Elettra
Sincrotrone
Trieste

Scientific Data Management

Roberto Pugliese

roberto.pugliese@elettra.eu
ICT Group

CERTIFIED MANAGEMENT SYSTEM
CQY CERTIQUALITY
UNI EN ISO 9001:2015
UNI ISO 45001:2018

Outline

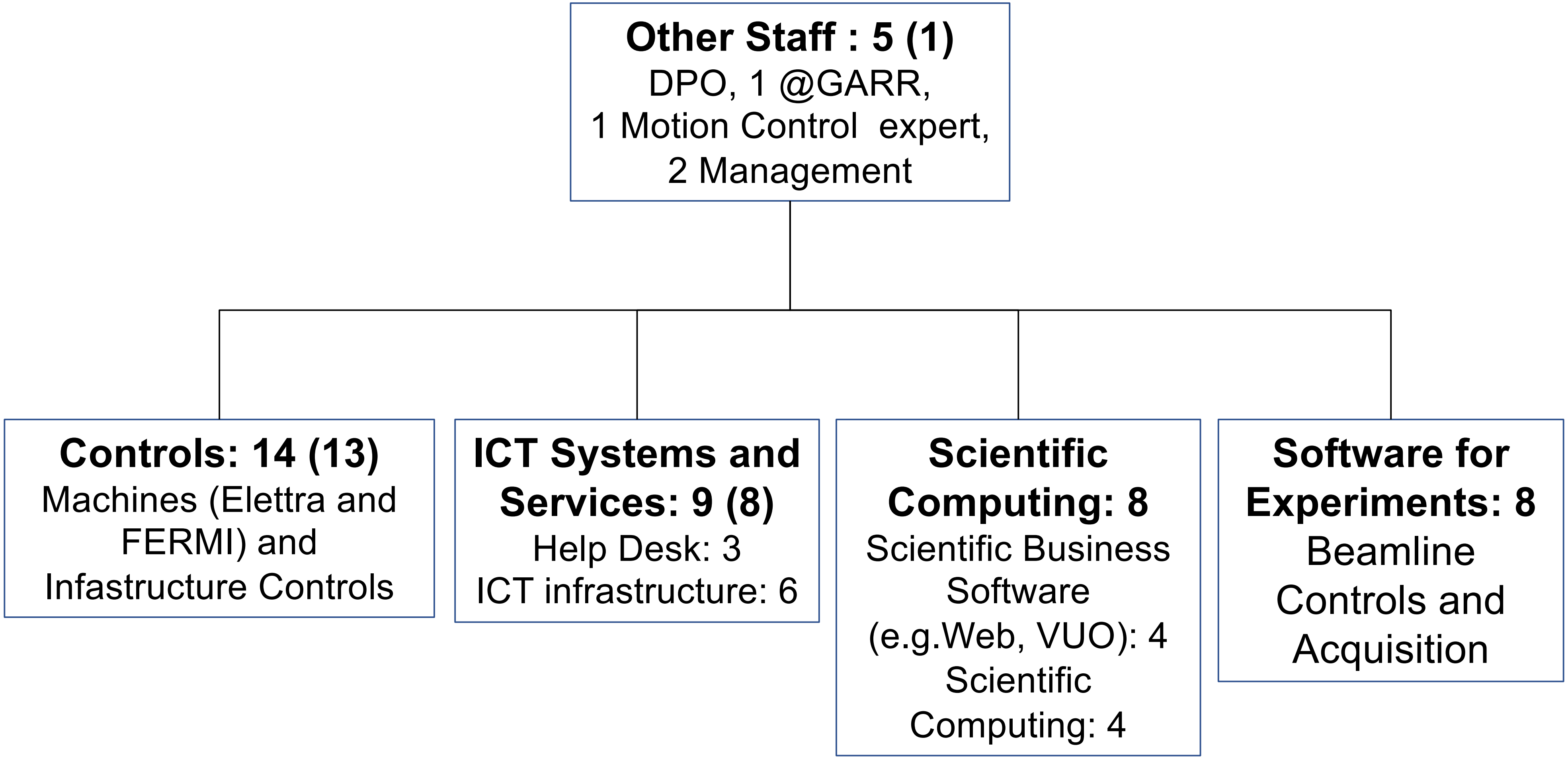
✓Elettra ICT Group

✓ICT Infrastructure

✓FAIR Data

- Sientific Data Policy
- DOIs and Scientific Data Storage implementation
- KPIs

Elettra ICT Group: 44 (38)



