



User and Data Journey, Fair Data and Data Curation

Let's talk about data

SWAP team (Scientific Web Applications)

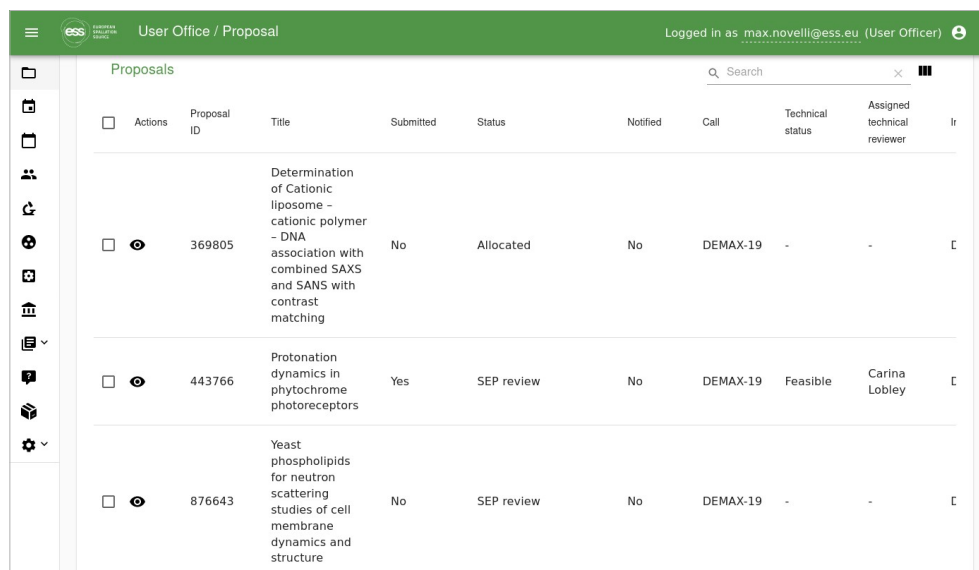
PRESENTED BY MASSIMILIANO NOVELLI, DATA CURATION SCIENTIST

2023-10-04

SWAP Software

User Office

Onboarding user with their visit at ESS and manage all the required steps



The screenshot shows the 'User Office / Proposal' interface. At the top, it says 'Logged in as max.novelli@ess.eu (User Officer)'. Below this is a table titled 'Proposals' with columns: Actions, Proposal ID, Title, Submitted, Status, Notified, Call, Technical status, Assigned technical reviewer, and Ir. There are three rows of proposals.

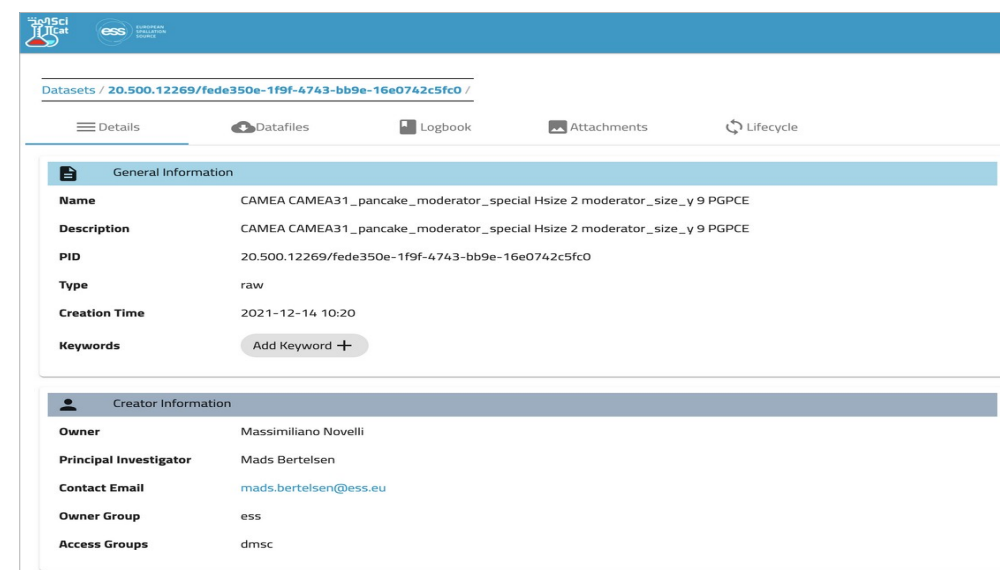
Actions	Proposal ID	Title	Submitted	Status	Notified	Call	Technical status	Assigned technical reviewer	Ir
<input type="checkbox"/>	369805	Determination of Cationic liposome - cationic polymer - DNA association with combined SAXS and SANS with contrast matching	No	Allocated	No	DEMAX-19	-	-	[
<input type="checkbox"/>	443766	Protonation dynamics in phytochrome photoreceptors	Yes	SEP review	No	DEMAX-19	Feasible	Carina Lobley	[
<input type="checkbox"/>	876643	Yeast phospholipids for neutron scattering studies of cell membrane dynamics and structure	No	SEP review	No	DEMAX-19	-	-	[

