PAMFILOS FOKIANOS

CERN ANALYSIS PRESERVATION PORTAL

IMPROVE PRESERVATION AND RE-USE OF RESEARCH RESULTS AT CERN

CERN
Analysis Preservation

CERN ANALYSIS PRESERVATION PORTAL

tool for physicists to capture, preserve and make their analysis components reusable for the future

ANALYSIS PRESERVATION @ CERN

The Analysis Preservation effort aims at responding to two parallel demands:

(1) Inside of the Collaborations:
The high complexity of the analyses create major challenges in terms of capturing and preserving the analysis and the knowledge around it.

+ CERN Open Data Policy for the LHC

+ Landscape project initiative

(2) Externally:
Increasing number of funding agencies have put in place data management policies demanding the development of comprehensive data management frameworks for data and knowledge preservation, and for future reuse (or even reinterpretation and reproducibility) of research outcomes.

CERN

ANALYSIS PRESERVATION @ CERN

In general, preservation efforts and needs at CERN change from one experiment to another.

Even between the same collaboration, there are different working groups that work and do analyses in their own way.

CERN ANALYSIS PRESERVATION PORTAL

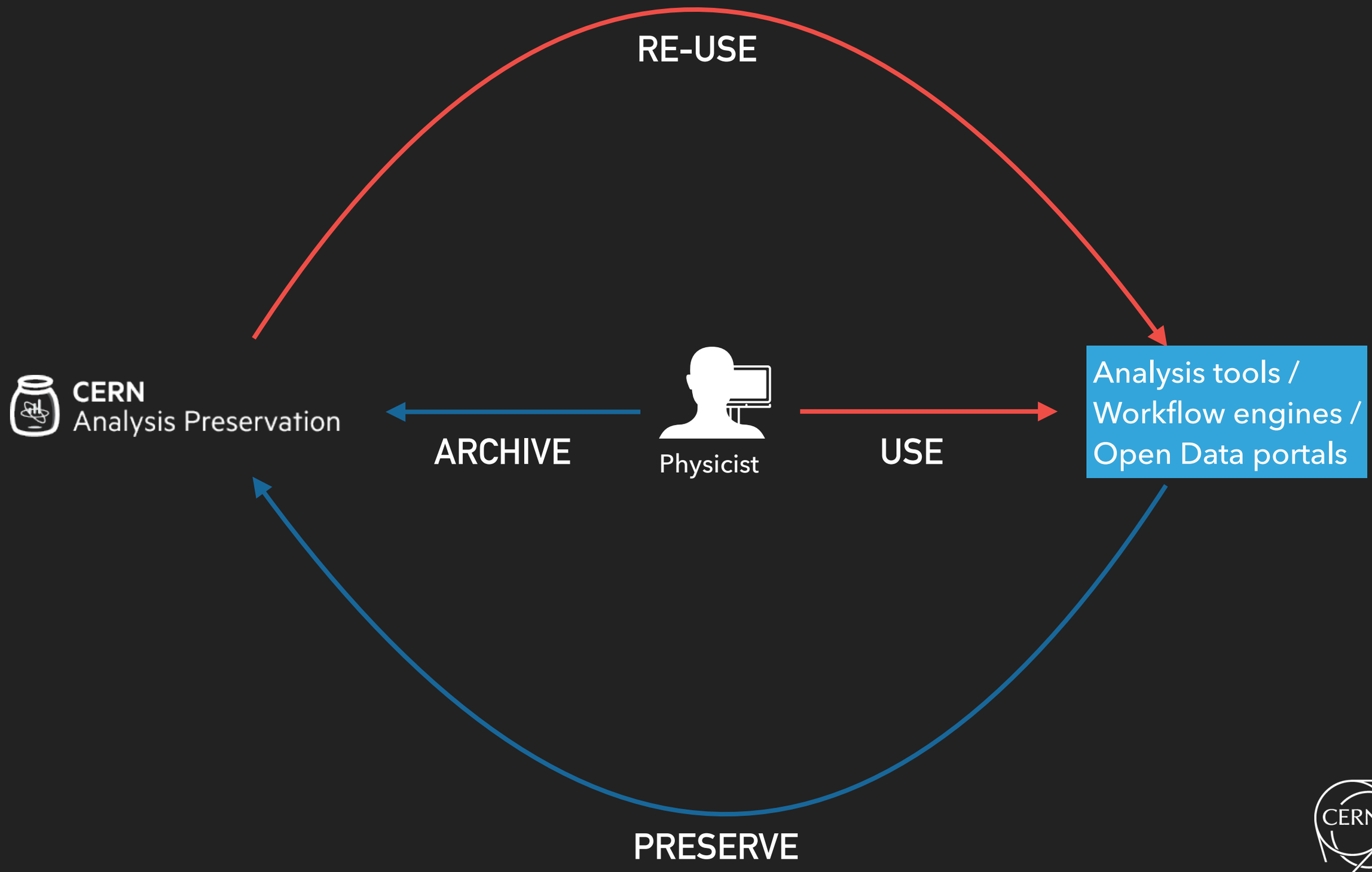
With CERN Analysis Preservation, we have built a tool that tries to accommodate these needs, depending on the specifications and requirements of individual groups (e.g. they can have their own data models/collections, to preserve their analysis knowledge and make it reusable and understandable)

