Scientific Information Service

Addressing the challenges of research data management, reuse and collaboration: the case for CERN Analysis Preservation and FAIR data services.

Sunje Dallmeier-Tiessen, Pamfilos Fokianos and Artemis Lavasa @CERN

May 2023

“…and the results of its experimental and theoretical work shall be published or otherwise made generally available”

CERN Founding Convention (1953)

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

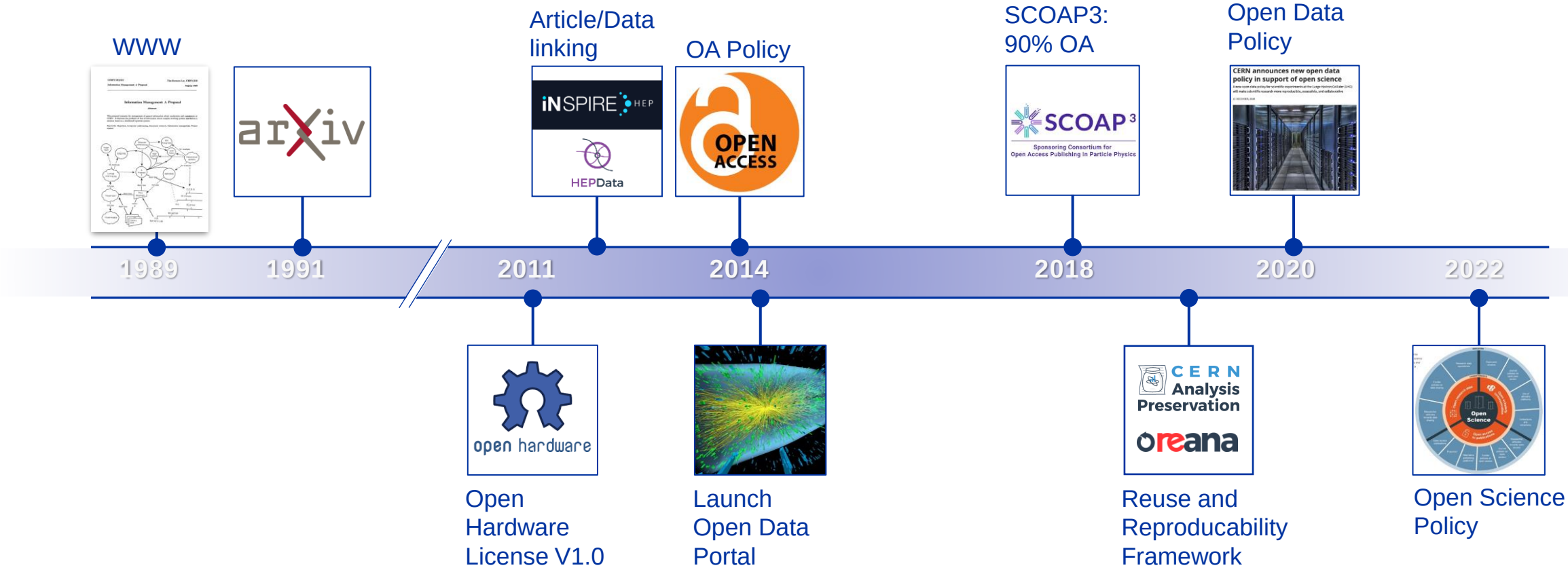
CONVENTION

FOR THE ESTABLISHMENT OF A EUROPEAN ORGANIZATION
FOR NUCLEAR RESEARCH

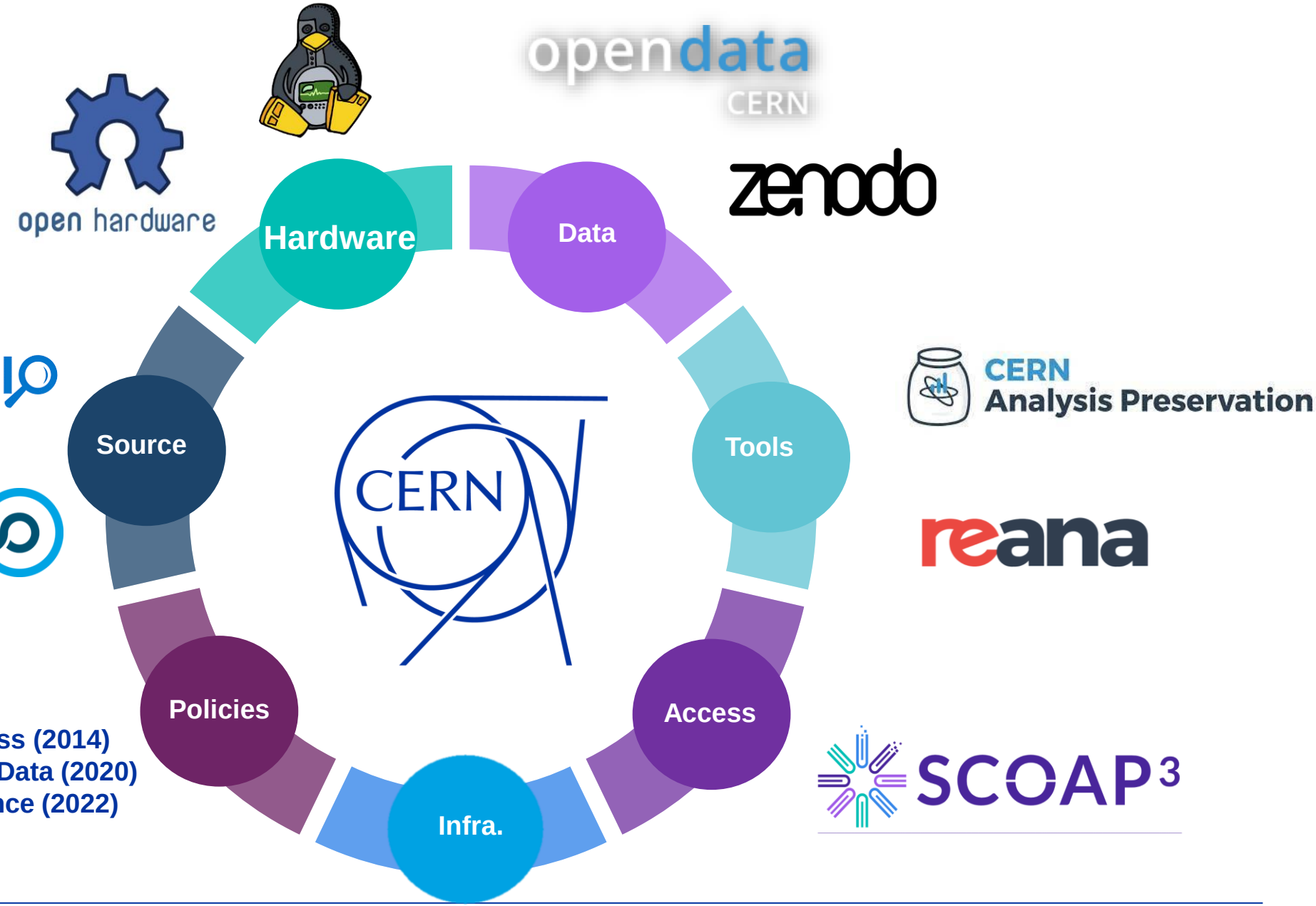
PARIS, 1st JULY, 1953

Illustration by Stephanie van de Sandt

CERN – on the path to universal Open Science

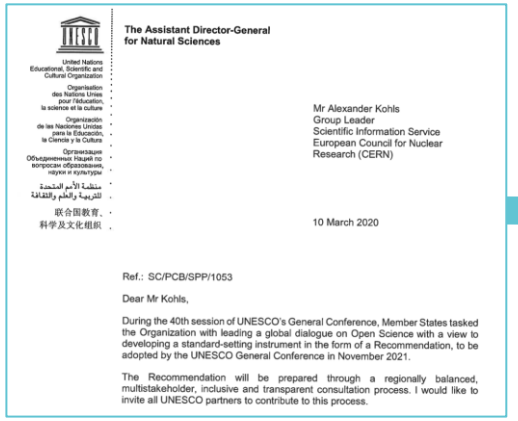


Open Science “products” at CERN



Open Access (2014)
LHC Open Data (2020)
Open Science (2022)

UNESCO Recommendations on Open Science created the momentum



The Assistant Director-General
for Natural Sciences

United Nations Educational, Scientific and Cultural Organization

Organisation des Nations Unies pour l'éducation, la science et la culture

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

Организация Объединенных Наций по вопросам образования, науки и культуры

منظمة الأمم المتحدة للتربية والعلم والثقافة

联合国教育、科学及文化组织

Mr Alexander Kohls
Group Leader
Scientific Information Service
European Council for Nuclear Research (CERN)

10 March 2020

Ref.: SC/PCB/SPP/1053

Dear Mr Kohls,

During the 40th session of UNESCO's General Conference, Member States tasked the Organization with leading a global dialogue on Open Science with a view to developing a standard-setting instrument in the form of a Recommendation, to be adopted by the UNESCO General Conference in November 2021.