Technologies:

- React, Angular
- TypeScript, Python
- Node
- PostgreSQL, MongoDB

SciCat

Scientific data management and access.
Data publishing and sharing.



The screenshot shows the 'SciCat' interface. At the top, it says 'Datasets / 20.500.12269/fede350e-1f9f-4743-bb9e-16e0742c5fc0 /'. Below this is a table titled 'General Information' with columns: Name, Description, PID, Type, Creation Time, and Keywords. There is also a section titled 'Creator Information' with columns: Owner, Principal Investigator, Contact Email, Owner Group, and Access Groups.

General Information	
Name	CAMEA CAMEA31_pancake_moderator_special Hsize 2 moderator_size_v 9 PGPCE
Description	CAMEA CAMEA31_pancake_moderator_special Hsize 2 moderator_size_v 9 PGPCE
PID	20.500.12269/fede350e-1f9f-4743-bb9e-16e0742c5fc0
Type	raw
Creation Time	2021-12-14 10:20
Keywords	<input type="text" value="Add Keyword"/>

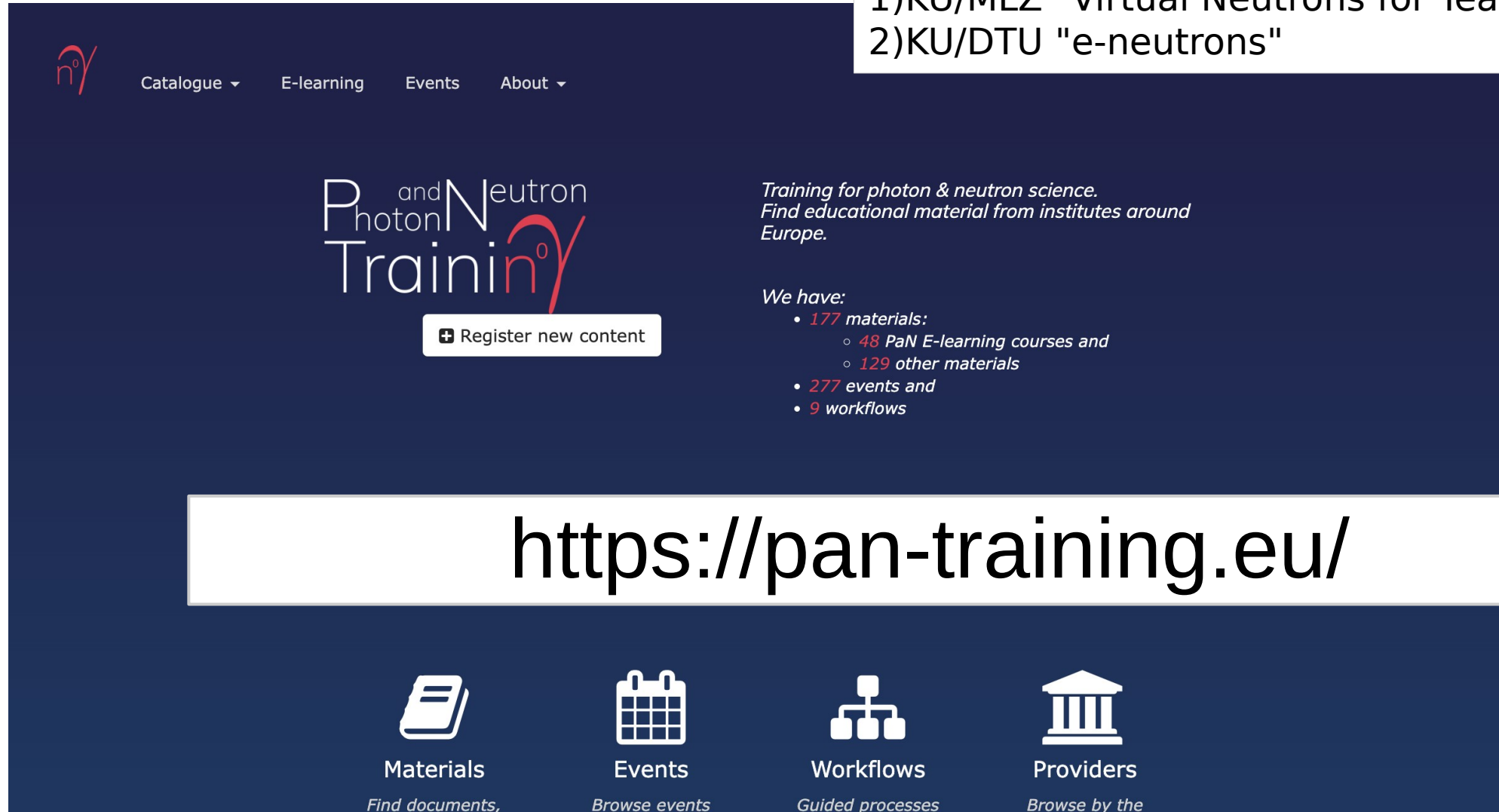
Creator Information	
Owner	Massimiliano Novelli
Principal Investigator	Mads Bertelsen
Contact Email	mads.bertelsen@ess.eu
Owner Group	ess
Access Groups	dmsc