- **Goal**: capture as many elements needed to understand and reiterate an analysis even several years later and link them together persistently
- **Goal**: make analysis components easily reusable (e.g. from scripts or CI/CD, writing tools, workflow engines, push to other services)

CERN ANALYSIS PRESERVATION PORTAL

- Flexible data models (JSON-Schemas), FAIR practices
- Customisable forms
- Versioning of metadata & files, using the publishing draft/record model
- Review process
- Integration with related scientific services and universal identifiers (i.e. Github, Gitlab, Zenodo, ORCID, ROR, etc)
- Sync with and indexing of various experiment DBs/APIs/sources for searching and integration
- Powerful API (records, files, users, search)

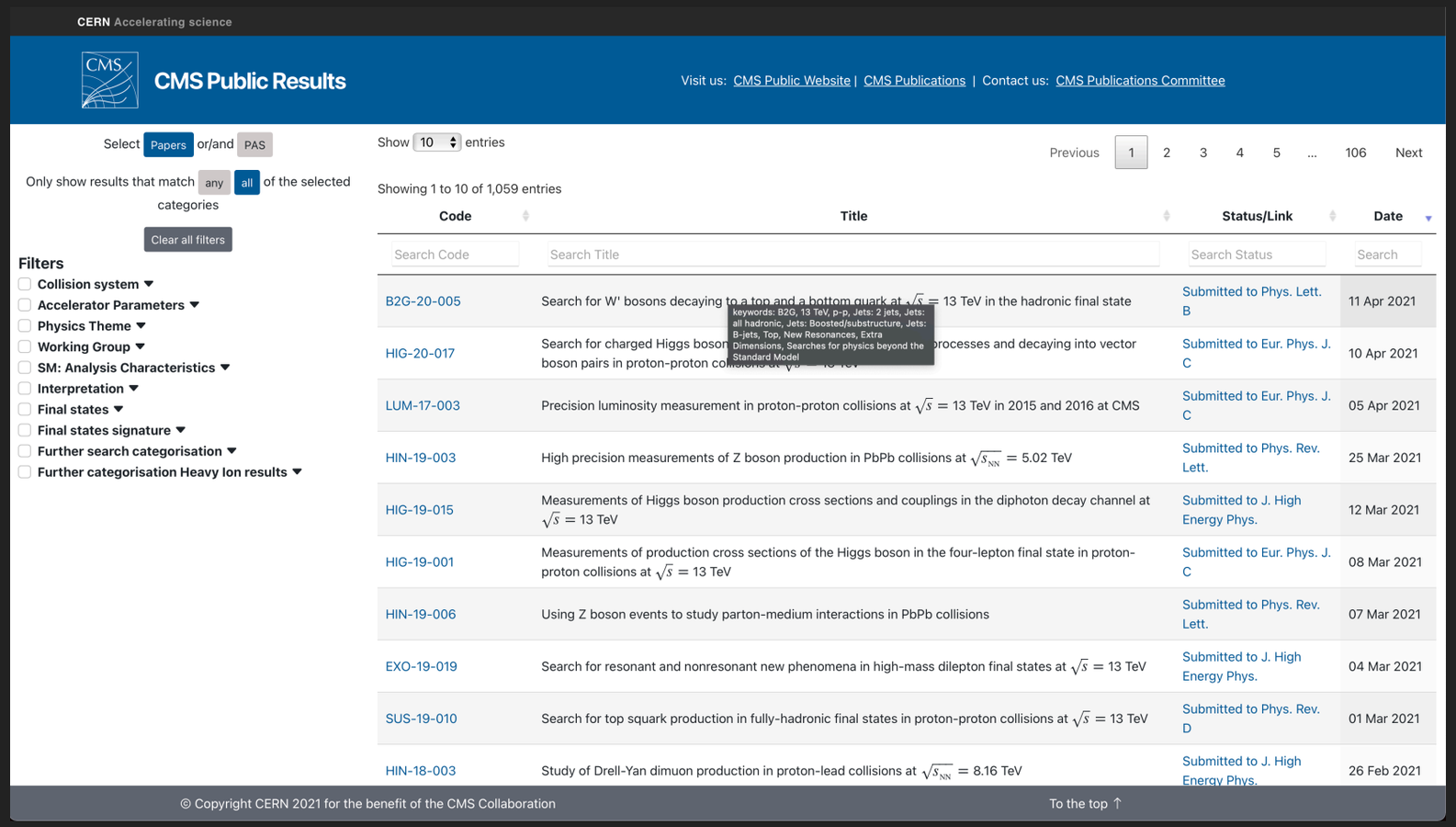
PRESERVATION ⇄ RE-USABILITY