The Recommendation will be prepared through a regionally balanced, multistakeholder, inclusive and transparent consultation process. I would like to invite all UNESCO partners to contribute to this process.

Pushing the Boundaries of Open Science at CERN: Submission to the UNESCO Open Science Consultation

July 2020

Kamran Naim, Tullio Basaglia, Jelena Brankovic, Pamfilos Fokianos, Jose Benito Gonzalez Lopez, Maria Grazia Pia, Alexander Kohls, Artemis Lavasa, Lars Holm Nielsen, Stephanie van de Sandt, Javier Serrano, Tim Smith

CERN, the European Organization for Nuclear Research, is the world's largest high-energy physics (HEP) laboratory. Since its founding in 1954, the Laboratory has made significant contributions to our understanding of the world and the universe. The mission of the Organization is to: provide a unique range of particle accelerator facilities that enable research at the forefront of human knowledge, perform world-class research in fundamental physics, and unite people from all over the world to push the frontiers of science and technology, for the benefit of all. Supported through a global partnership of 23 member states, CERN is home to the world's largest scientific instrument, the Large Hadron Collider (LHC), and hosts over 12,000 scientists and engineers from across the world.

The frontier research conducted at CERN has long embodied the values that have more recently come to be defined as the Open Science movement, which describes "research and development that is collaborative, transparent and reproducible and whose outputs are publicly available" (European Commission (2018)). Indeed these values were enshrined in 1953 in CERN's founding Convention, which states that "...the results of its experimental and theoretical work shall be published or otherwise made generally available," (CERN Council (1953)) providing the Organization with an early Open Science manifesto.

The embrace of the values of Open Science at CERN is not considered an obligation, but rather an expression of the collective moral and financial responsibility to member states and the global scientific community to advance the boundaries of human knowledge. The community at CERN helped build the Open Internet, were early adopters of Open Source, helped usher in the preprint culture, and helped to pioneer initiatives to allow Open Access to scientific publications. Yet, the complexity of the Laboratory makes the comprehensive pursuit of Open Science an ongoing and ever-evolving effort. Supported by long-term human and financial investments by the international community, CERN hosts dynamic research teams consisting of thousands of researchers from around the world, who collectively publish almost 1,000 research articles per year. These scientists are engaged in multiple experimental collaborations that operate largely independently of each other and have different tools and practices. The high-energy physics research conducted at CERN is one of the most data-intensive branches of science: particle collisions at CERN produce about 90 petabytes of collision data per year. The complexity, scale and value of these data presents unique and unprecedented challenges for the research community at CERN. The responsible stewardship, analysis and preservation of these data requires a supporting organizational research culture, but also a range of tools and services to optimize how data are used, verified, and built upon to advance knowledge.

On June 19th, 2020, CERN's highest governing authority, the CERN Council announced its updated strategy for particle physics in the global landscape. This document, the European Strategy for Particle Physics represents the most important strategic document for CERN, outlining the future strategy for the Organization. Open Science was given a strong endorsement from the CERN Council, with the strategy stating that; "European science policy is quickly moving towards Open Science, which promotes and accelerates the sharing of scientific knowledge with the community at large. Particle physics has been a pioneer in several aspects of Open Science. The particle physics community should work with the relevant authorities to help shape the emerging consensus on Open Science to be adopted for publicly funded research, and should then implement a policy of Open Science for the field"(European Strategy Group (2020)).

This paper aims to describe the ecosystem of initiatives, projects and technologies that have been developed at CERN to maximize the impact of our research through building an Open Science infrastructure that is effective, collaborative, and responsive to the needs of the scientific community. We aim to demonstrate that despite the complexity of the research undertaken at CERN, Open Science can be advanced through concerted efforts and as such, the CERN example could serve as an inspiration for the global scientific community.

United Nations Educational, Scientific and Cultural Organization

Organisation des Nations Unies pour l'éducation, la science et la culture

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

Организация Объединенных Наций по вопросам образования, науки и культуры

منظمة الأمم المتحدة للتربية والعلم والثقافة

联合国教育、科学及文化组织

ESOF2020 EUROSCIENCE OPEN FORUM TRIESTE

EuroDIG European Dialogue on Internet Governance

Towards a Global Consensus on Open Science

Online Regional Consultation for Western Europe and North America to the UNESCO Recommendation on Open Science

Online Consultation on Open Science

UNESCO invites scientists, publishers, science policymakers or anyone with experience and interest in Open Science to participate in the online consultation on implications, benefits and challenges of Open Science across the globe.

https://doi.org/10.17181/CERN.1SYT.9RGJ

CERN is proud to have joined UNESCO on the journey towards Recommendations on Open Science

Policy framework for Open Science at CERN

CERN Open Access Policy (2014)

- All CERN research articles published OA (CC-BY)
- Central fund available
- Different routes (SCOAP^3, Read & Publish, APC payment)

LHC Open Data Policy (2020)

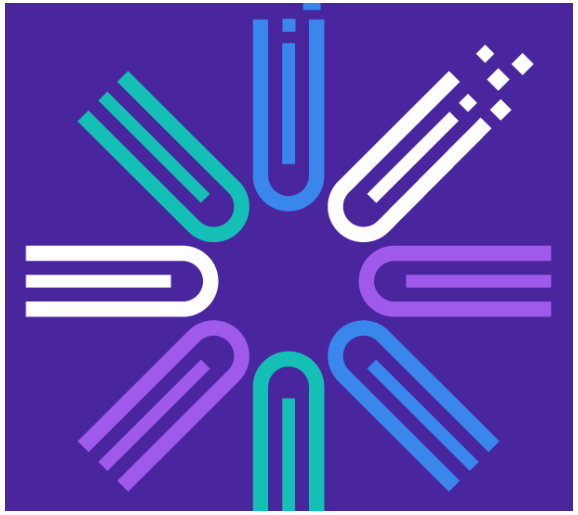
- 4 LHC collaborations will release all level 3 data (+ level 1 and 2)
- Gradual release will start ~5 years after collection
- Other experiments to follow

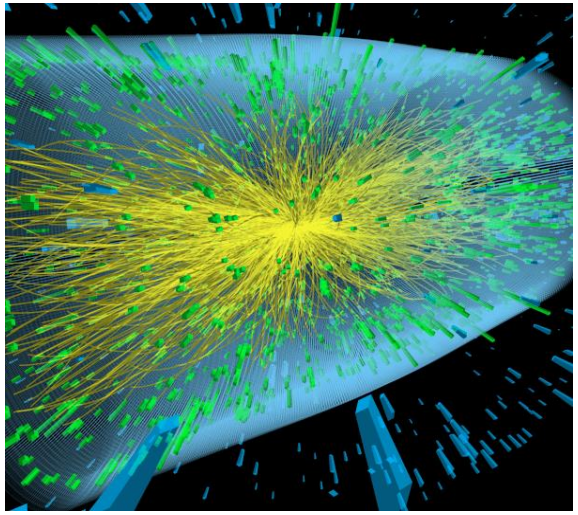
European Strategy for Particle Physics(2020)

