann, P. Tafforeau, P. D. Lee

Technique
Hierarchical Phase-Contrast Tomography (HiP-CT)

Instrument
BM05, ESRF

Visualize in Neuroglancer

Download files with Globus

ZIP 25.25um_LADAF-2021-17_brain_complete-organ_pag-0.01_0.03_jp2 — 31.0 GB

ZIP 50.5um_LADAF-2021-17_brain_complete-organ_pag-0.01_0.03_jp2 — 3.9 GB

ZIP 101.0um_LADAF-2021-17_brain_complete-organ_pag-0.01_0.03_jp2 — 485.6 MB

ZIP 202.0um_LADAF-2021-17_brain_complete-organ_pag-0.01_0.03_jp2 — 60.8 MB

TXT LADAF-2021-17_brain_complete-organ_25.25um_bm05 — 3.0 kB

These datasets are made freely available under the [CC-BY-4.0 licence](https://creativecommons.org/licenses/by/4.0/). Publications using these datasets are required to [cite the DOI](https://creativecommons.org/licenses/by/4.0/).

Source Rendering Annotations
n5://gs://ucl-hip-ct-35468991/esa2e932010444e8358940b/LADAF-2021-17/brain/25.25um_complete-organ_bm05/

Enable default subsample set

uint16 volume

[bounds] default annotations

	Source dimensions		
	Lower	x	y
Upper	0	5280	4686
Scale	25.25um	25.25um	25.25um

Output dimensions

	x	y	z
25.25um	1	0	0
25.25um	0	1	0
25.25um	0	0	1

+

Search (according to the keywords and metadata of the data) and display data in HOA

Paleontology Database

Data obtained from X-ray Computed Tomography experiments. Started in 2011, revamped in 2024.

Paleontology database EXPLORE SEARCH HELP

Domains

- Vertebrate Paleontology
- Invertebrate Paleontology
- Vertebrate Biology
- Paleoanthropology
- Archaeology
- Paleobotany
- Ichnology

Samples

- Cephalopoda
- Amber inclusions
- Doushantuo embryos
- Ammonites
- Ostracodes


Datasets

AMNH55901

Invertebrate paleontology North America

United State of America Mesozoic Cretaceous

3D reconstruction for Baculites sp. AMNH55901. For further details regarding samples, measurements and data analysis, refer to the files and the article.




AMNH66267

Invertebrate paleontology North America

United State of America Mesozoic Cretaceous

3D reconstruction for Baculites sp. AMNH66267. For further details regarding samples, measurements and data analysis, refer to the files and the article.




AMNH66253

Invertebrate paleontology North America

United State of America Mesozoic Cretaceous

3D reconstruction for Baculites sp. AMNH66253. For further details regarding samples, measurements and data analysis, refer to the files and the article.



Paleontology database EXPLORE SEARCH HELP ← BACK TO RESULTS

burrow BP1-5558 Thrinaxodon Broomistega

Description

3D reconstruction for burrow BP-1-5558 Thrinaxodon Broomistega. For further details regarding samples, measurements and data analysis, refer to the files and the article.

Comment: fossil burrow aged of 250 million years from the Karoo basin in South Africa. It contains two skeletons in anatomical connection: the mammalian reptile Thrinaxodon and the amphibian temnospondyl Broomistega

Scientific domain	vertebrate paleontology
Repository institution	Evolutionary Studies Institute, Johannesburg, South Africa
DOI	DOI 10.15151/ESRF-DC-1634314047
Reference	Fernandez, V., Abdala, F., Carlson, K.J., Cook, D.C., Rubidge, B.S., Yates, A. and Tafforeau, P., 2013. Synchrotron reveals Early Triassic odd couple: injured amphibian and aestivating therapsid share burrow. PLoS One, 8(6), p.e64978.
Segmentation	Vincent Fernandez
Instrument	ID17, ESRF

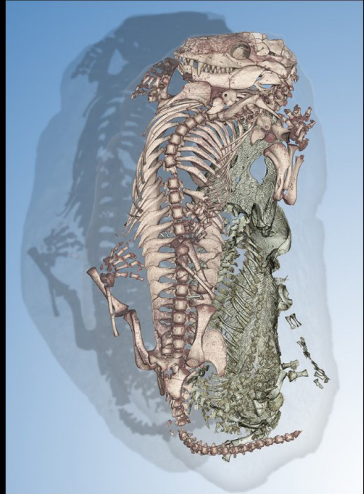
Location

Continental Region	Africa
Country	South Africa
Localisation	Karoo
Formation	unknown

Geological Time

Era	Mesozoic
Period	Triassic
Epoch	Early Triassic

Classification



Download files with Globus

ZIP	org_slices/BP_1_5558_45.5um_16bits_5500_6221	— 10.5 GB	↓
ZIP	org_slices/BP_1_5558_45.5um_16bits_0000_0499	— 7.3 GB	↓
ZIP	org_slices/BP_1_5558_45.5um_16bits_0500_0999	— 7.2 GB	↓
ZIP	org_slices/BP_1_5558_45.5um_16bits_5000_5499	— 7.1 GB	↓
ZIP	org_slices/BP_1_5558_45.5um_16bits_1000_1499	— 7.1 GB	↓

Explore and display data in paleontology database. Geographical, geological, temporal, and taxonomical keywords to facilitate searches

Conclusion

- A robust architecture has been implemented to handle multiple types of processed data.
- Handling processed data enhances the FAIRness of the data: structured metadata, full transparency of data provenance, clear documentation of processing steps.
- Real-time assessment of data quality increases productivity and ensures efficient use of beamtime: enabling rapid adjustments to the experimental setup and prioritizing samples or settings that provide the most valuable data.

Next steps

- Integrate more techniques (SSX, TOMO, etc.). Linked to the ongoing implementation of the **PANET ontology***
- Improve processing plan
- **Development** of a new dedicated portal for Cultural Heritage.
- Keep track of data processing and analysis in the e-logbook.

* 14h10 Wout De Nolf: Ontological definition of experimental techniques for FAIR data



OSCAR
Open Science, Culture, Action
for Research & Society

ESRF Data management team

Alex de Maria

Marjolaine Bodin

Maël Gaonach

Guillaume Gaisné

Andy Götz



Thank you for your attention!
Questions?