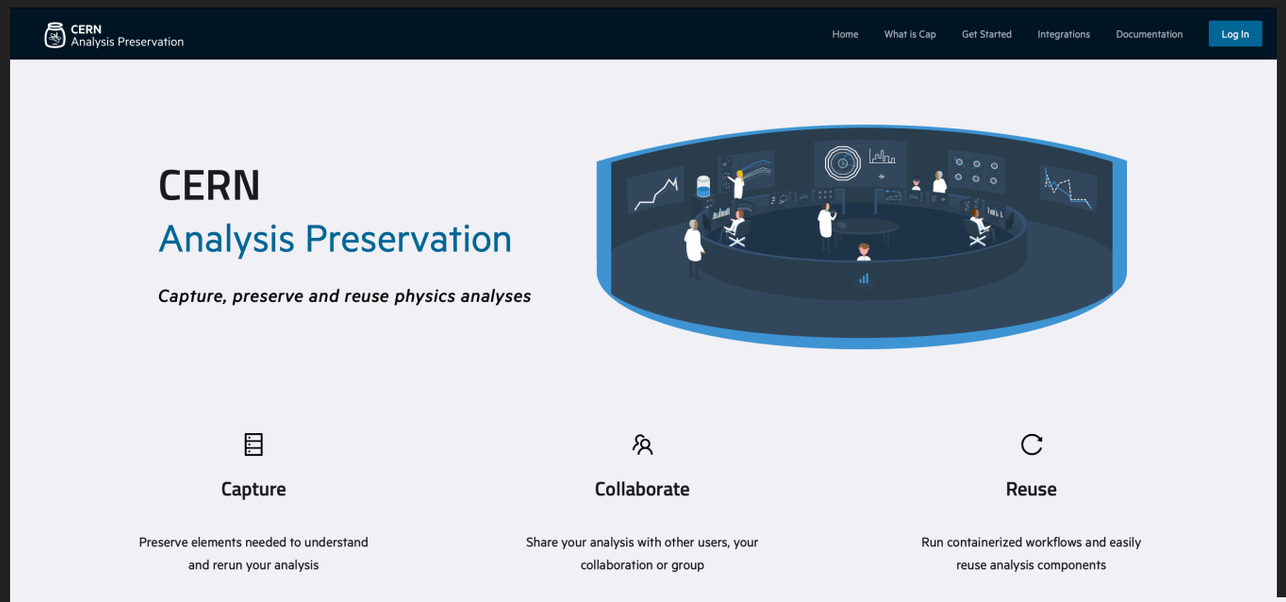
USE CASE: REUSE

Reuse of CAP data

- CMS public results page
  - Example of reuse of information
  - Dedicated website
  - Fetching data from CAP REST API



CERN ANALYSIS PRESERVATION PORTAL



What

Metadata

Your analysis description, input data, sources, referenced analyses, collaborators

Files

Plots, tables, formulas, likelihoods

Code

Scripts, instructions, repositories

Workflows

Containerized images, workflows

Documentation

Publications, presentations, conferences, notes

How

Web Interface

Login from your browser and explore all the features

Command Line

Use our command line client to automate the preservation process and make it part of your everyday work cycle

RESTful Interface

Integrate CERN Analysis Preservation with your existing services and tools by using our REST API