- RabbitMQ, Kafka
- Docker, Kubernetes
- Git
- CI/CD

E-learning - pan-training.eu

Origins:

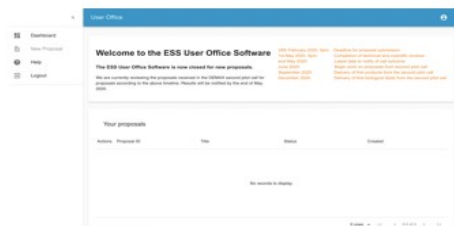
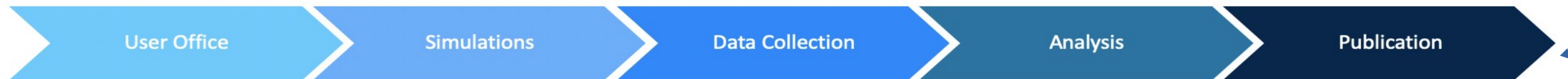
- 1) KU/MLZ "Virtual Neutrons for Teaching"
- 2) KU/DTU "e-neutrons"



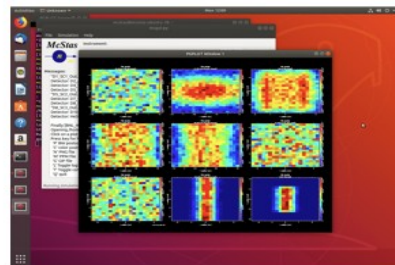
<https://pan-training.eu/>

User Journey

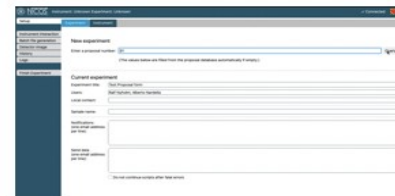
Which systems the users interact with at ESS



