

PaNOSC Data Catalogue Work Package 3 Catalogue Integration Best Practices Meeting

<https://indico.eli-laser.eu/event/3/>

18th – 20th May 2021

Tobias Richter (ESS) – Leader WP3 / Lajos Schrettner (ELI) – Host



Outline

→ **Welcome!**

→ **PaNOSC**

→ **Work Package 3 – Data Catalogues**

→ **Catalogue Integration & Best**

Practices



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



PaNOSC Factsheet

Description: cluster of ESFRI Photon and Neutron sources

Partners: ESRF, ILL, XFEL.EU, ESS, CERIC-ERIC, ELI, EGI

Coordinator: ESRF

Observers/non-funded: GÉANT, EUDAT, national RIs

Call: Horizon 2020 InfraEOSC-04

Budget: 12 M€

Started: 2018-12-01

Duration: 4 years

GitHub: <https://github.com/panosc-eu>

Home page: <https://panosc.eu>

Twitter: @PaNOSC_eu #PaNOSC



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



PaNOSC is one of 5 ESFRI clusters



The Science Clusters [ENVRI-FAIR](#), [EOSC-LIFE](#), [ESCAPE](#), [PaNOSC](#), [SSHOC](#) bring together 72 world-class research infrastructures from the ESFRI roadmap and beyond to work on FAIR data management and connecting their user communities to the EOSC. The European e-infrastructures [EGI](#), [EUDAT](#), [GÉANT](#), and [OpenAIRE](#) provide inclusive services across Europe and beyond to the benefit of science.

“EOSC - a tool for enabling Open Science in Europe”

<https://doi.org/10.5281/zenodo.4044009>



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



PaNOSC Goals



Full FAIR compliance of PaN scientific data

Support in shaping EOSC services for users needs

Increase of RIs' impact by encouraging data reuse

Innovative data services at RIs and as part of the EOSC

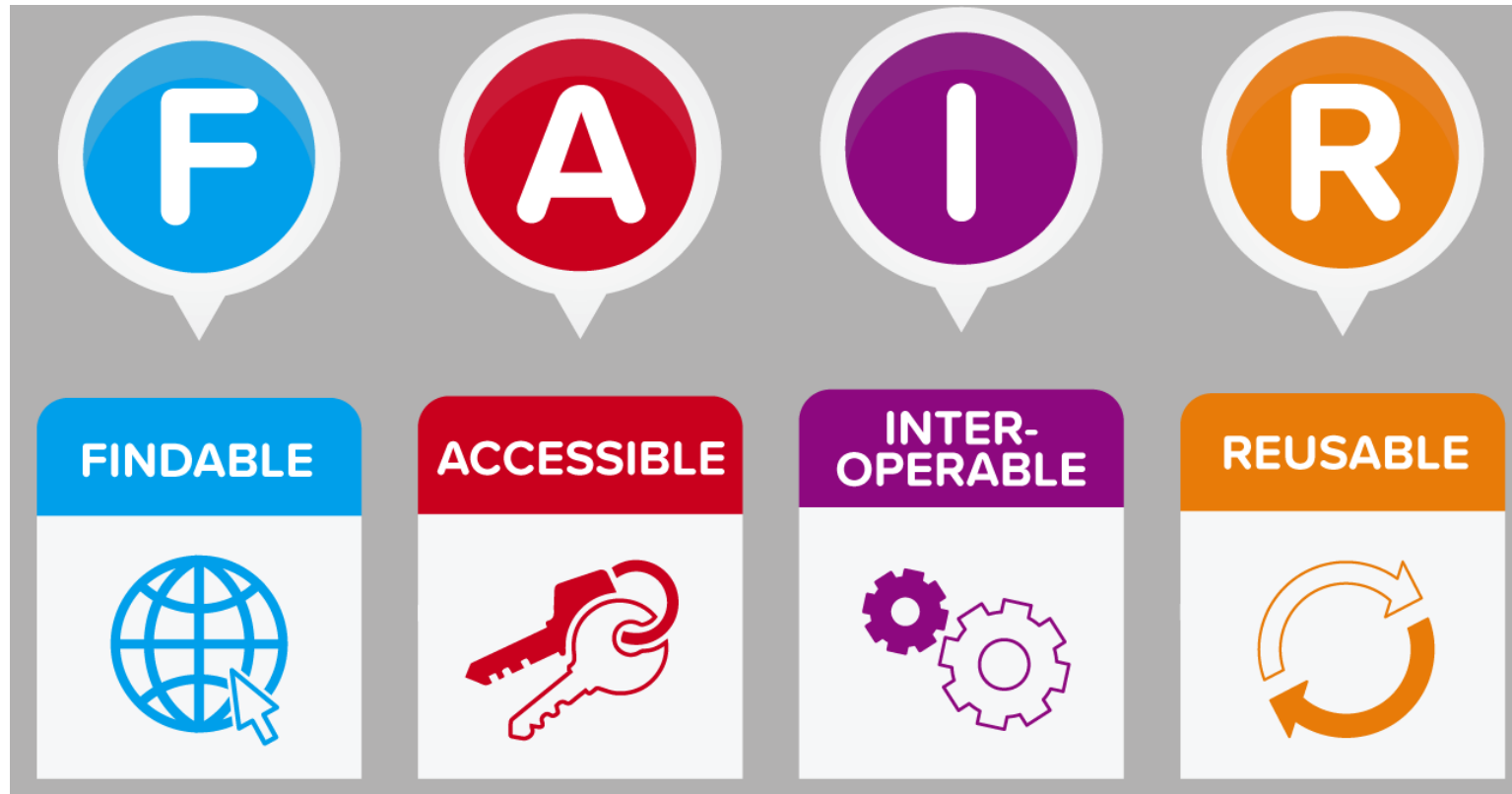
Sharing of best practices for open data policies

