

Data Processing and Management Activities at ELI-ALPS

Lajos Schrettner, Balázs Bagó

13 September 2022

SZÉCHENYI 2020



European Union European Regional Development Fund



Hungarian Government

INVESTING IN YOUR FUTURE



- ELI-ALPS
 - Mission
 - Research infrastructure
- Data collection
 - Phases of an investigation
 - Sources of data
- Data storage and processing
- Data management
 - Policies, principles, integration

ELI – Extreme Light Infrastructure

ELI-ERIC ELI European Research Infrastructure Consortium

ELI-Beamlines Dolni Brezany, Czech Republic
ELI-ALPS Szeged, Hungary
ELI-NP Magurele, Romania

ELI laser facilities aim at hosting some of the most intense lasers world-wide, develop new interdisciplinary research opportunities with light from these lasers and secondary radiation derived from them, and make them available to an international scientific user community.

ELI-ERIC webpage at www.eli-laser.eu

) ei

ELI-ALPS Mission





Visit our webpage at www.eli-alps.hu/en/Gallery

ELI-ALPS Mission

- to generate X-UV and X-ray femtosecond and attosecond pulses, for temporal investigation of electron dynamics in atoms, molecules, plasmas and solids at the attosecond scale.
- to contribute to the technological development towards high average power, high peak intensity lasers.

Implementation poses many challenges, among them are

- IT infrastructure (data network, storage, computing)
- Control system software

)))ei

• Data processing, data management



Extreme Light Infrastructure - Attosecond Light Pulse Source



Research infrastructure



Research infrastructure

Special construction

I eli

- Vibration "free" base
- Building in a building, isolation
- Regulated environment
 - Cleanrooms
 - Vacuum chambers and tubing
- Safety measures
 - Card readers
 - Lab clothing
 - Protective glasses



Control system

I eli



Data Collection

Phases of an investigation

- Proposal submission, evaluation
- Preparation, simulation
- Experiment, measurement
- Analysis

) ei

Publication

Different kinds of data are collected in each phase Data management has to consider and deal with all

- Raw measurement data is not enough

Data Collection

Types of data

) ei

- Documents related to the investigation
 - Proposal, related research, published papers
- Simulation descriptions and results
- Measurements during the experiment (data acquisition)
 - Sensor data voltages, histograms, images, ...
 - Logbook entries blog of an experiment
 - Lab environment data
- Analysis descriptions and results charts, calculations
- Reports, publications

AIM: Verifiability & reproducibility of experiments

(Meta)Data Collection

Environment of an experiment

• Pulse characteristics

ei

- Vibration sensor data
- Vacuum/gas pressure data
 - (Lack of) chamber pressure
 - Technological gas pressure
- Air quality data
 - Temperature / humidity
 - Particle count
 - Relative pressure
- Cooling water





Electronic LogBook

- Inhouse developed WEB application
- **BLOG like structure (logbook, entry, comments)**
- **REST / Python API**

ELI-ALPS Test LogBook	Home Export LogBook Templates Sear	Content v Q	1 bagob Ge Lo				
Successfully logged in							
Home / LogBook							
+ Add LogBook entry			ELI-ALPS Test LogBook	Home Export LogBook Templates	Search	Content v Q	L bagob C Log out
LogBooks		test entry 4 🧭	Home / LogBook / April 27,	2022			
All 🗸	Posted by: ALPS Control at 2022-08-12 09:22:51 LogBook: SE test	12		Viewing posts from April 27, 2022			
Entries	1 The second sec		+ Add LogBook entry				
2022	e de formanie 20 Strange autor		LogBooks		Test p	post 🥖	
···· 🖬 12 (2) (··· 📰 11 (5)			All 🕶	Posted by: admin at 2022-04-27 11:01:27 LogBook:	SE test2		read more / 2 con
- 🔲 10 (1)		test entry 4 🧸	Entries				
· 📰 21 (4)	Posted by: ALPS Control at 2022-08-12 09:17:36 LogBook: SE test	12	2022				
06 (4)	E. The second seco		12 (2)	2.625 bar			
- 🚍 June (6)			11 (5)				
· 📰 08 (2) · 📰 07 (2)	a di ter para		🚈 July (11)	2.600 bar			ער זיזעער איי
April (1)			III 21 (4) III 08 (3)				
- 2021		test entry 3 🧭	P 🔳 06 (4)	2.575 bar			
February (2)	Posted by: ALPS Control at 2022-08-11 11:38:07 LogBook: SE test	12	🚍 June (6)	2 550 bar			
- 🖪 16 (1)	5. International data water and the second sec		08 (2)				
- September (2)	·		i 🗐 07 (2)	2.525 bar			
(🔝 06 (2)	(2 Marane) y - ghoin		j 🗐 27 (1)				
- Cotober (1)				2.500 bar		N TH WAY WALL WA	
	1	test entrv 2 🖌	- I 17 (1)				
) 🔳 16 (1)	2.475 bar 10:30 10:32 10:34 10:36			10:50 10:52 10:54 10:54
			September (2)	 cooling_water_feedback.mean {serial_number: 098424. 	A459} — cooling_water_feedback.mean	(serial_number: 102724A459) — cooling_water_f	eedback.mean {serial_number: 307924A030} 🛛 🚽
			06 (2)				
			2018	34 Hz			
			p 🔳 04 (1)				
			Draft entry	32 Hz			
			,				