Who

Researchers

It doesn't matter on which stage of your analysis you are, it's never too early or too late to preserve your work

Collaborators

Share your work with others and invite them to contribute

Reviewers

Access analyses and all their components from one central place

Students

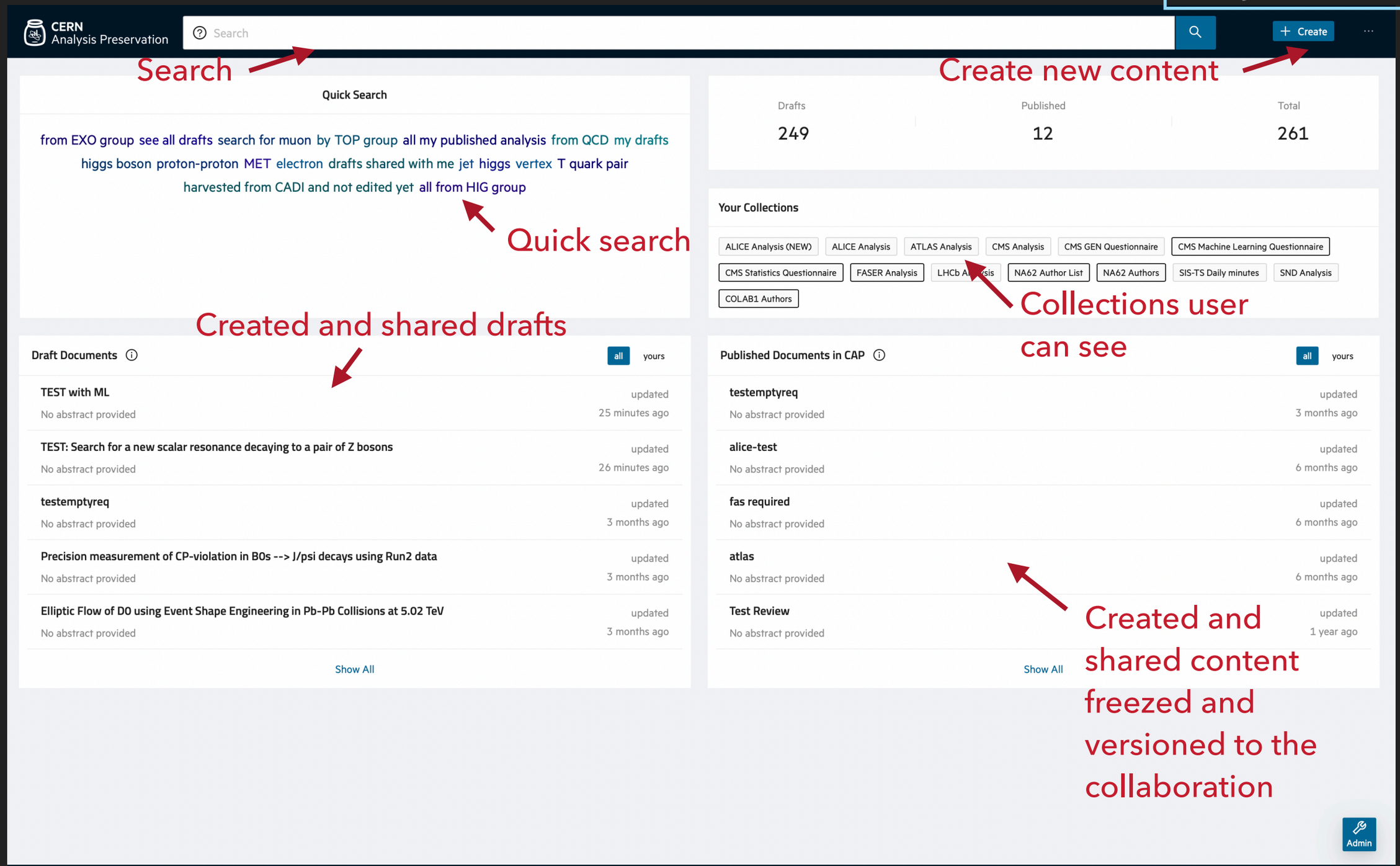
Search through older analyses, share with your supervisors, and preserve your work so that it never gets lost