Other Staff : 5 (1)
DPO, 1 @GARR,
1 Motion Control expert,
2 Management

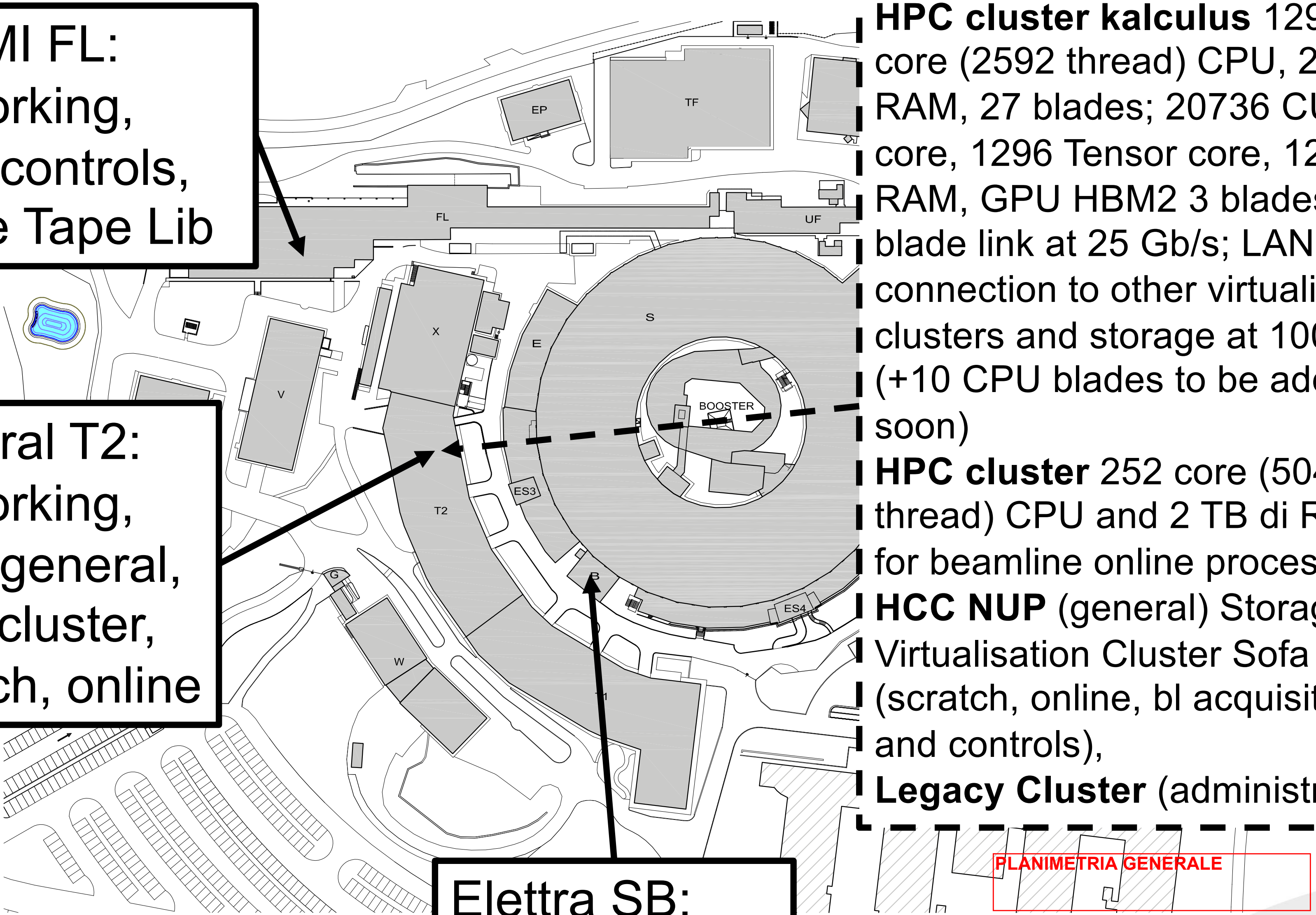
Controls: 14 (13)
Machines (Elettra and FERMI) and Infastructure Controls

ICT Systems and Services: 9 (8)
Help Desk: 3
ICT infrastructure: 6

Scientific Computing: 8
Scientific Business Software (e.g.Web, VUO): 4
Scientific Computing: 4