Collaboration with EOSC projects to share outcomes

PaNOSC works closely with the national PaN sources in Europe via ExPaNDS to develop common policies, strategies and solutions in the area of FAIR data policy, data management and data services, integrating t



Data Catalog Services – Essential for FAIR Data

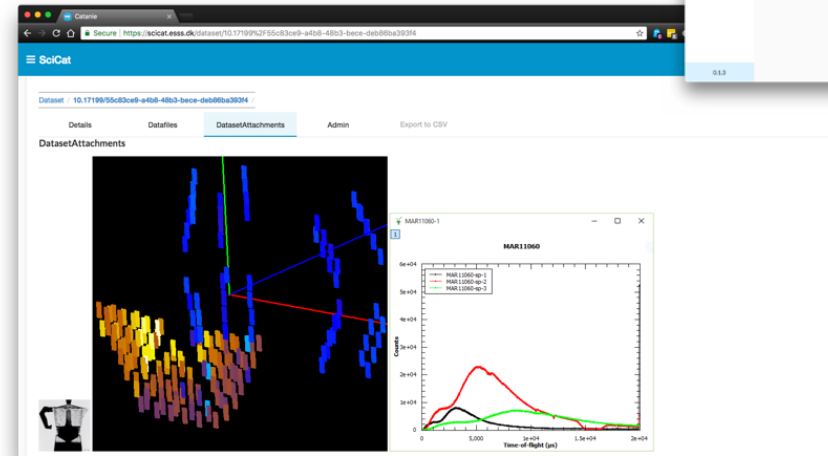


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852



WP3 Data Catalogues – Common Functions

- Store reference to raw data and enough metadata to identify the “right” dataset to work with.
- Information available often includes:
 - Proposal Information
 - Sample Information
 - Scientific Metadata
 - Previews
 - Provenance
 - Relationship to other data (derived data)
 - Relationship to people
- Catalogues also Facilitate Data Sharing and Publication



The screenshot shows a SciCat search results page with 'Filter Results: 60'. The table below lists dataset entries with columns for 'Dataset Name', 'Exp. ID', 'Acquisition Date', 'Status', 'Proposal ID', 'Archive Status', and 'Metadata Status'. The table contains 10 rows of data.