DASHBOARD

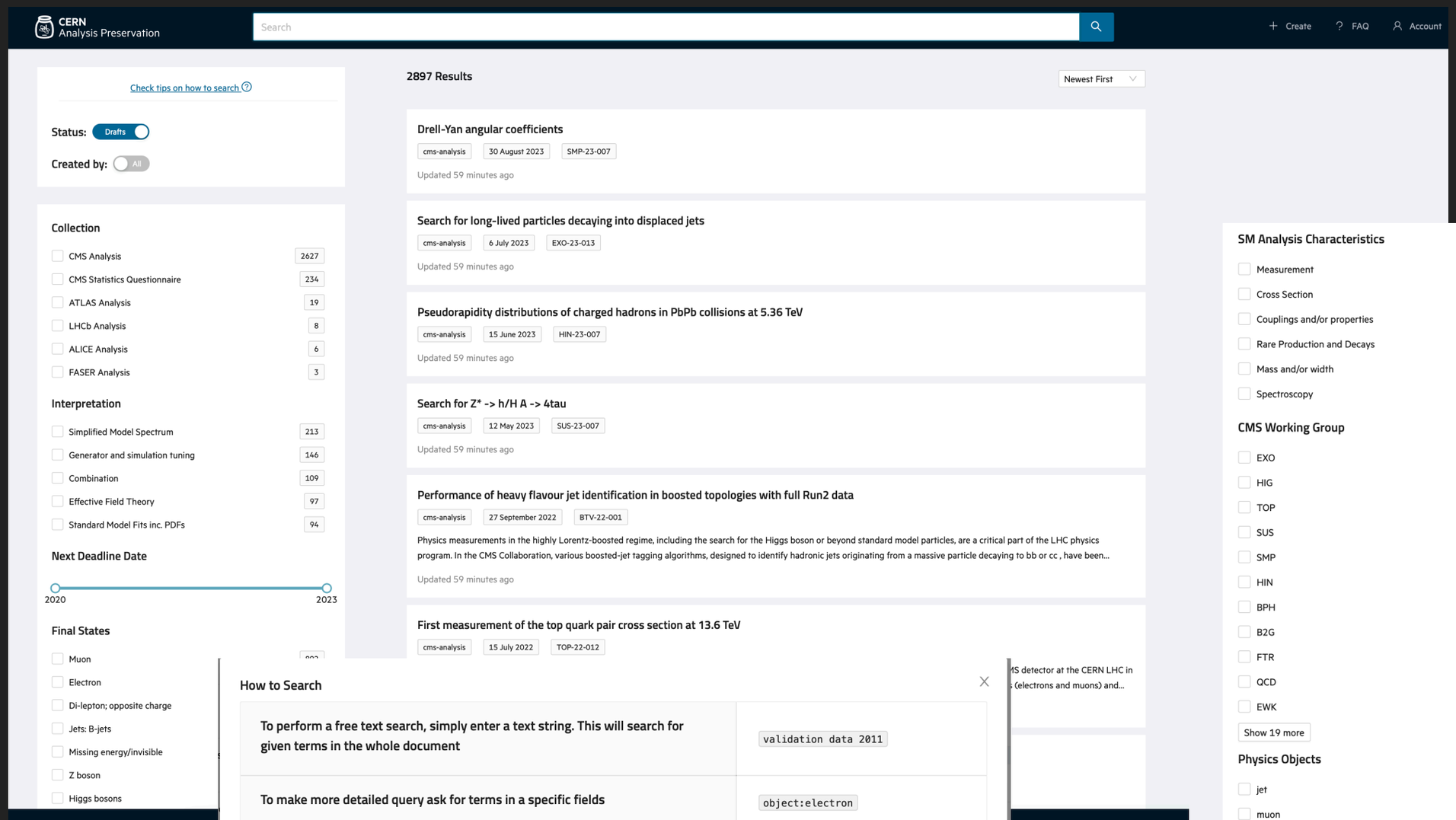
Intro page: overview of work done and shared with the user

Login through CERN

Anyone with a CERN account can login. Depending on the e-groups in his account - content and experience can vary