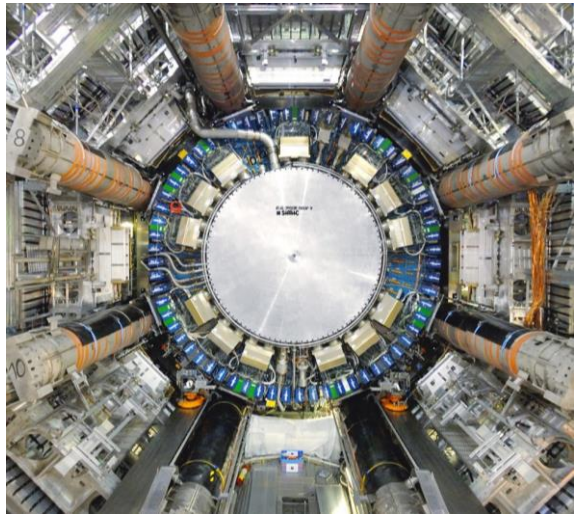
- OS recognized as organizational issue for the discipline
- Should develop and implement an OS policy for the field

Funder Open Science Policies

- Funding agencies supporting experimental collaborations have specific open data requirements
- CERN will establish central support office for compliance



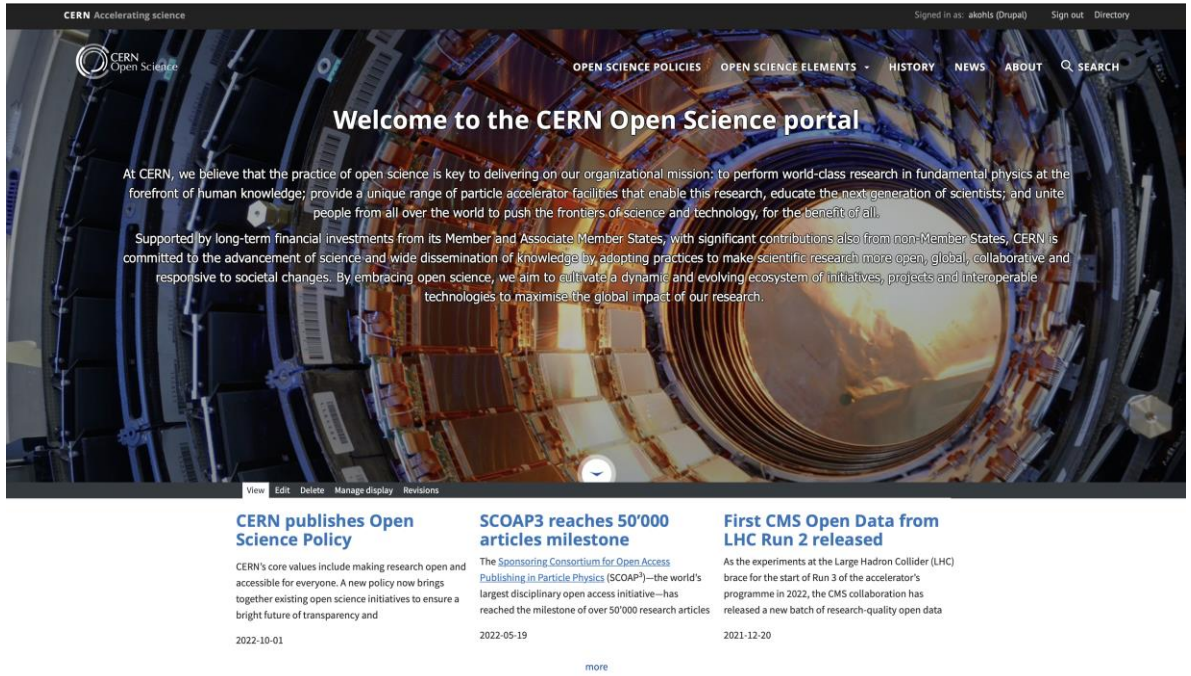




CERN Open Science Policy Published October 1st, 2022

After 12 months of consultations and collaborative drafting, CERNs first institutional Open Science Policy was formally adopted by CERN Council

- August/September: Presentations and discussions with directorate
- 29th September: Policy presented to CERN Council
- 1st October: Policy formally in place



More information: https://openscience.cern

CERN Open Science Policy

- Captures current practice and states progressive vision across multiple Open Science domains:
  - Open Access to Publications
  - Open Research Data
  - Open Software
  - Open Hardware
  - Research Integrity, Reuse & Reproducibility
  - Infrastructure for Open Science
  - Research Assessment & Evaluation
  - Education, Training & Outreach
  - Citizen Science
- Policy to be regularly updated to reflect changes in landscape, practices, funder requirements & community demands
- Policy and its implementation plan are developed and governed by the community.
- V1.0, Oct 2022: https://cds.cern.ch/record/2835057

Research integrity, reuse and reproducibility

5. Research integrity, reuse and reproducibility

CERN is committed to ensuring the integrity of research. In order to facilitate the reuse of its research products, CERN provides infrastructures to accommodate the scale and complexity of its research outputs. Reuse and reproducibility are facilitated by practising comprehensive analysis preservation to capture relevant research objects, such as research data releases with supporting metadata, auxiliary data, linked software, reproducible analysis workflows, documentation, etc.

https://cds.cern.ch/record/2835057

CERN Analysis Preservation: what is it?

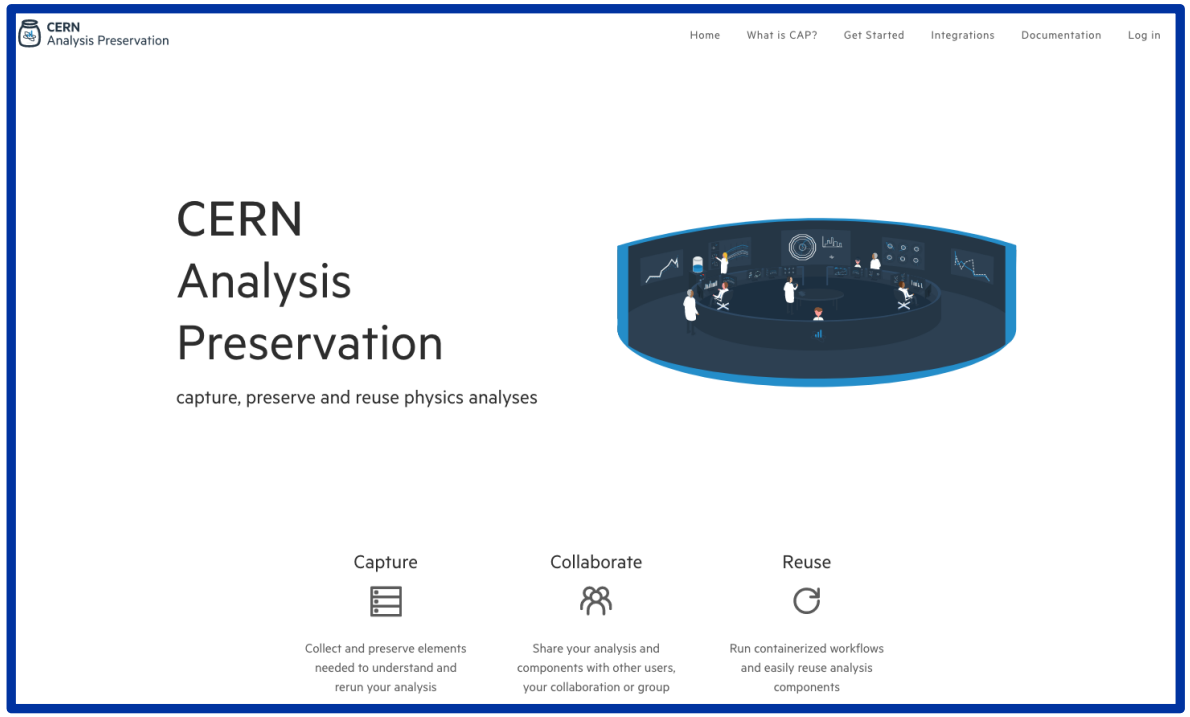
Flexible and collaborative tool to link and preserve “everything” around an analysis, metadata, data, software...

Version content and metadata. Link persistently all the elements of an analysis needed to understand and rerun an analysis several years later.