Dataset Name	Exp. ID	Acquisition Date	Status	Proposal ID	Archive Status	Metadata Status
HL-Siow05	15	2015-05-15T09:34:29.550Z	open05	128	archivedStatus	retiredStatus
HL-Siow16	20	2015-05-15T09:34:29.550Z	open16	128	archivedStatus	retiredStatus
HL-Siow03	10	2015-05-15T09:34:29.550Z	open03	128	archivedStatus	retiredStatus
HL-Siow10	20	2015-05-15T09:34:29.550Z	open10	133	archivedStatus	retiredStatus
HL-Siow13	20	2015-05-15T09:34:29.550Z	open13	138	archivedStatus	retiredStatus
HL-Siow08	55	2015-05-15T09:34:29.550Z	open08	158	archivedStatus	retiredStatus
HL-Siow08	48	2015-05-15T09:34:29.550Z	open08	158	archivedStatus	retiredStatus
HL-Siow14	24	2015-05-15T09:34:29.550Z	open14	137	archivedStatus	retiredStatus
HL-Siow01	11	2015-05-15T09:34:29.550Z	open01	124	archivedStatus	retiredStatus



Work Package Goals

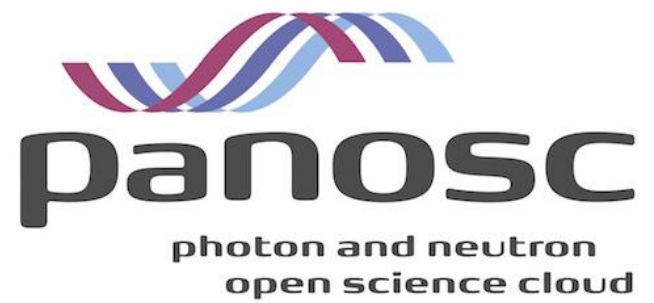
- Open Data – Make Existing and Future Datasets FAIR
 - How: Data Harvesting endpoints and Search API
 - Target: PaN users and Open Science communities at large
 - Why: Relevant data is the way into using PaNOSC services
- Improve Metadata Quality
 - Define common PaN Dictionaries (Ontologies) and Metadata Schemas
 - Increase and improve common use of the NeXus File Format
- Connecting Up
 - Production facilities (beamlines)
 - PaNOSC Service Portal (WP4 & WP6)
 - EOSC (B2Find, OpenAIRE)
 - Further Facilities – ExPaNDS
 - Metadata Bodies (RDA, NeXus)