Software for Experiments: 8
Beamline Controls and Acquisition

IT computing and storage resources



FERMI FL: networking, HCC controls, offline Tape Lib

General T2: networking, HCC general, HPC cluster, scratch, online

Elettra SB: networking, HCC controls

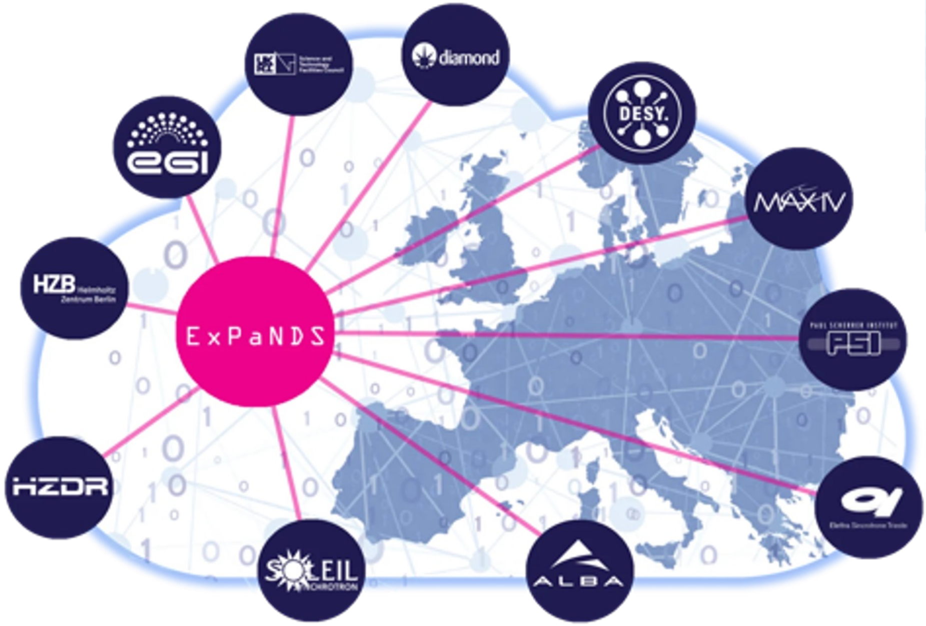
HPC cluster kalculus 1296 core (2592 thread) CPU, 20 TB RAM, 27 blades; 20736 CUDA core, 1296 Tensor core, 120GB RAM, GPU HBM2 3 blades; blade link at 25 Gb/s; LAN connection to other virtualisation clusters and storage at 100 Gb/s (+10 CPU blades to be added soon)

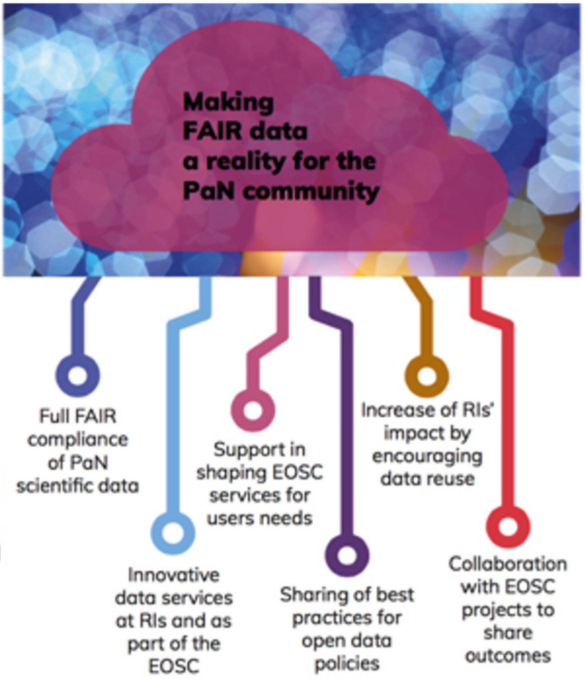
HPC cluster 252 core (504 thread) CPU and 2 TB di RAM for beamline online processing