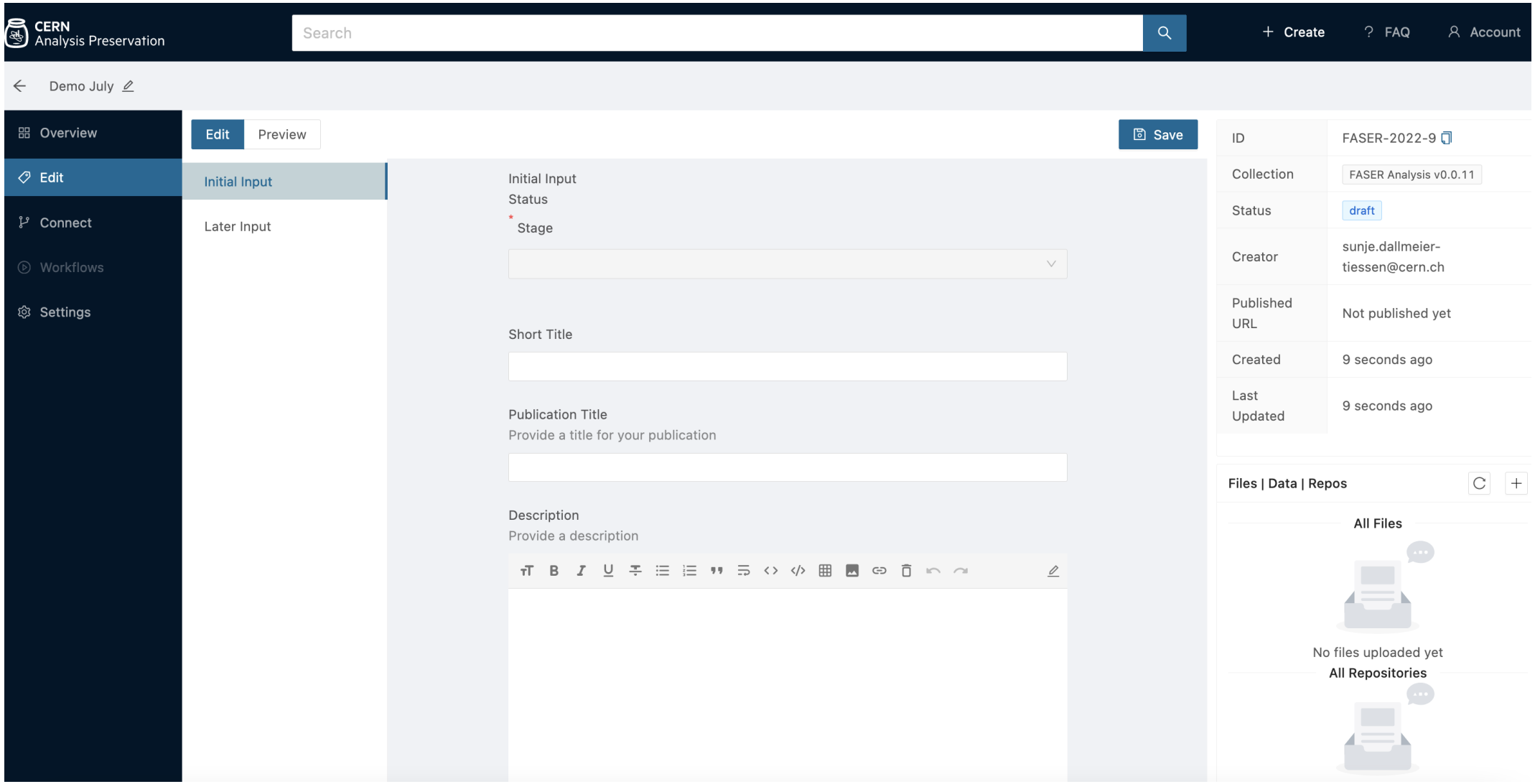
Standardize analysis components so that they are reusable (ex. scripts or CI/CD, writing tools, workflow engines, push to other services).

Ensure that users are always in control of when and if their work is shared.

Accommodate the needs of each collaboration or team to integrate into local tools



Starting an analysis in CERN Analysis Preservation



FAIR @ CERN Analysis Preservation

Findable

Each analysis with a unique ID, with rich metadata, that faciliates advanced search

Accessible

Metadata are accessible and retrievable (if permitted by the experiment/team)

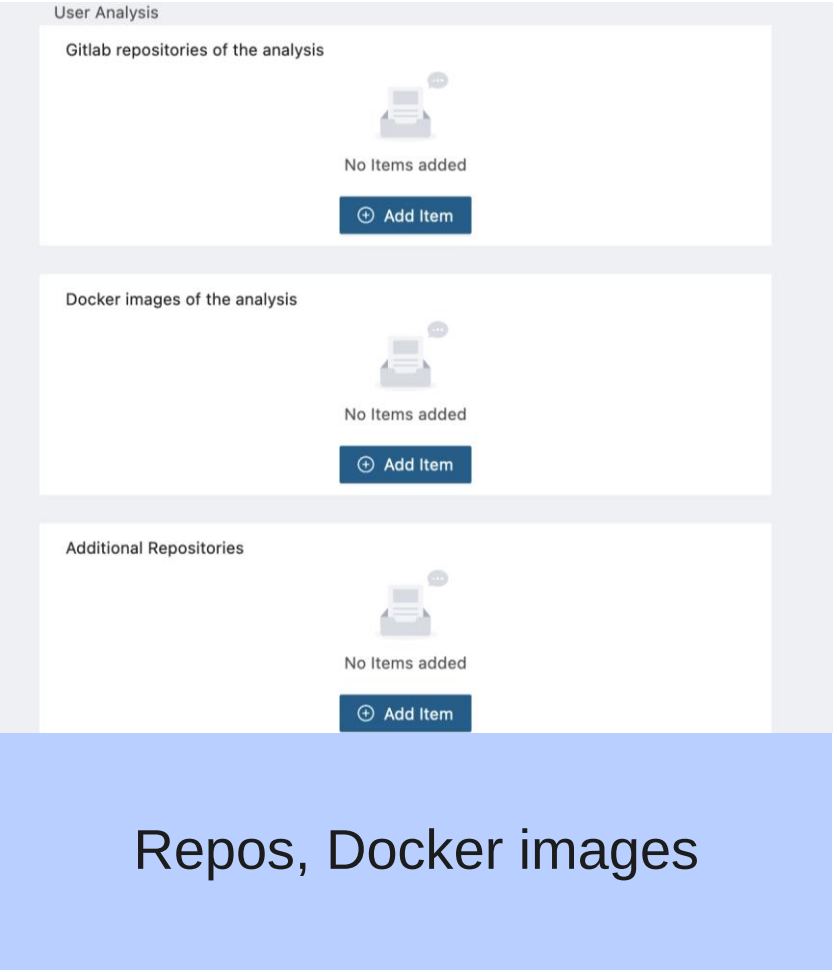
Interoperable

JSON schemas. Challenge are diverse community terminologies…

Reusable

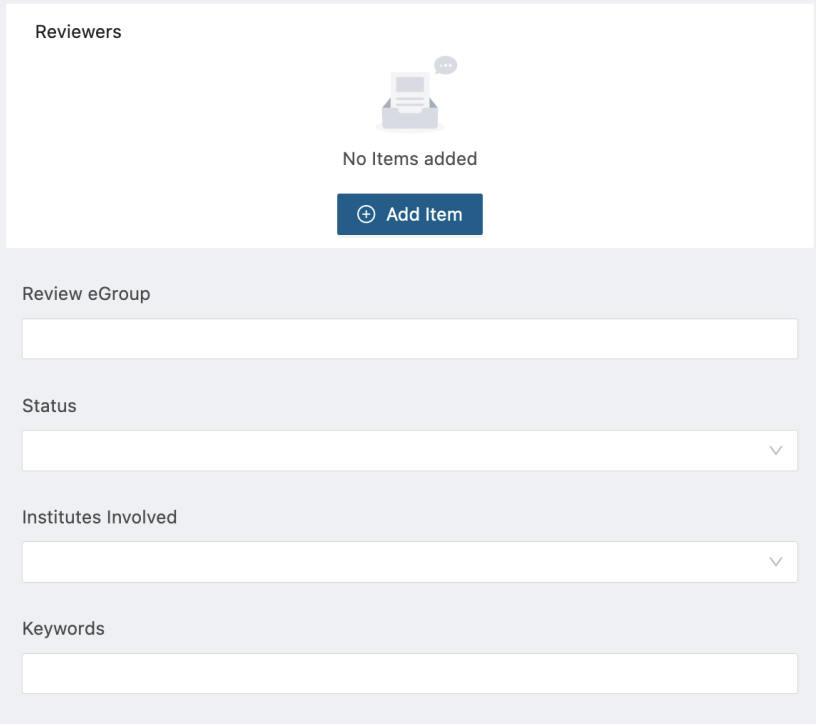
Rich metadata for each analysis, i.e. automated ingestion from experiment tools, contextual linking