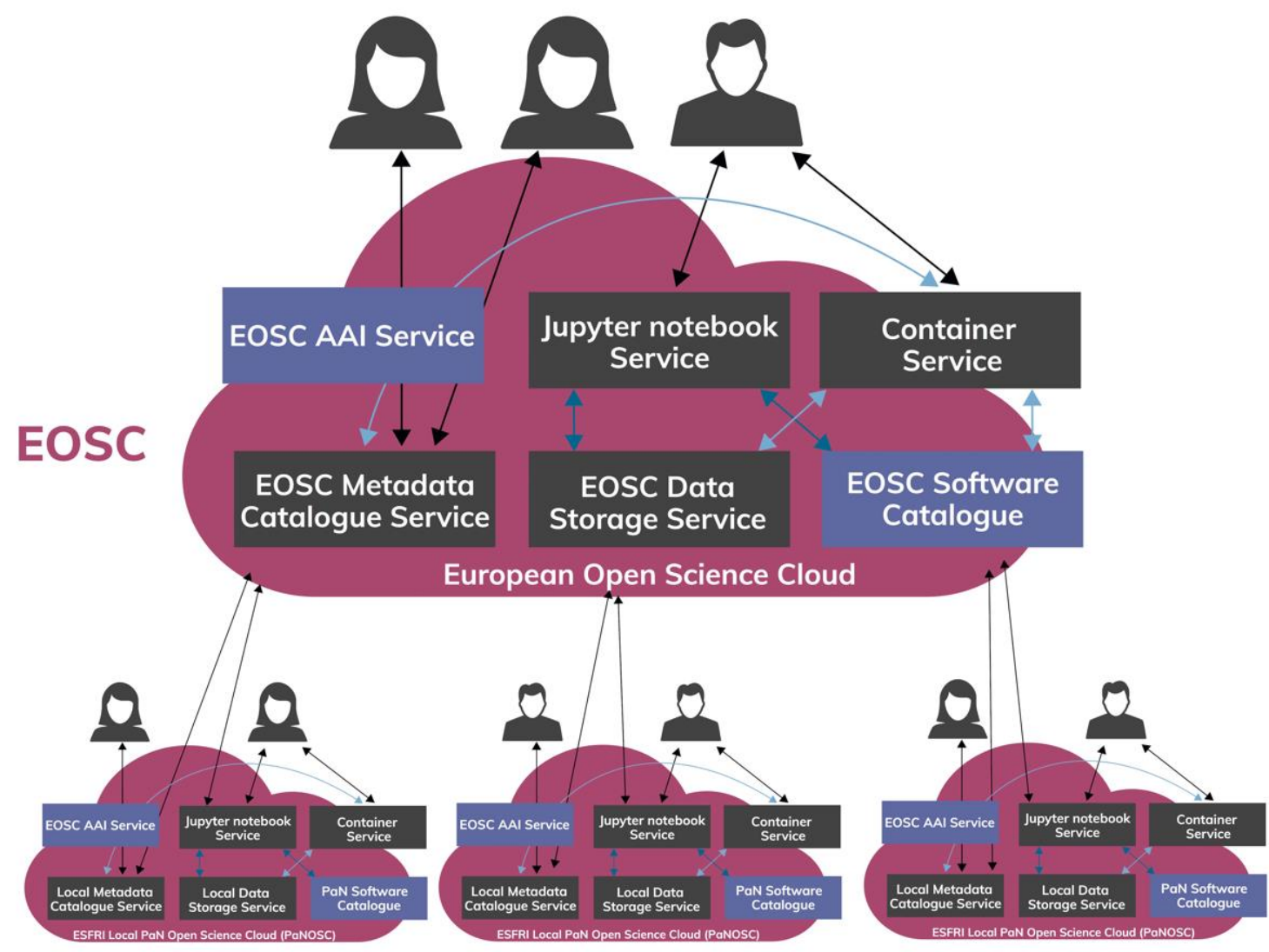


Common API for 6 catalogues

Common API to search across all PaNOSC catalogues



EOSC + PaNOSC Clouds



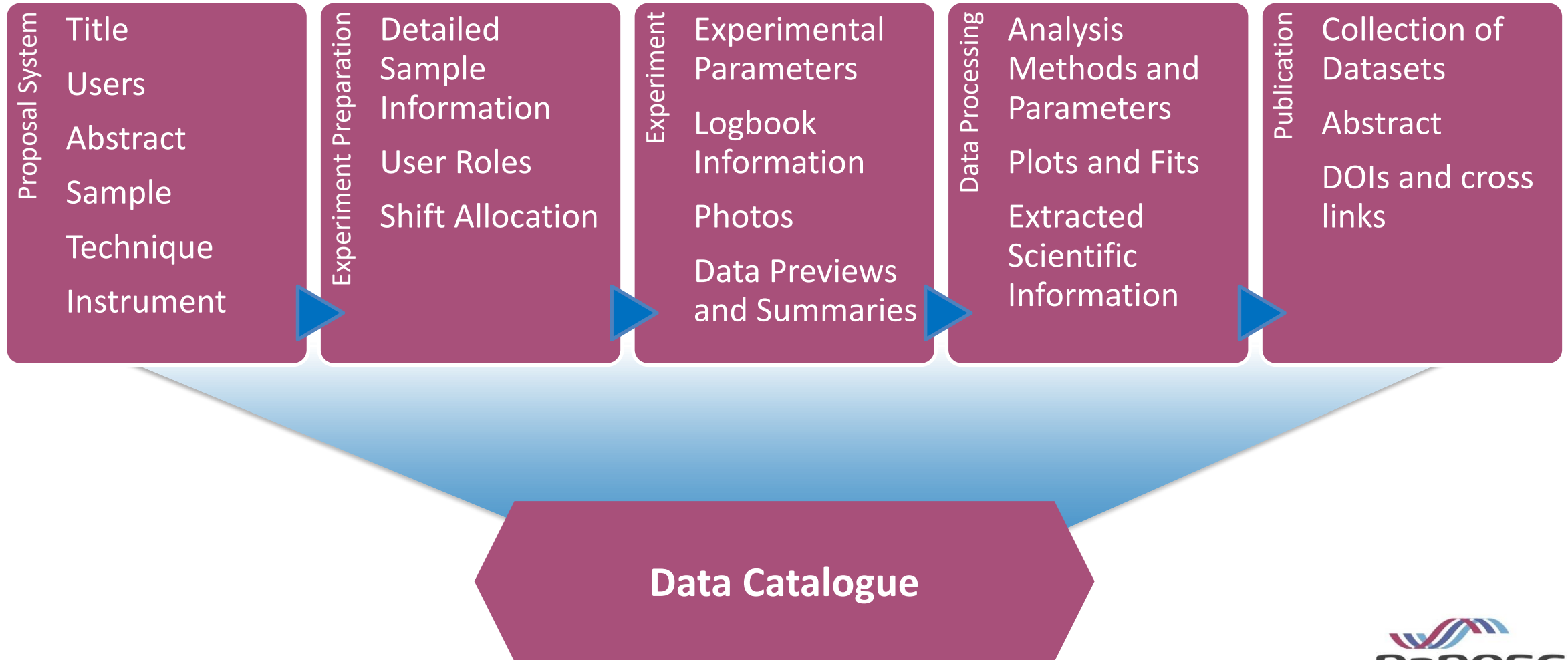
How do PaN facilities work?