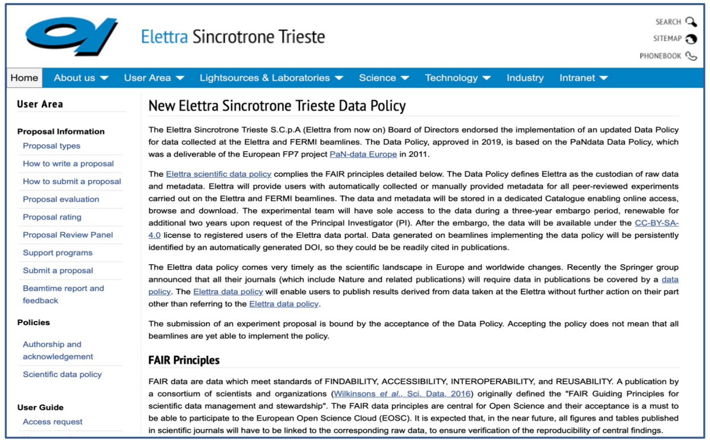
HCC NUP (general) Storage + Virtualisation Cluster Sofa (scratch, online, bl acquisition and controls),

Legacy Cluster (administration)

FAIR Data Projects: ExPaNDS, PaNOSC


ExPaNDS
EGI
Science and Technology Facilities Council
diamond
DESY.
MAX IV
PAUL SCHERRER INSTITUT PSI
Elettra Sincrotrone Trieste
ALBA
SOLEIL SYNCHROTRON
HZDR
HZB Helmholtz Zentrum Berlin


Making FAIR data a reality for the PaN community
Full FAIR compliance of PaN scientific data
Innovative data services at RIs and as part of the EOSC
Support in shaping EOSC services for users needs
Sharing of best practices for open data policies
Increase of RIs' impact by encouraging data reuse
Collaboration with EOSC projects to share outcomes



User Area

Proposal Information

- Proposal types
- How to write a proposal
- How to submit a proposal
- Proposal evaluation
- Proposal rating
- Proposal Review Panel
- Support programs
- Submit a proposal
- Beamtime report and feedback

Policies

- Authorship and acknowledgement
- Scientific data policy

User Guide

- Access request

New Elettra Sincrotrone Trieste Data Policy

The Elettra Sincrotrone Trieste S.C.p.A (Elettra from now on) Board of Directors endorsed the implementation of an updated Data Policy for data collected at the Elettra and FERMI beamlines. The Data Policy, approved in 2019, is based on the PaNdata Data Policy, which was a deliverable of the European FP7 project PaN-data Europe in 2011.

The Elettra scientific data policy complies the FAIR principles detailed below. The Data Policy defines Elettra as the custodian of raw data and metadata. Elettra will provide users with automatically collected or manually provided metadata for all peer-reviewed experiments carried out on the Elettra and FERMI beamlines. The data and metadata will be stored in a dedicated Catalogue enabling online access, browse and download. The experimental team will have sole access to the data during a three-year embargo period, renewable for additional two years upon request of the Principal Investigator (PI). After the embargo, the data will be available under the CC-BY-SA-4.0 license to registered users of the Elettra data portal. Data generated on beamlines implementing the data policy will be persistently identified by an automatically generated DOI, so they could be be readily cited in publications.

The Elettra data policy comes very timely as the scientific landscape in Europe and worldwide changes. Recently the Springer group announced that all their journals (which include Nature and related publications) will require data in publications be covered by a data policy. The Elettra data policy will enable users to publish results derived from data taken at the Elettra without further action on their part other than referring to the Elettra data policy.

The submission of an experiment proposal is bound by the acceptance of the Data Policy. Accepting the policy does not mean that all beamlines are yet able to implement the policy.

FAIR Principles

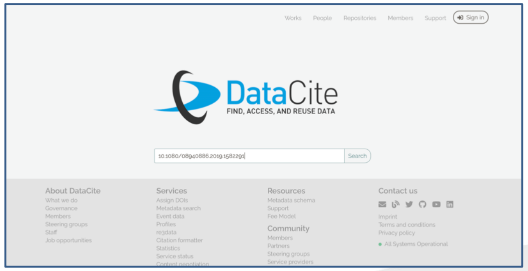
FAIR data are data which meet standards of FINDABILITY, ACCESSIBILITY, INTEROPERABILITY, and REUSABILITY. A publication by a consortium of scientists and organizations (Wilkinsons *et al.*, Sci. Data, 2016) originally defined the "FAIR Guiding Principles for scientific data management and stewardship". The FAIR data principles are central for Open Science and their acceptance is a must to be able to participate to the European Open Science Cloud (EOSC). It is expected that, in the near future, all figures and tables published in scientific journals will have to be linked to the corresponding raw data, to ensure verification of the reproducibility of central findings.