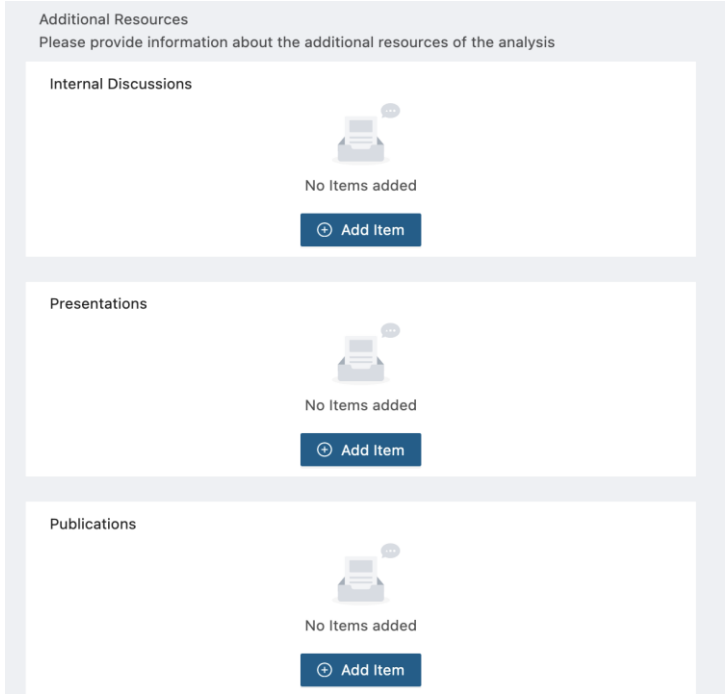
CAP: preserving context



Repos, Docker images

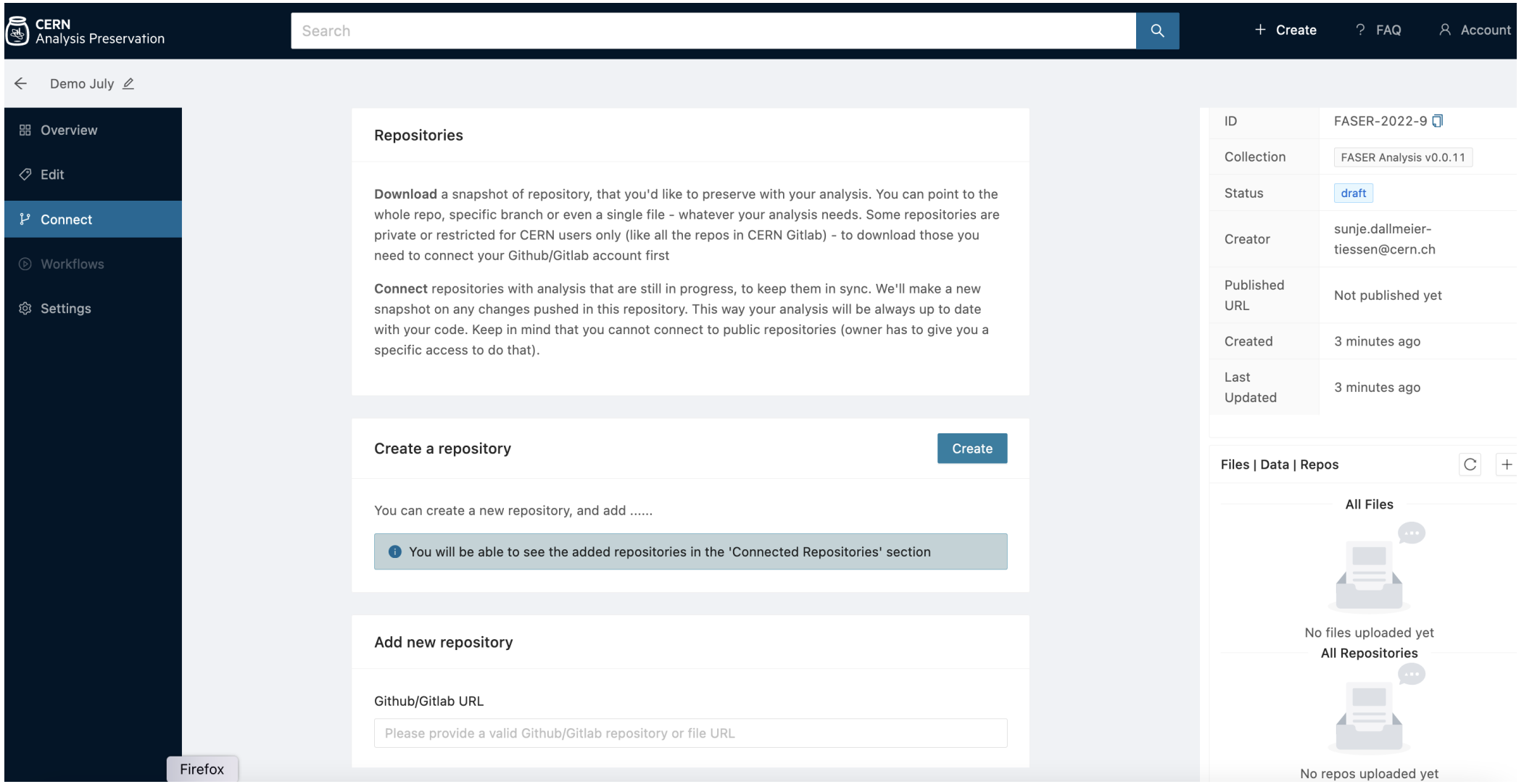
Metadata



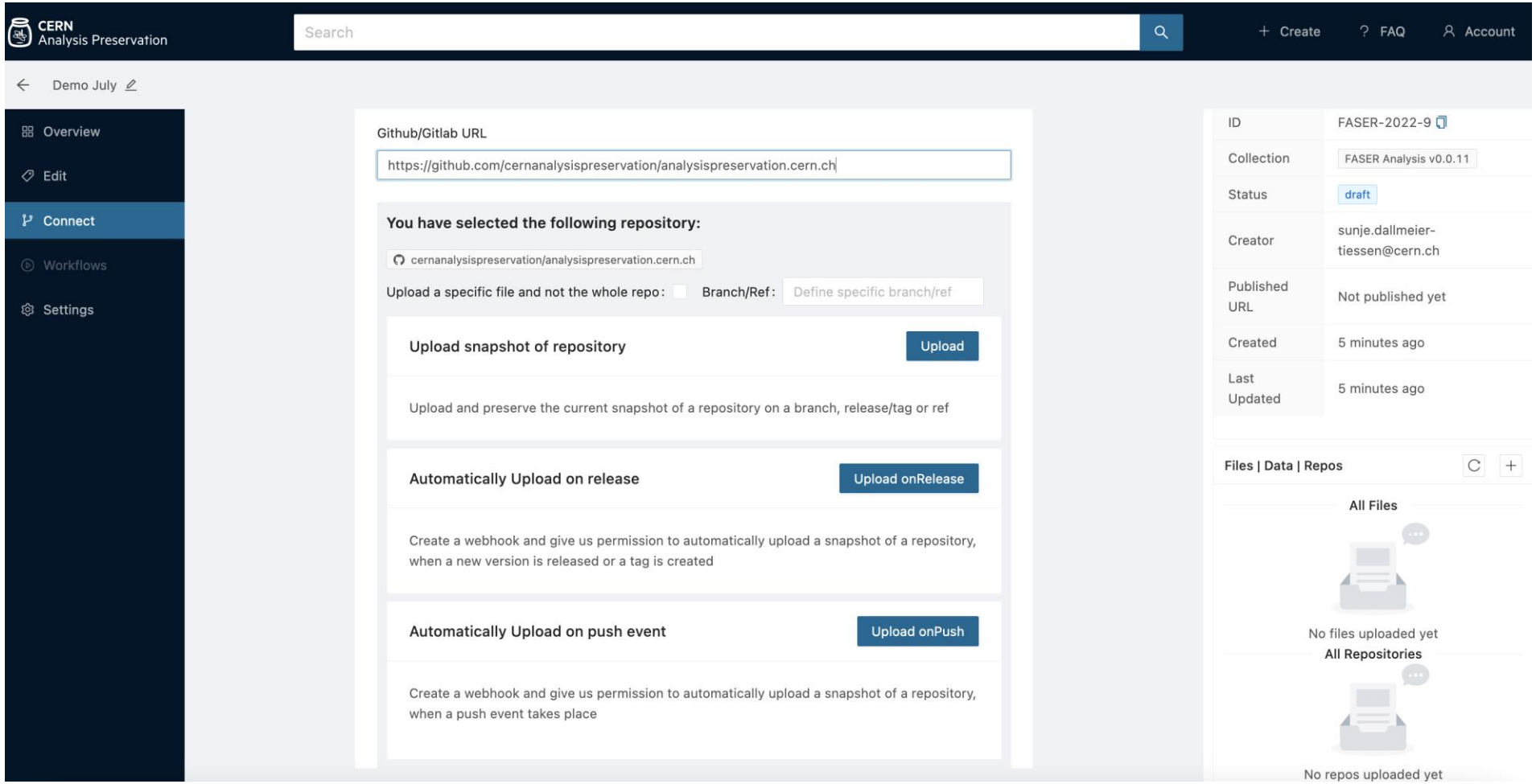


Presentations, Discussions