ADVANCED SEARCH

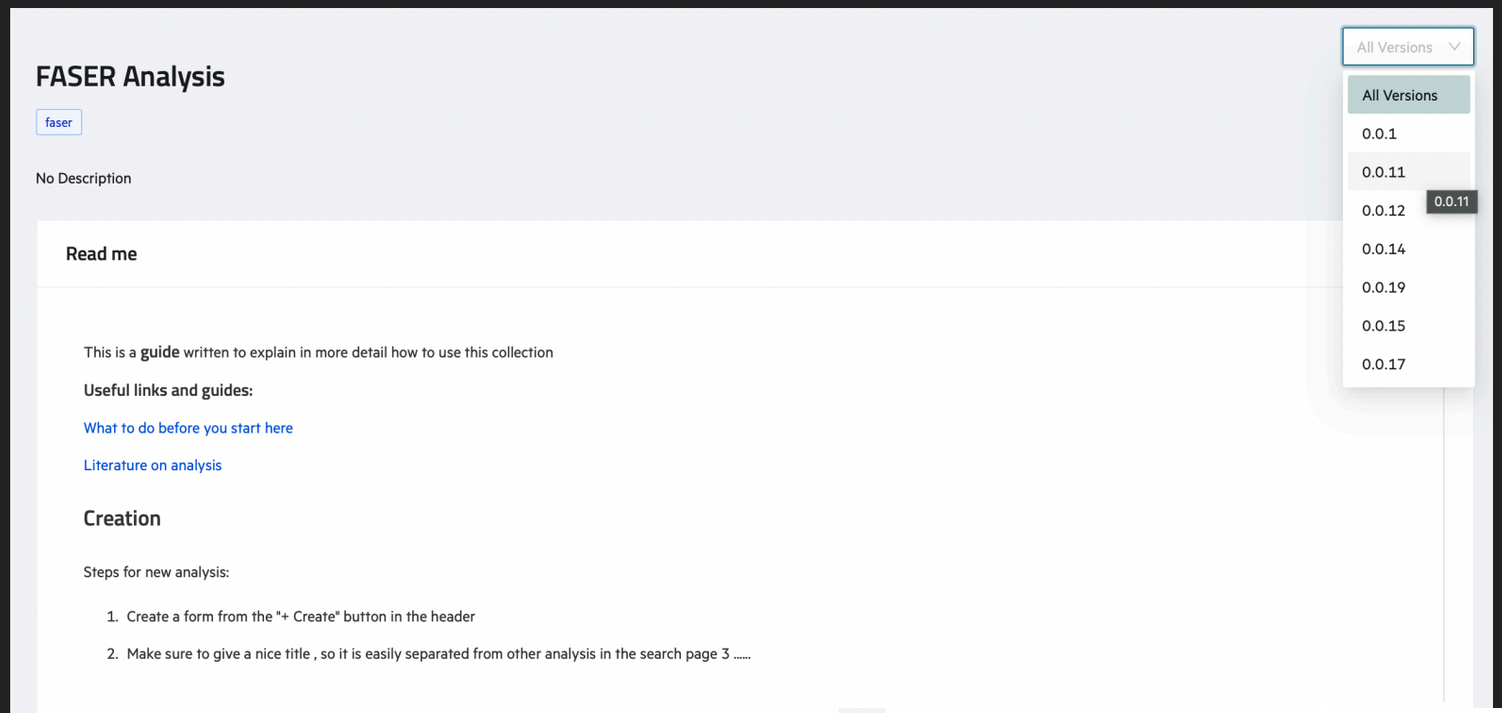


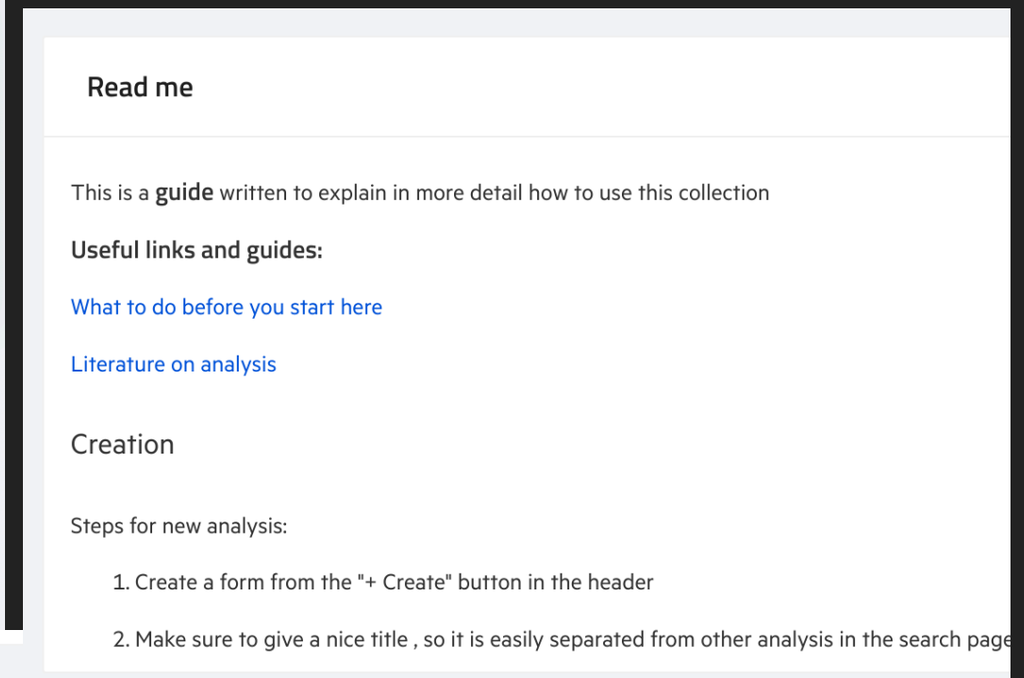
How to Search

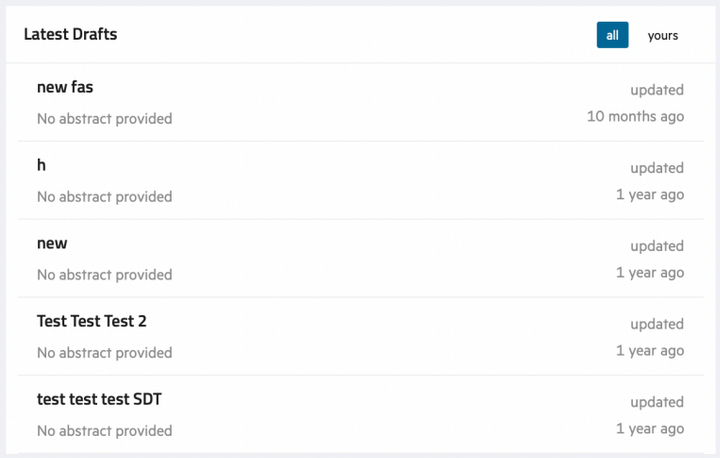
To perform a free text search, simply enter a text string. This will search for given terms in the whole document	`validation data 2011`
To make more detailed query ask for terms in a specific fields	`object:electron`
To point to nested fields use `.` operator or one of many available aliases	`researcher reviewer ananote arxiv status keyword dataset trigger object`