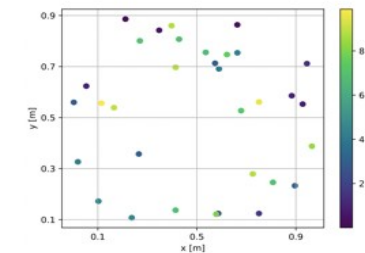
User Office



McStas



NICOS

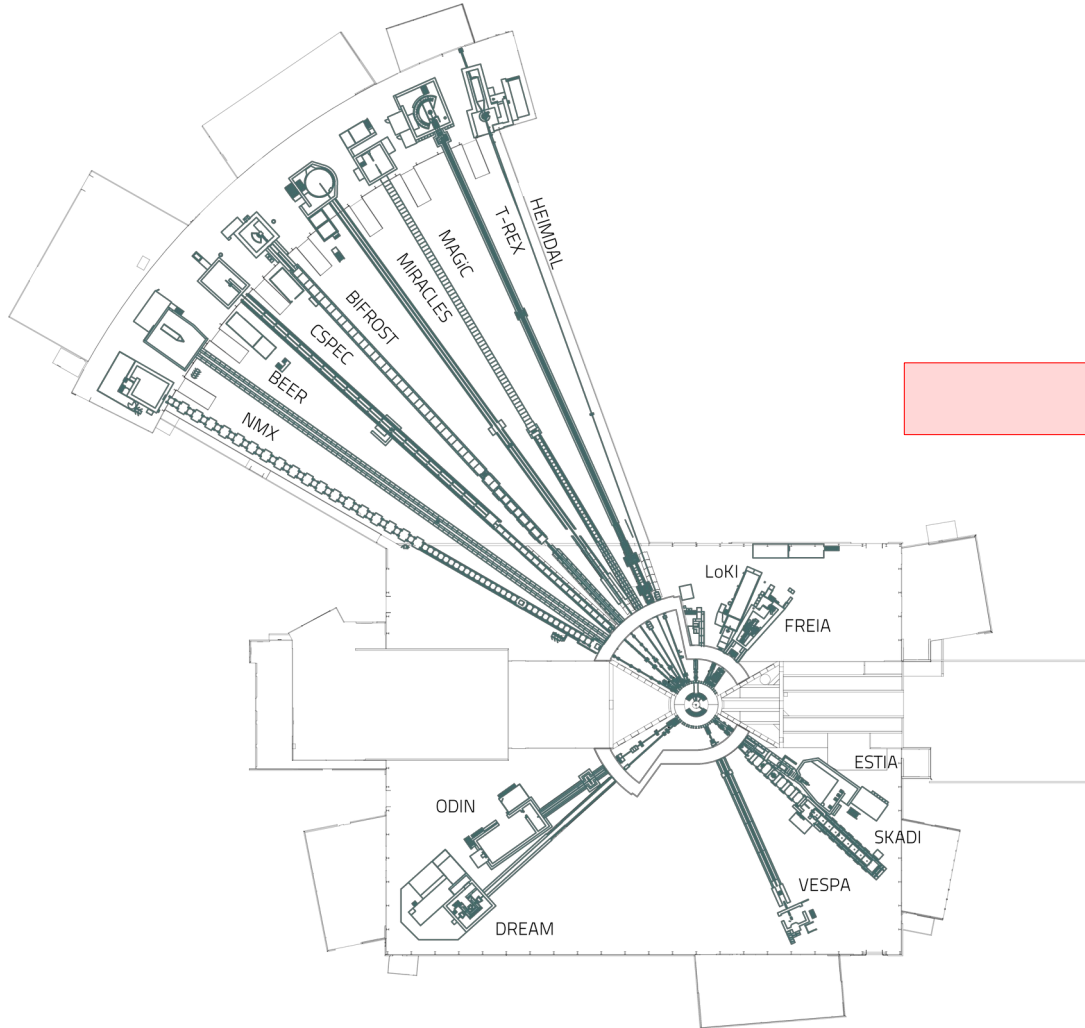


Scipp + analysis



DOI Landing Page

Data Journey: Experiment



<input type="checkbox"/>		Name		Run No.		Source Folder		Size		Start Time		Type		Image		Science Metadata		Proposal ID
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:47		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:46		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:45		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:44		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:43		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:42		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:41		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:40		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:39		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:37		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:36		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:35		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:34		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			
<input type="checkbox"/>		can_we_crash_the_flowriter		..ymir/2022		10 MB		2022-11-09 Wed 19:33		raw		tra__approx_file_size_mb(object Object)	tra_entry_instrument_fir_camera_image_key_average_va		038243			