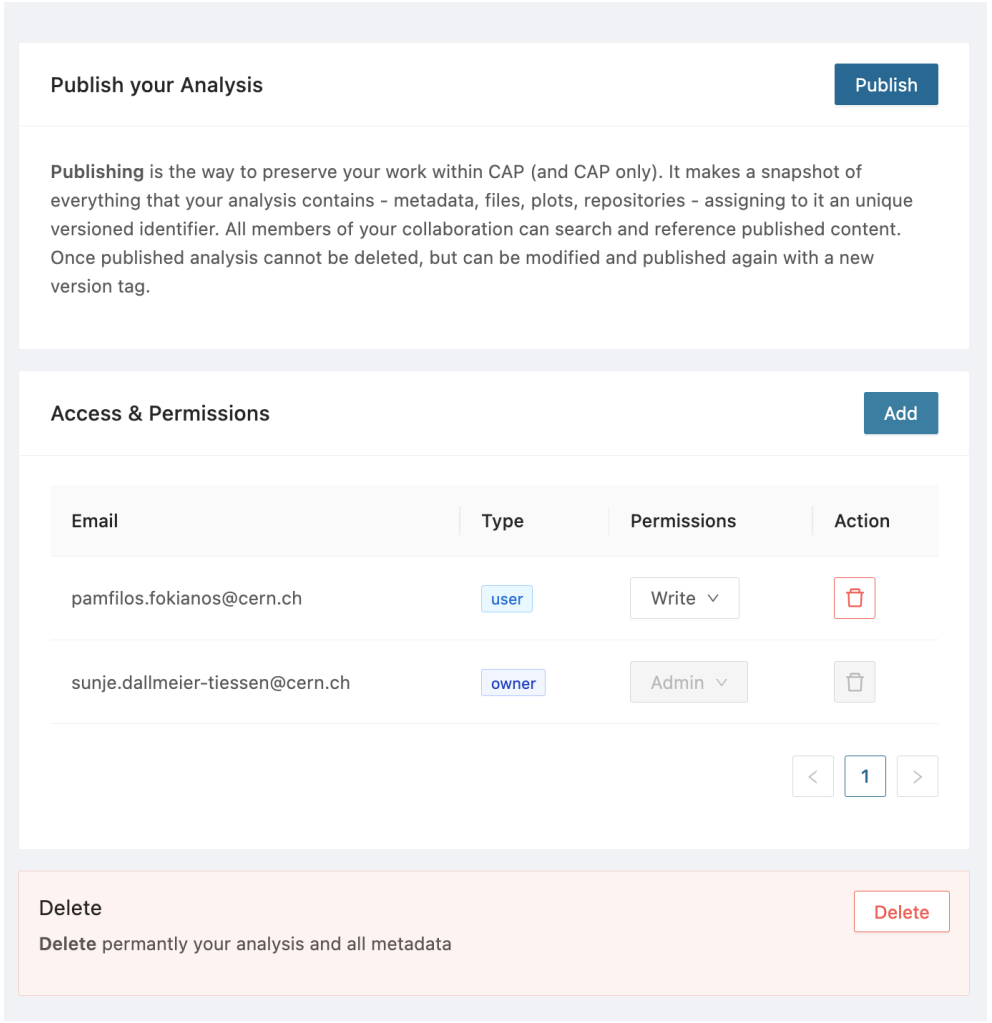
Preserving repositories

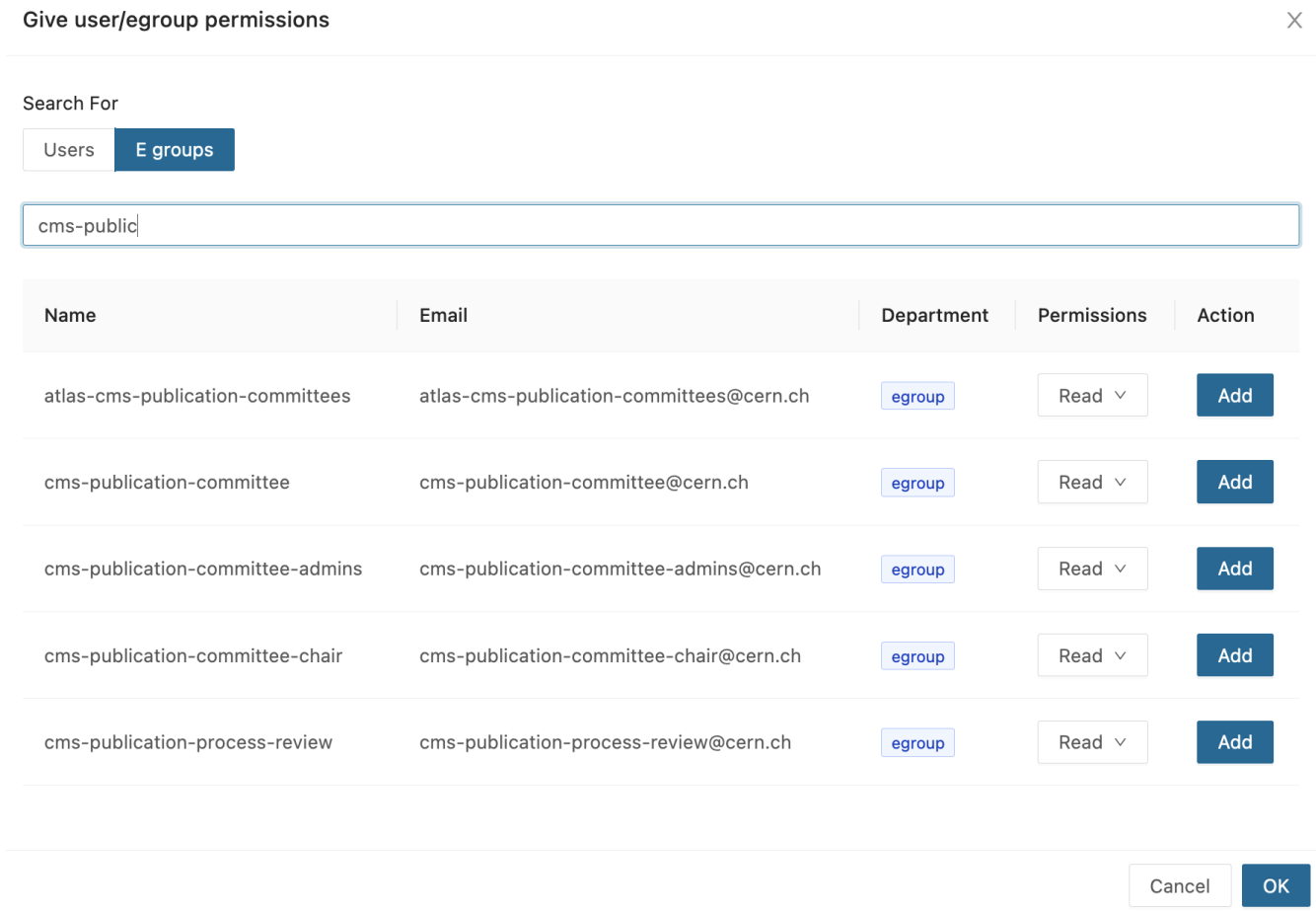


Capturing repositories

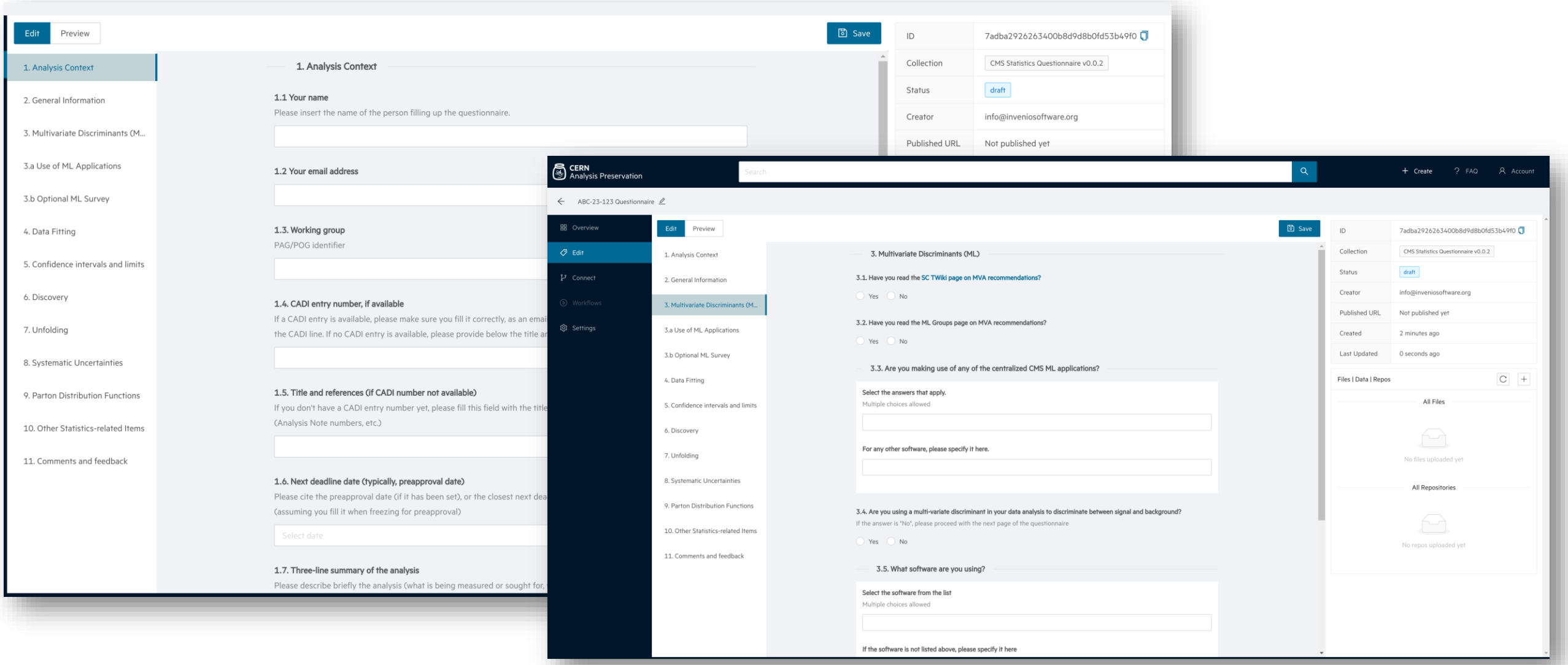


Full control for the users: permissions and reviews