Elettra
Sincrotrone
Trieste

Scientific Data Storage Architecture

DATA CATALOGUE	Without Metadata	With Metadata	
STANDARD STATION	1 (Data Uploader)	2	3
AUTOMATIZED STATION		1	2
STORAGE SYSTEM	SCRATCH Can be organized Can be processed Can be transferred	ONLINE Can be organized Can be processed Can be transferred	OFFLINE Can be transferred
CAPACITY	2PB (CEPH – Rep. 3)	4PB (CEPH – Rep. 3) (48%)	6 PB Up to 60PB (Replica 4)

✓ When online datasets can have an associated DOI and you can search the datasets in the DataCite, PaNOSC and EOSC portals

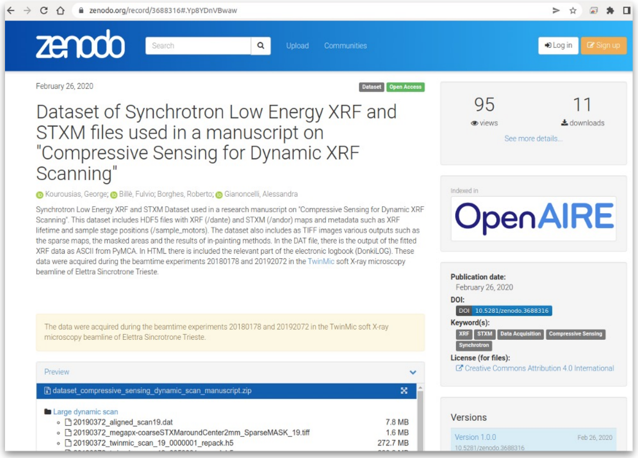


10.1080/08940886.2019.1582291

In summary what happens now …

- ✓ **Before proposal submission**
  - The new scientific data policy is available on the web
- ✓ **During proposal submission**
  - The principal investigator has to accept the new scientific data policy
- ✓ **Before beamtime**
  - A chat group is be created 6 months before the first experiment
- ✓ **During beamtime**
  - chat and all local and remote support services are available to whole experimental team
- ✓ **After beamtime**
  - 2 weeks after the end data is copied offline to the tape library,
  - BEST (the day after), achievements (after 30 days) check DOI text, generate **DOI**, generate **DMP**, chat 1 year after the last experiment closes is closed
- ✓ **Data become open access:** 3y[+1y[+1y]] or on request by the principal investigator
- ✓ **When a data access request arrives**: the principal investigator is informed
- ✓ **Moreover** … scratch is pruned when space occupation is above 75%, on-line when above 75%, when deadline expires (10 years …) and when above 75% off-line space will be pruned while metadata is kept forever … best effort.

FAIR: scientist feedback



February 26, 2020 | Dataset | Open Access

Dataset of Synchrotron Low Energy XRF and STXM files used in a manuscript on "Compressive Sensing for Dynamic XRF Scanning"

Kourousias, George; Billè, Fulvio; Borghes, Roberto; Gianoncelli, Alessandra

Synchrotron Low Energy XRF and STXM Dataset used in a research manuscript on "Compressive Sensing for Dynamic XRF Scanning". This dataset includes HDF5 files with XRF (/dante) and STXM (/andor) maps and metadata such as XRF lifetime and sample stage positions (/sample_motors). The dataset also includes as TIFF images various outputs such as the sparse maps, the masked areas and the results of in-painting methods. In the DAT file, there is the output of the fitted XRF data as ASCII from PyMCA. In HTML there is included the relevant part of the electronic logbook (DonkiLOG). These data were acquired during the beamtime experiments 20180178 and 20192072 in the TwinMic soft X-ray microscopy beamline of Elettra Sincrotrone Trieste.

The data were acquired during the beamtime experiments 20180178 and 20192072 in the TwinMic soft X-ray microscopy beamline of Elettra Sincrotrone Trieste.

Preview

dataset_compressive_sensing_dynamic_scan_manuscript.zip

Large dynamic scan

20190372_aligned_scan19.dat	7.8 MB
20190372_megapx-coarseSTXMaroundCenter2mm_SparseMASK_19.tiff	1.6 MB
20190372_twinmic_scan_19_0000001_repack.h5	272.7 MB

95 views | 11 downloads

See more details...