Data Journey: Curation



<input type="checkbox"/>	Name	Run No.	Source Folder	Size	Start Time	Type	Image	Science Metadata	Proposal ID
<input type="checkbox"/>	can_we_crash_the_filewriter		.../ymir/2022	10 MB	2022-11-09	raw	SciCat	Extra Entry Instrument Flir Camera Image Key Average Value	038243
<input type="checkbox"/>	can_we_crash_the_filewriter		.../ymir/2022	10 MB				Extra Entry Instrument Flir Camera Image Key Maximum Value	
<input type="checkbox"/>	can_we_crash_the_filewriter		.../ymir/2022	10 MB				Extra Entry Instrument Flir Camera Image Key Minimum Value	
<input type="checkbox"/>	can_we_crash_the_filewriter		.../ymir/2022	10 MB				Extra Entry Instrument Flir Camera X Pixel Size Average Value	
<input type="checkbox"/>	can_we_crash_the_filewriter		.../ymir/2022	10 MB				Extra Entry Instrument Flir Camera X Pixel Size Maximum Value	
<input type="checkbox"/>	can_we_crash_the_filewriter		.../ymir/2022	10 MB				Extra Entry Instrument Flir Camera X Pixel Size Minimum Value	
<input type="checkbox"/>	can_we_crash_the_filewriter		.../ymir/2022	10 MB				Extra Entry Instrument Flir Camera Y Pixel Size Average Value	

Scientific Metadata

View

Edit

Name	Value	Unit
Extra Entry Instrument Flir Camera Image Key Average Value	0	--
Extra Entry Instrument Flir Camera Image Key Maximum Value	0	--
Extra Entry Instrument Flir Camera Image Key Minimum Value	0	--
Extra Entry Instrument Flir Camera X Pixel Size Average Value	0	--
Extra Entry Instrument Flir Camera X Pixel Size Maximum Value	0	--
Extra Entry Instrument Flir Camera X Pixel Size Minimum Value	0	--
Extra Entry Instrument Flir Camera Y Pixel Size Average Value	0	--

How do users find their data?

Thousands Datasets, Hundreds of metadata and files

Y Maximum Value	0.06070312500003183	--
Extra Entry Instrument Laser Aperature loc Timestamps	0.060625000000015916	--
Y Minimum Value		
Extra Entry Instrument Laser Aperature loc Timestamps		

Data Journey: Search, Analyse, Publish



Datasets /

Search Clear

Text Search

Location

Group

Type

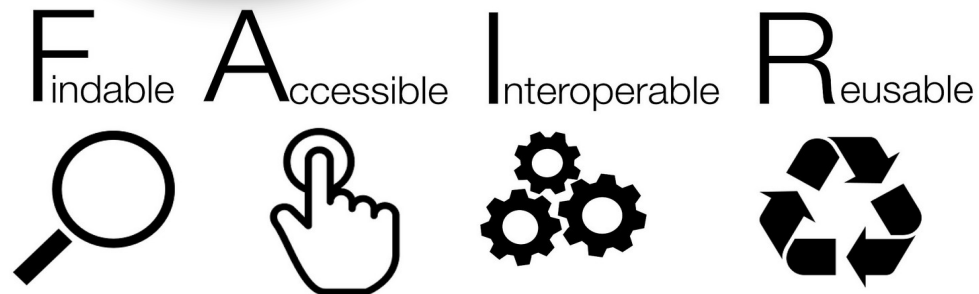
Keywords

Start Date

[+ Add Condition](#)



Publication Status	
Status	pending_registration
Register	
About the published data	
Title	Demo
Creator List	clement
Abstract	Demo
DOI	10.5072/2d629944-9792-449c-ab54-3feb9bd5ccaa
URL	doi2.psi.ch/detail/10.5072/2d629944-9792-449c-ab54-3feb9bd5ccaa
Administrative metadata	
User	Idap.Henrik Johansson
Creator	clement
Data Description	None
Dataset IDs	20.500.12269/BCMFEEAcceptance_297_2020-07-14-17-04-37_3, 20.500.12269/BCMFEEAcceptance_297_2020-07-14-17-04-37_2, 20.500.12269/BCMFEEAcceptance_297_2020-07-14-17-04-37_1
Publisher	ESS
Resource Type	raw



Data Journey: Enrich and Enhance

Datasets / 20.500.12269/34125a73-4eb4-435b-973a-00cf1f1acf4a /

Details Datafiles Logbook Attachments Lifecycle

General Information

Name CAMEA 20, Pancake Moderator Test moderator_size_y 3 PGPCP 3moderator size y run1

Description CAMEA 20, Pancake Moderator Test moderator_size_y 3 PGPCP 3moderator size y run1 20221107152012