CAP is already being used by some experiments



Control centre for a project or an experiment (aka the “admin panel”)

What is this?

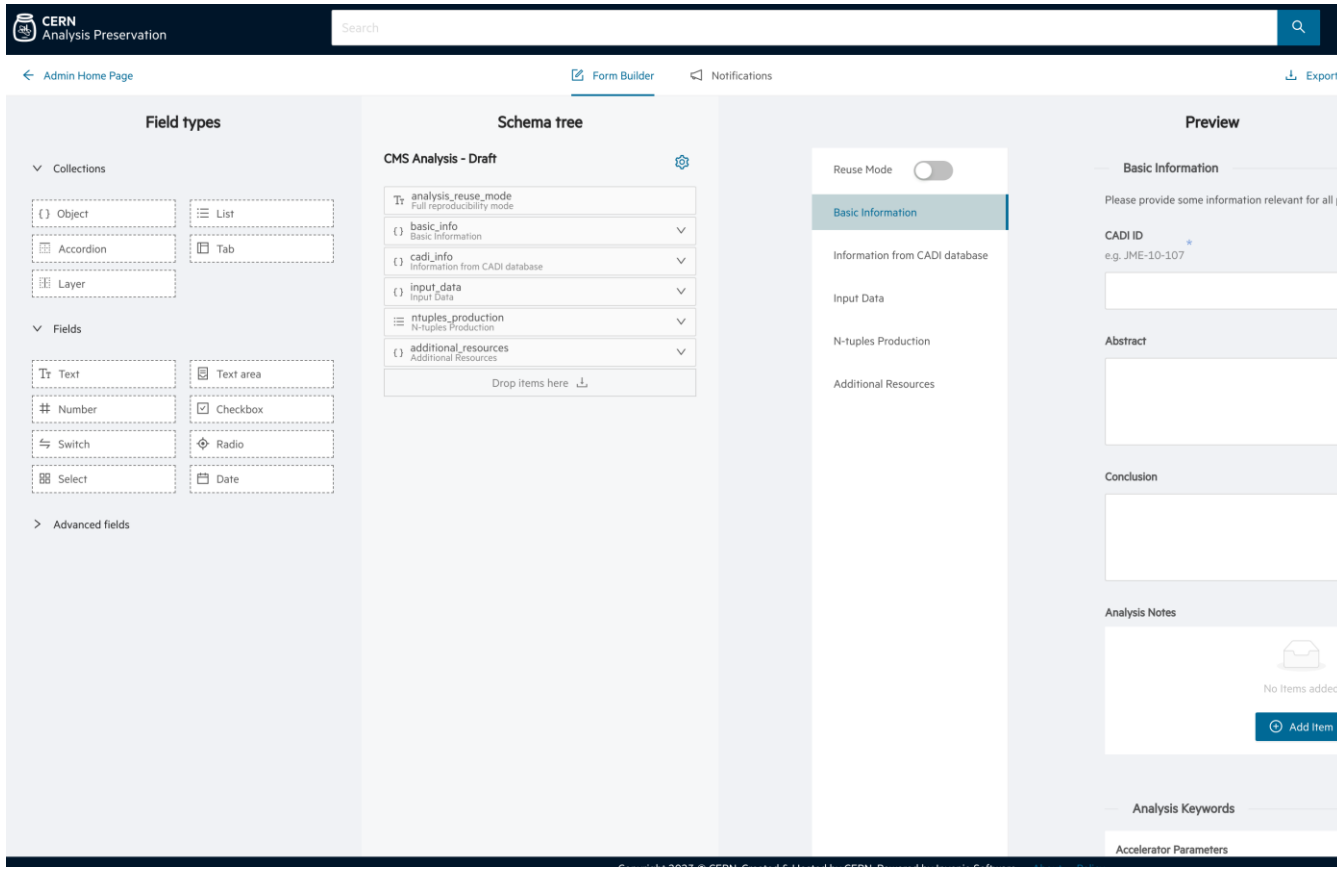
Easier control over communities, schemas, interfaces

Who is this for?