- 1. User:**
has an idea / need to study a sample
- 2. Proposal:**
User writes a proposal for one of the PaN facilities
- 3. Review committee:**
Reviews proposal and rates scientific quality
- 4. Beamline scientist:**
Review proposal and checks feasibility
- 5. Beamtime allocated:**
User travels to facility / sends sample
- 6. Experiment:**
Sample(s) are exposed to beam + data collected
- 7. Analysis:**
Data is reduced, analysed + curated (DOI)
- 8. Publication:**
User publishes results (DOI) in peer review journal



Opportunities to Integrate Catalogues or to Capture Metadata

Not an exclusive list. Every facility is different.



<https://panosc.eu/events/panosc-wp3-catalogue-integration-best-practices-meeting/>

The screenshot shows a web browser window with the URL <https://www.panosc.eu/events/panosc-wp3-catalogue-integration-best-practices-meeting/>. The page header features the PaNOSC logo and the text "photon and neutron open science cloud". A navigation menu includes links for About, Data, Services, Training, Materials, News & Events, and EOSC. The breadcrumb trail is "Home > Events > PaNOSC WP3 Catalogue Integration Best Practices Meeting".

The main content area is titled "Events" and prominently displays the date "18/05/21" in large blue text. Below this, the event title "PaNOSC WP3 Catalogue Integration Best Practices Meeting" is shown. A text block states: "Save the date for the PaNOSC WP3 Catalogue Integration Best Practices Meeting, taking place online on 18-20 May 2021." Another text block explains: "In PaNOSC [Work Package 3 – Data Catalogue Services](#), the partners create means for users and third parties to find datasets from photon and neutron sources using domain-specific search terms." A black button labeled "Timetable" is positioned at the bottom left of the main content area.

On the right side, there is a sidebar with the event title and a "Read →" link. Below this, three categories are listed with icons and "Read →" links: "Data Storage" (with an E=mc² icon), "Data Analysis" (with a magnifying glass icon), and "Data Simulation" (with a pendulum icon). At the bottom of the sidebar, a "Data Catalogue" link is partially visible with a "Read →" link.

A dark cookie consent banner is at the bottom of the page, containing the text "By using this site, you accept the use of cookies and third-party cookies [Learn more](#)" and an "Accept" button.

**Let's have interesting
presentations and discussions!**

tobias.richter@ess.eu

lajos.schrettner@eli-alps.hu