PID 20.500.12269/34125a73-4eb4-435b-973a-00cf1f1acf4a

Type raw

Creation Time 2022-11-07 15:20

Keywords Simulation Monte Carlo ray tracing Swarm optimization SLURM iFit McStas neutron guide CAMEA BIFROST CAMEA 20 CAMEA 20 Pancake Moderator Test Pancake Moderator Add Keyword +

Creator Information

Owner Massimiliano Novelli

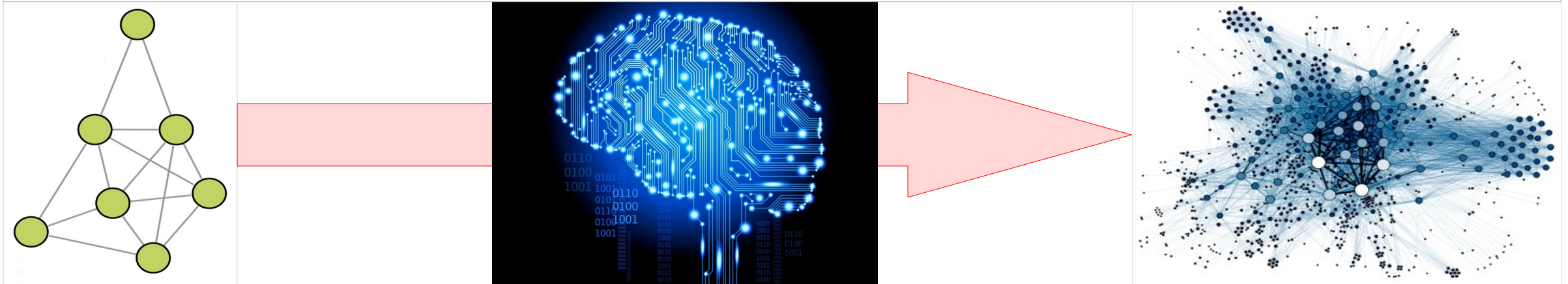
Principal Investigator Mads Bertelsen

Scientific Metadata

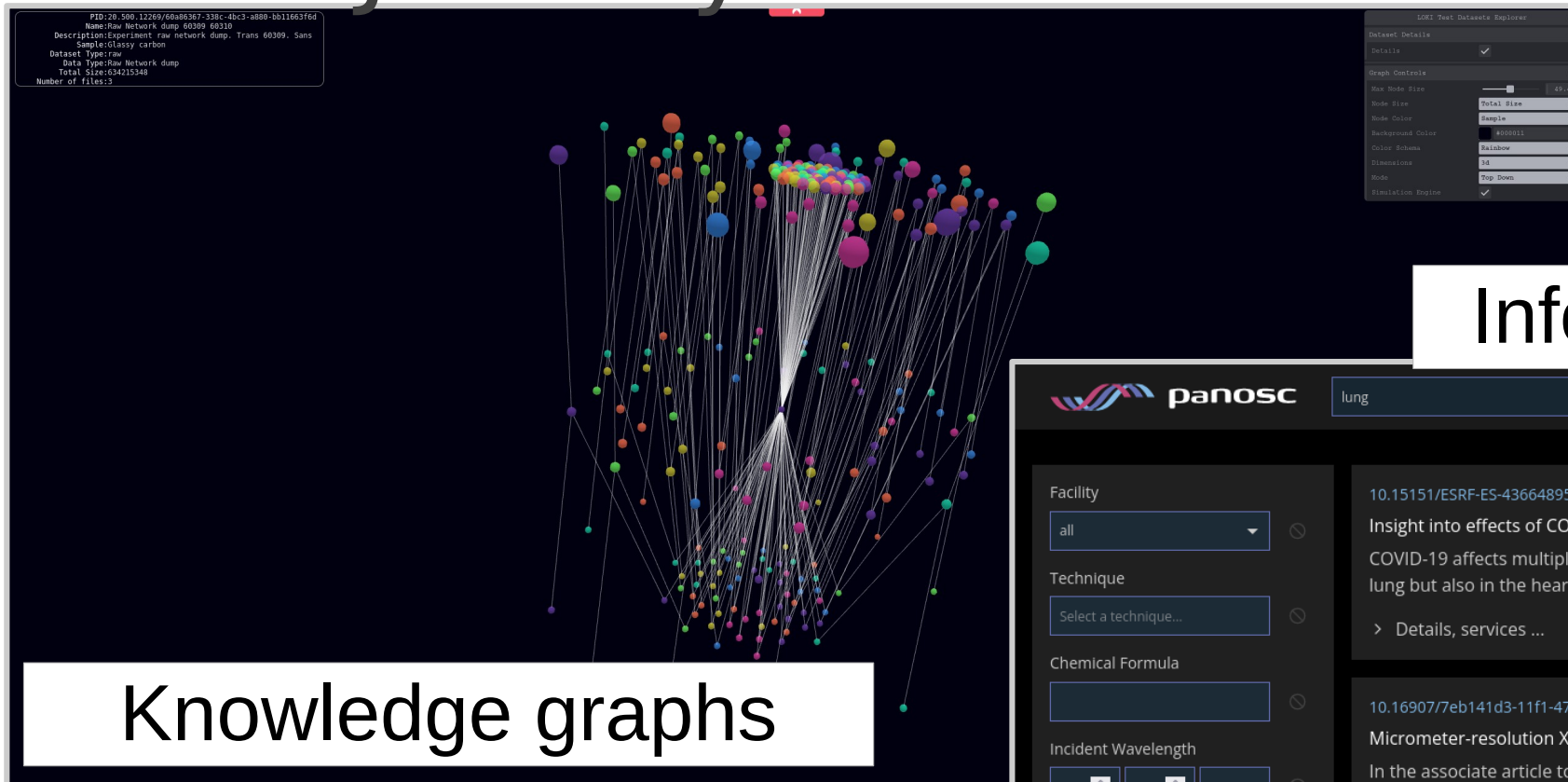
View Edit

Name	Value	Unit
Extra Entry Instrument Flir Camera Image Key Average Value	0	--
Extra Entry Instrument Flir Camera Image Key Maximum Value	0	--
Extra Entry Instrument Flir Camera Image Key Minimum Value	0	--
Extra Entry Instrument Flir Camera X Pixel Size Average Value	0	--
Extra Entry Instrument Flir Camera X Pixel Size Maximum Value	0	--
Extra Entry Instrument Flir Camera X Pixel Size Minimum Value	0	--
Extra Entry Instrument Flir Camera Y Pixel Size Average Value	0	--

Machine Learning and Artificial Intelligence



Data Journey: Use and Dissemination



Information Retrieval

panosc

lung

Facility

all

Technique

Select a technique...

Chemical Formula

Incident Wavelength

min max nm

Incident Photon Energy

min max eV

Temperature

min max K

Pressure

min max Pa

10.15151/ESRF-ES-436648953

Insight into effects of COVID-19 on human heart and lung tissue at sub-cellular level

COVID-19 affects multiple vital organs and leads to drastic changes in tissue architecture not only in the lung but also in the heart. The understanding of these alterations is currently very limited and no high...

> Details, services ...