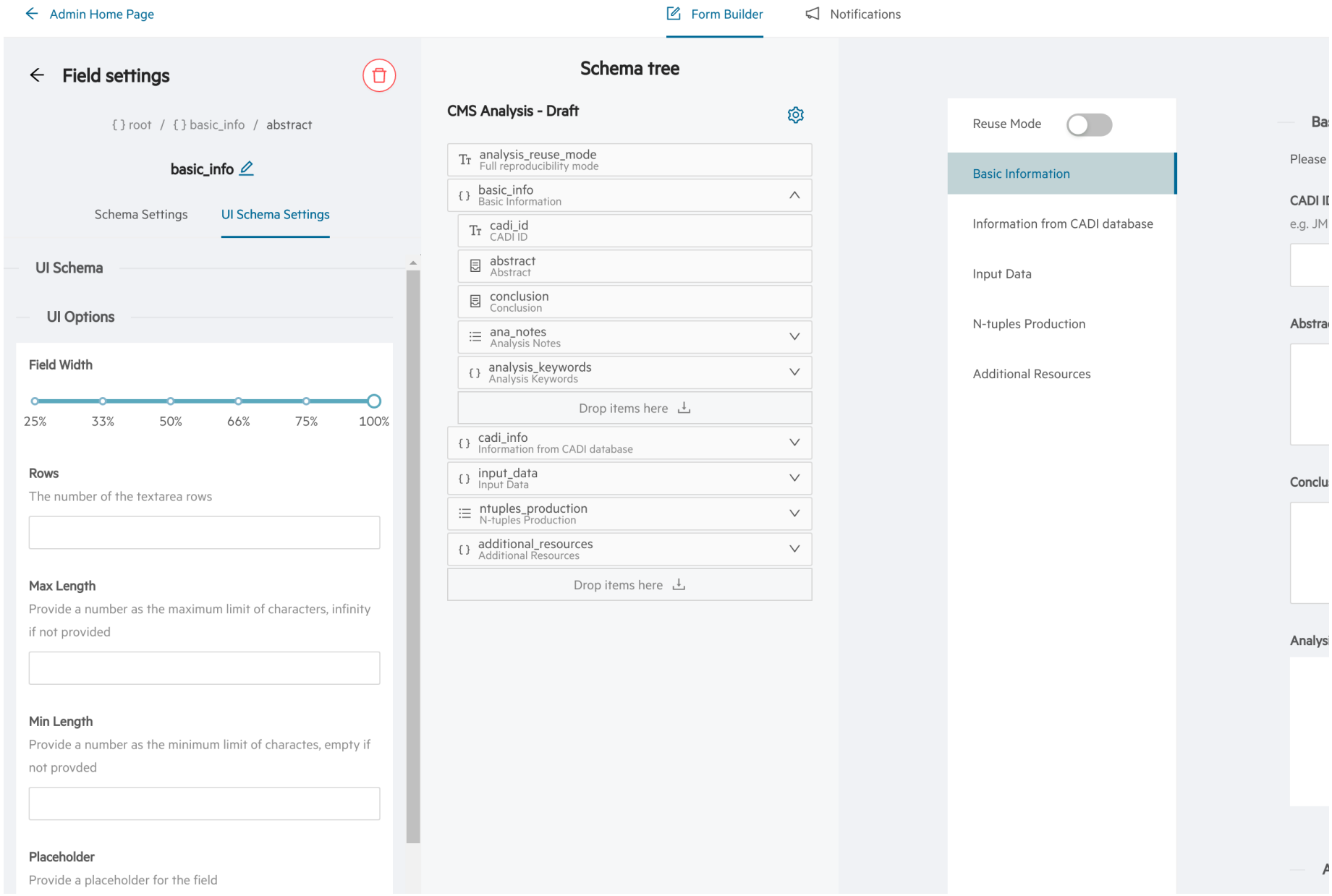
Conveners, managers, boards…

Why?

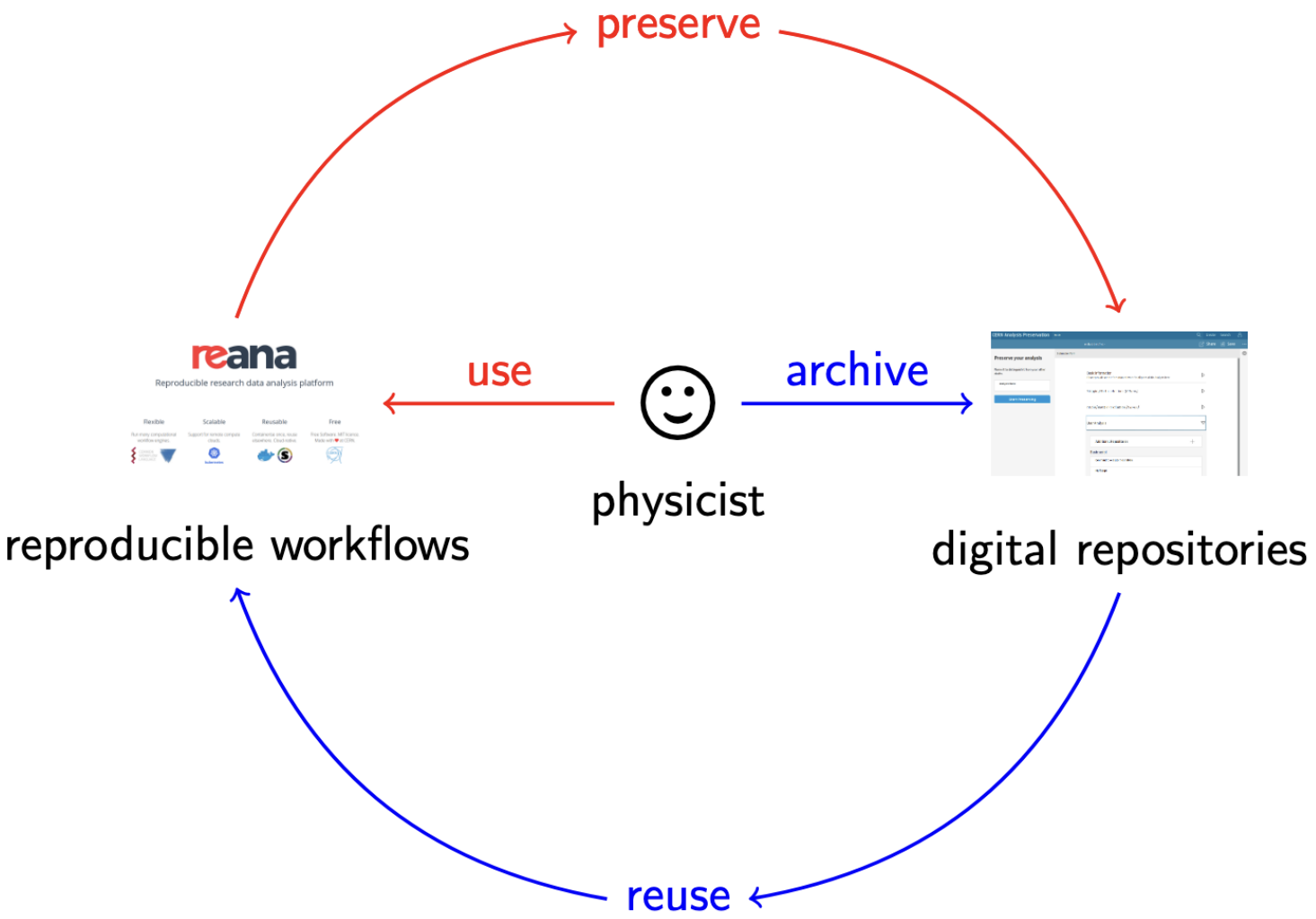
No need to ask us for smaller changes, you got full control over communities and settings



Example admin panel for a CMS analysis



Reproducibility ⇌ Preservation



@tiborsimko

Conclusions

- CERN Analysis Preservation a tool to manage, preserve and share internally research “products” around an analysis
- Preserving integrity of research, enable FAIR research, and prepare for Open Science
- Already being used by some experiments

Future:

- Monitor and enhance FAIR developments
- Integrate more with the experiments, i.e. exploit potential of “admin” panel
- Integrate with REANA

Research integrity: access and reusability

Foundation of research integrity: accessibility of all components, quality assurance, reusability

Enabling research integrity and openness through preservation the “entire” analysis workflow. Ensuring the value of our research beyond contract durations, experiment lifetime etc.

Many funders, political bodies and research organisations stressed the importance of research integrity, open science and FAIR

Thank You!

Questions?
analysis-preservation-support@cern.ch

Scientific
Information
Service

scientific-info.cern