To make your query more generic, use wildcards: `?` for a single character `*` for multiple ones	`/DoubleMu*/*/AOD`
To search for terms contatining special characters escape them with `/`	`* ? . : ! ( ) [ ] " ~`
To search for a range of dates, put them in brackets, using the keyword `TO` between them. The dates follow the `YYYY-MM-DD` standard.	`deadline:[2018-01-20 TO 2020-02-01]`

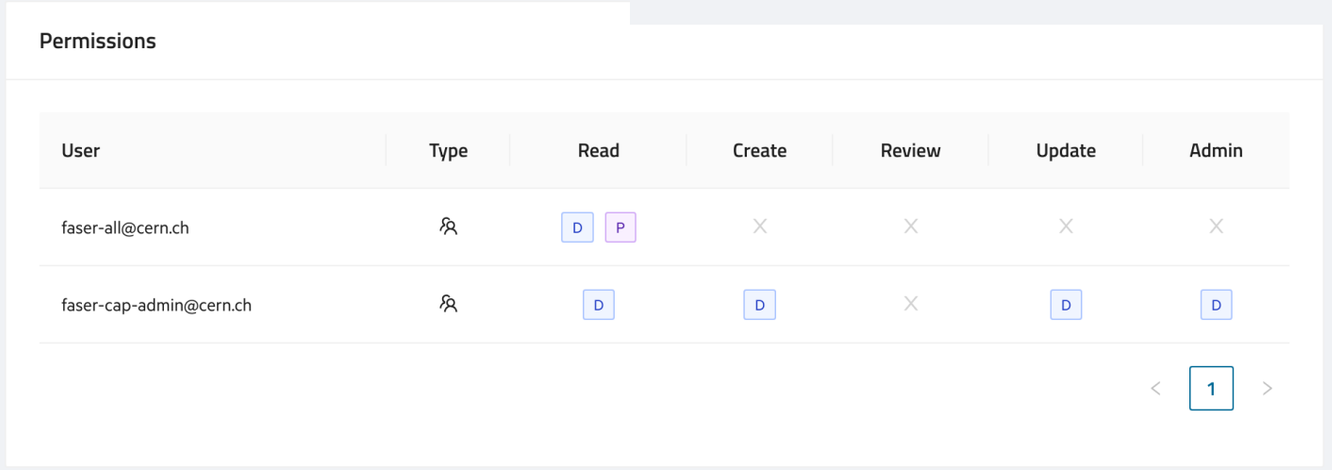
COLLECTION PAGE

- quick link and overview of each collection
- detailed page, with extra guides and information
- collection permissions overview