Data Storage and Processing



Data Storage and Processing

Data storage and processing

- 1PB storage disk space on site
- HPC cluster (CPUs, GPUs) for scientific batch computation
- IT network

i ei

- 10 Gbit/s in lab, core network with 40 Gbit/s switches
- Ca. 50 Workstations, ca. 30 control system servers
- Servers for interactive computation
 - Scientific applications (MATLAB, ComSol, ...)
 - Jupyter notebooks

FAIR data principles

• Findable

- Unique Ids
- Rich metadata
- Accessible
 - Retrieval by standard, open protocol
- Interoperable
 - Formats are widely accepted, standard
 - Vocabularies
- Reusable
 - Accurate and relevant content
 - Licenses



EOSC integration - PaNOSC

- **PaNOSC Photon and Neutron Open Science Cloud**
 - Initiative integrate photon and neutron sources to **EOSC – 6 members from across Europe**
 - 9 work packages to make FAIR a reality
 - **ELI-ALPS** contributions to

)))ei

- **WP3: Data Catalogue services**
- **WP4: Data Analysis services**
- WP5: Virtual Neutron and X-ray Laboratory
- WP6: EOSC integration



PaNOSC webpage at

www.panosc.eu

Invenio RDM

• Survey

- Aspects
- Catalogue systems
 - SciCat, iCat and Invenio RDM
- Invenio RDM
 - DataCite metadata schema

NVENIO	RDM	Lenso Q. Uphan	4) Login (2) Serup
		29 result(s) found	Newest -
Versions		May 22, 2027 NO 52: 2Pages Yorking and	@ View
View all versions		Carter, Chung and Ward's gallery	
Access status Metadata-only	304	Determine the Long County of the American Structure and American Str	ity that them. Speak center partner, Do long professor color.
🖌 Open		THAT TRANSPORT STORE TRANSPORT PROGRAM, SHE WITH FEDGUARD DEFIN SPACE INDEED ME. 1 Definition in August 6, 2021	
Restricted	۲		
Embargoed Resource types		Riddle-Simon's gallery	die View
Publication	-	Grant, William Chebroson, Kenneth, O Henry, Edward Suffer part so. Speak pass green meet young read future. Traditional box least energy special central. So	mething give owner. Would
> 🗌 image	10	sensitient support state formation these care power care of the time time time, care on the security instant memory and an exercise of democratic include room time time time ten our student. Decade bit professor trail tablesing in August 6, 2021	ng min. mortain crieck
Y Dataset	۲		
Other	۲	Dec 16, 1943 - Dec 21, 2014 Md-2.5	Gir Viewe
Poster	(8)	Cross and Sons's gallery	
Presentation		Across table voice box him power. Close pay company season. Movie support wonder debate enjoy. Clo	se fill opportunity wind white
Lesson	۲	protessor this compare. Up task per I. Worry international ready fear tend frim. Participant your wait soli cover line billion assume. Order water recent member will traditional. Congress th	der heart Indeed. Give human
Video/Audio	۲	Upticaled to Augent 6, 2021	
Software	۲	2014/07/2012/2012/01/11	dit View
iearch guide		Chen Ltd's gallery	

Data management status

- ELI ERIC data policy
 - Available on Zenodo
- Data management plan
 - User

- DAQ machines
- Folder
- Remote access
- Data pilot project with NanoESCA

Planned FAIR Data architecture





THANK YOU FOR YOUR ATTENTION!





European Union European Regional Development Fund



Hungarian Government

INVESTING IN YOUR FUTURE