Indexed in OpenAIRE

Publication date:
February 26, 2020

DOI:
DOI 10.5281/zenodo.3688316

Keyword(s):
XRF | STXM | Data Acquisition | Compressive Sensing | Synchrotron

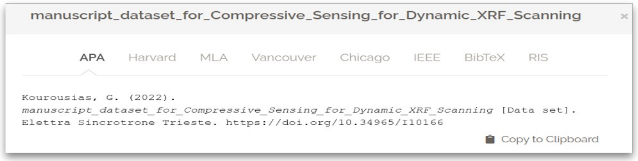
License (for files):
Creative Commons Attribution 4.0 International

Versions

Version 1.0.0 — Feb 26, 2020
10.5281/zenodo.3688316

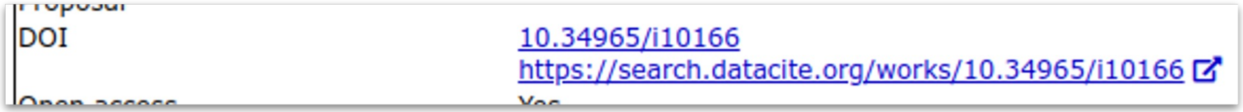
- ✓ During a BLEX-IT meeting and while doing beamtime together we discovered that users need something:
  - Similar to **Zenodo**
  - Allows for **public data access** but **not anonymous**
  - Hosted in **Elettra**
  - Compatible with major **scientific Journals** and EU projects
  - In accordance with the Elettra Scientific **Data Policy**
  - **Open** and FAIR
  - Provides the data owner with useful **information and control**

You upload or select your data and get two links, one to the VUO where you can access the data (http://dx.doi.org/10.34965/i10166) and the other one to a public "search" portal.

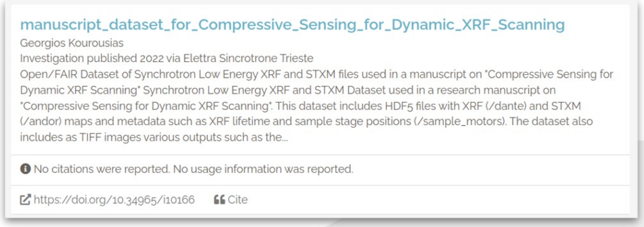


manuscript_dataset_for_Compressive_Sensing_for_Dynamic_XRF_Scanning

```
Kourousias, G. (2022).
manuscript_dataset_for_Compressive_Sensing_for_Dynamic_XRF_Scanning [Data set].
Elettra Sincrotrone Trieste. https://doi.org/10.34965/I10166
```



Proposal	
DOI	10.34965/i10166 https://search.datacite.org/works/10.34965/i10166
Open access	Yes