Released by ESRF on May 7th 2024

10.16907/7eb141d3-11f1-47a6-9d0e-76f8832ed1b2

Micrometer-resolution X-ray tomographic imaging of a complete intact post mortem juvenile rat lung

In the associate article to these data sets, we present an X-ray tomographic imaging method that is well suited for pulmonary disease studies in animal models, to resolve the full pathway from gas intake to ga...

> Details, services ...

Released by PSI on January 1st 2020

10.5291/ILL-DATA.9-13-653

Structural changes to the air-water interface following cholesterol oxidation in mixed lung lipid monolay...

The presence of lung surfactant at the air-water interface of the lung is vital to prevent death from respiratory failure. Ozone, a pollutant in ambient air, is known to damage some components of lung...

> Details, services ...

Released by ILL on September 26th 2020

Potential Projects

Creating/curating/using a set of diverse data



Software developing

Data capturing Capture relevant metadata and do so automatically.

Validation Validate dataset metadata.

Curation Improve metadata and migrate to homogeneous metadata
Improve data F.A.I.R.ness.

Visualization Aid users in searching and evaluating their dataset.
Aid instrument scientists evaluating and improving their instrument.
Aid management improving data curation and FAIRness.
Knowledge graphs and discovery tools

ML/AI Make dataset and metadata machine readable.
Leverage ML/AI to enhance metadata and/or data.

Search Find data by leveraging Information Retrieval.
Find and disseminate Open Data.



Thank you

Questions?

Max Novelli
max.novelli@ess.eu