manuscript_dataset_for_Compressive_Sensing_for_Dynamic_XRF_Scanning

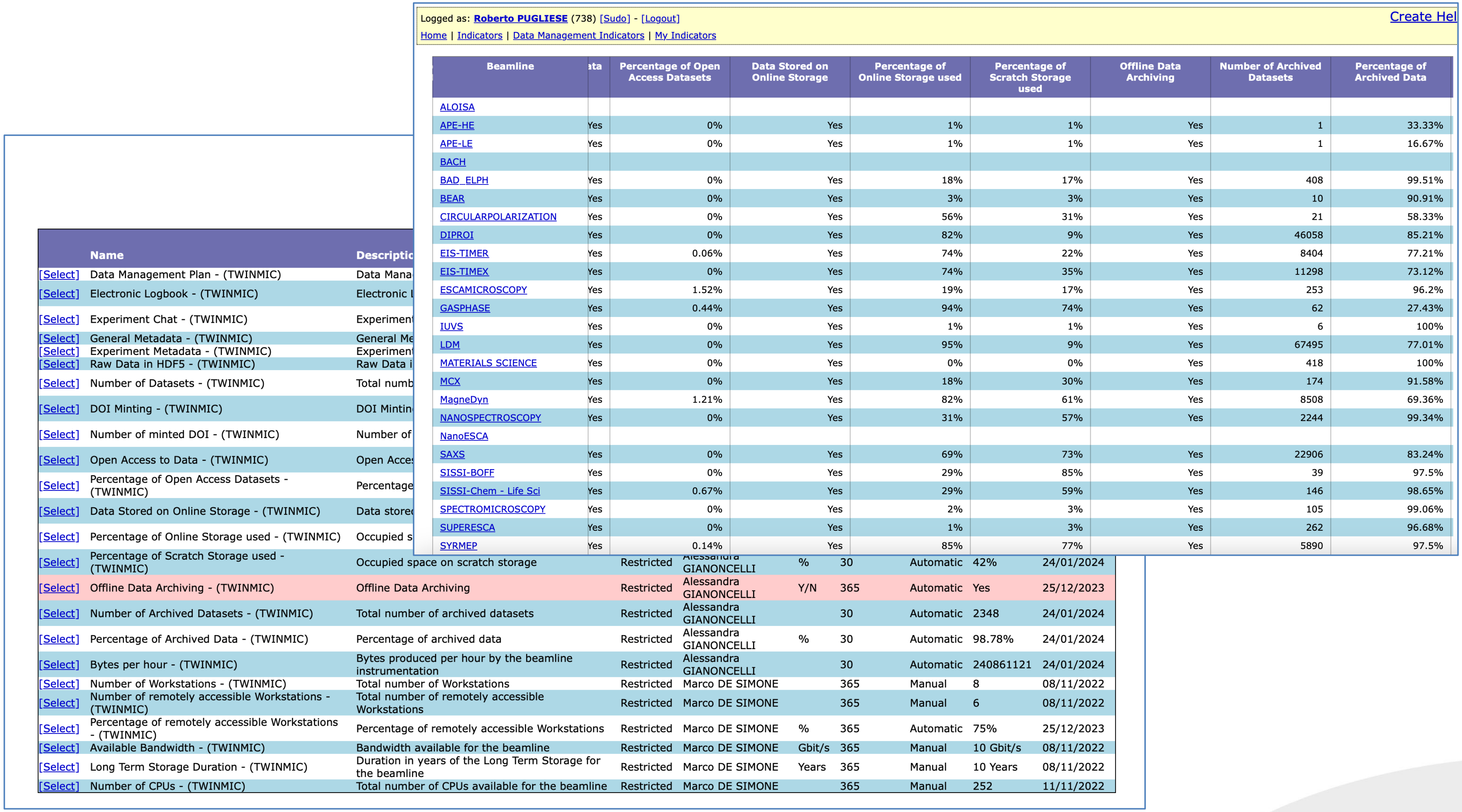
Georgios Kourousias

Investigation published 2022 via Elettra Sincrotrone Trieste

Open/FAIR Dataset of Synchrotron Low Energy XRF and STXM files used in a manuscript on "Compressive Sensing for Dynamic XRF Scanning" Synchrotron Low Energy XRF and STXM Dataset used in a research manuscript on "Compressive Sensing for Dynamic XRF Scanning". This dataset includes HDF5 files with XRF (/dante) and STXM (/andor) maps and metadata such as XRF lifetime and sample stage positions (/sample_motors). The dataset also includes as TIFF images various outputs such as the...

No citations were reported. No usage information was reported.

KPIs to monitor progress and achievements



Logged as: **Roberto PUGLIESE** (738) [Sudo] - [Logout]

Create Hel

Home | Indicators | Data Management Indicators | My Indicators

Beamline	ata	Percentage of Open Access Datasets	Data Stored on Online Storage	Percentage of Online Storage used	Percentage of Scratch Storage used	Offline Data Archiving	Number of Archived Datasets	Percentage of Archived Data
ALOISA								
APE-HE	Yes	0%	Yes	1%	1%	Yes	1	33.33%
APE-LE	Yes	0%	Yes	1%	1%	Yes	1	16.67%
BACH								
BAD_ELPH	Yes	0%	Yes	18%	17%	Yes	408	99.51%
BEAR	Yes	0%	Yes	3%	3%	Yes	10	90.91%
CIRCULARPOLARIZATION	Yes	0%	Yes	56%	31%	Yes	21	58.33%
DIPROI	Yes	0%	Yes	82%	9%	Yes	46058	85.21%
EIS-TIMER	Yes	0.06%	Yes	74%	22%	Yes	8404	77.21%
EIS-TIMEX	Yes	0%	Yes	74%	35%	Yes	11298	73.12%
ESCAMICROSCOPY	Yes	1.52%	Yes	19%	17%	Yes	253	96.2%
GASPHASE	Yes	0.44%	Yes	94%	74%	Yes	62	27.43%
IUVS	Yes	0%	Yes	1%	1%	Yes	6	100%
LDM	Yes	0%	Yes	95%	9%	Yes	67495	77.01%
MATERIALS SCIENCE	Yes	0%	Yes	0%	0%	Yes	418	100%
MCX	Yes	0%	Yes	18%	30%	Yes	174	91.58%
MagneDyn	Yes	1.21%	Yes	82%	61%	Yes	8508	69.36%
NANOSPECTROSCOPY	Yes	0%	Yes	31%	57%	Yes	2244	99.34%
NanoESCA								
SAXS	Yes	0%	Yes	69%	73%	Yes	22906	83.24%
SISSI-BOFF	Yes	0%	Yes	29%	85%	Yes	39	97.5%
SISSI-Chem - Life Sci	Yes	0.67%	Yes	29%	59%	Yes	146	98.65%
SPECTROMICROSCOPY	Yes	0%	Yes	2%	3%	Yes	105	99.06%
SUPERESCA	Yes	0%	Yes	1%	3%	Yes	262	96.68%
SYRMEP	Yes	0.14%	Yes	85%	77%	Yes	5890	97.5%

	Name	Descriptio							
[Select]	Data Management Plan - (TWINMIC)	Data Mana							
[Select]	Electronic Logbook - (TWINMIC)	Electronic L							
[Select]	Experiment Chat - (TWINMIC)	Experiment							
[Select]	General Metadata - (TWINMIC)	General Me							
[Select]	Experiment Metadata - (TWINMIC)	Experiment							
[Select]	Raw Data in HDF5 - (TWINMIC)	Raw Data i							
[Select]	Number of Datasets - (TWINMIC)	Total numb							
[Select]	DOI Minting - (TWINMIC)	DOI Mintin							
[Select]	Number of minted DOI - (TWINMIC)	Number of							
[Select]	Open Access to Data - (TWINMIC)	Open Acces							
[Select]	Percentage of Open Access Datasets - (TWINMIC)	Percentage							
[Select]	Data Stored on Online Storage - (TWINMIC)	Data stored							
[Select]	Percentage of Online Storage used - (TWINMIC)	Occupied s							
[Select]	Percentage of Scratch Storage used - (TWINMIC)	Occupied space on scratch storage	Restricted	Alessandra GIANONCELLI	%	30	Automatic	42%	24/01/2024
[Select]	Offline Data Archiving - (TWINMIC)	Offline Data Archiving	Restricted	Alessandra GIANONCELLI	Y/N	365	Automatic	Yes	25/12/2023
[Select]	Number of Archived Datasets - (TWINMIC)	Total number of archived datasets	Restricted	Alessandra GIANONCELLI		30	Automatic	2348	24/01/2024
[Select]	Percentage of Archived Data - (TWINMIC)	Percentage of archived data	Restricted	Alessandra GIANONCELLI	%	30	Automatic	98.78%	24/01/2024
[Select]	Bytes per hour - (TWINMIC)	Bytes produced per hour by the beamline instrumentation	Restricted	Alessandra GIANONCELLI		30	Automatic	240861121	24/01/2024
[Select]	Number of Workstations - (TWINMIC)	Total number of Workstations	Restricted	Marco DE SIMONE		365	Manual	8	08/11/2022
[Select]	Number of remotely accessible Workstations - (TWINMIC)	Total number of remotely accessible Workstations	Restricted	Marco DE SIMONE		365	Manual	6	08/11/2022
[Select]	Percentage of remotely accessible Workstations - (TWINMIC)	Percentage of remotely accessible Workstations	Restricted	Marco DE SIMONE	%	365	Automatic	75%	25/12/2023
[Select]	Available Bandwidth - (TWINMIC)	Bandwidth available for the beamline	Restricted	Marco DE SIMONE	Gbit/s	365	Manual	10 Gbit/s	08/11/2022
[Select]	Long Term Storage Duration - (TWINMIC)	Duration in years of the Long Term Storage for the beamline	Restricted	Marco DE SIMONE	Years	365	Manual	10 Years	08/11/2022
[Select]	Number of CPUs - (TWINMIC)	Total number of CPUs available for the beamline	Restricted	Marco DE SIMONE		365	Manual	252	11/11/2022

Elettra
Sincrotrone
Trieste

Thanks!
Questions?

www.elettra.eu