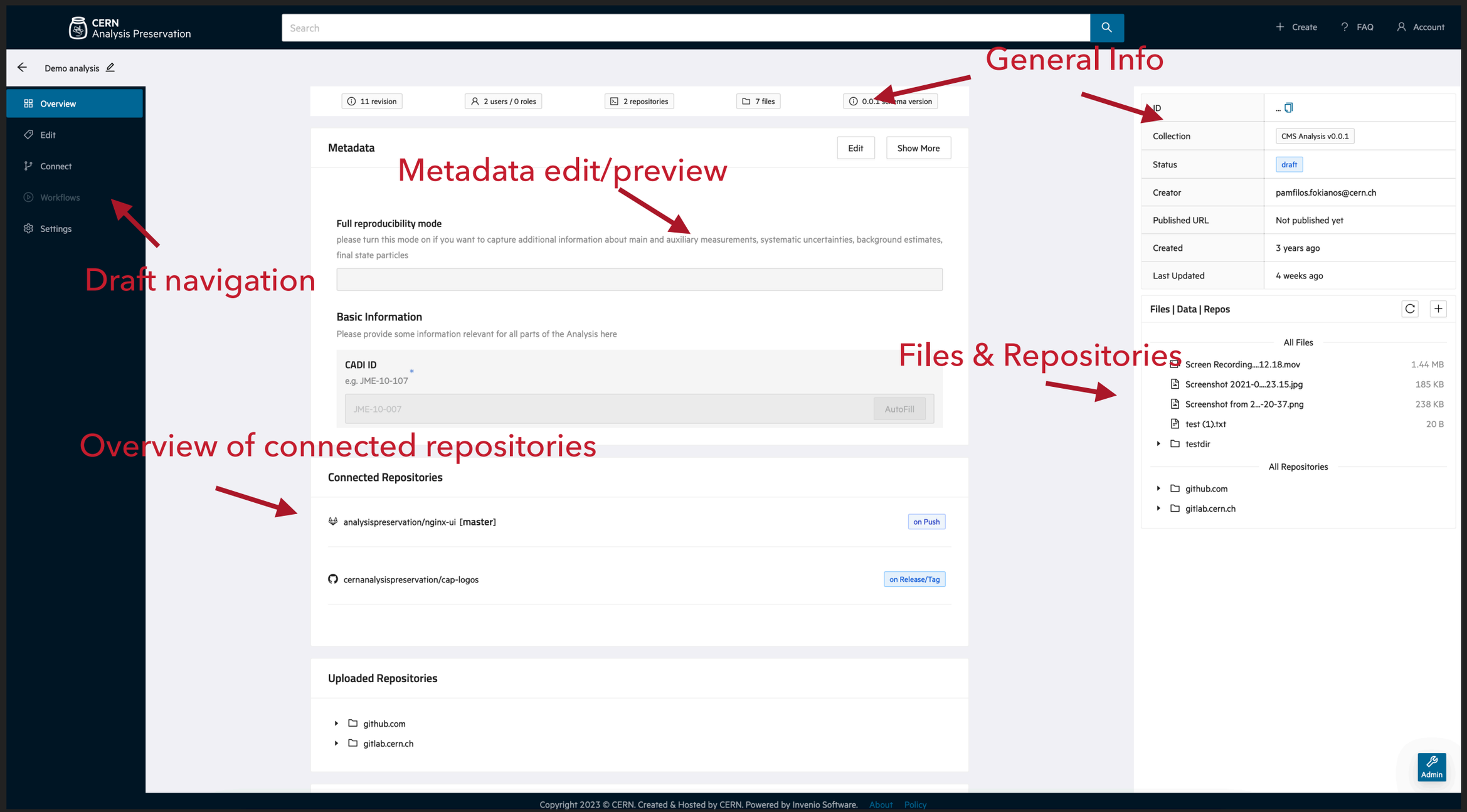






RECORD EDITOR/PREVIEW

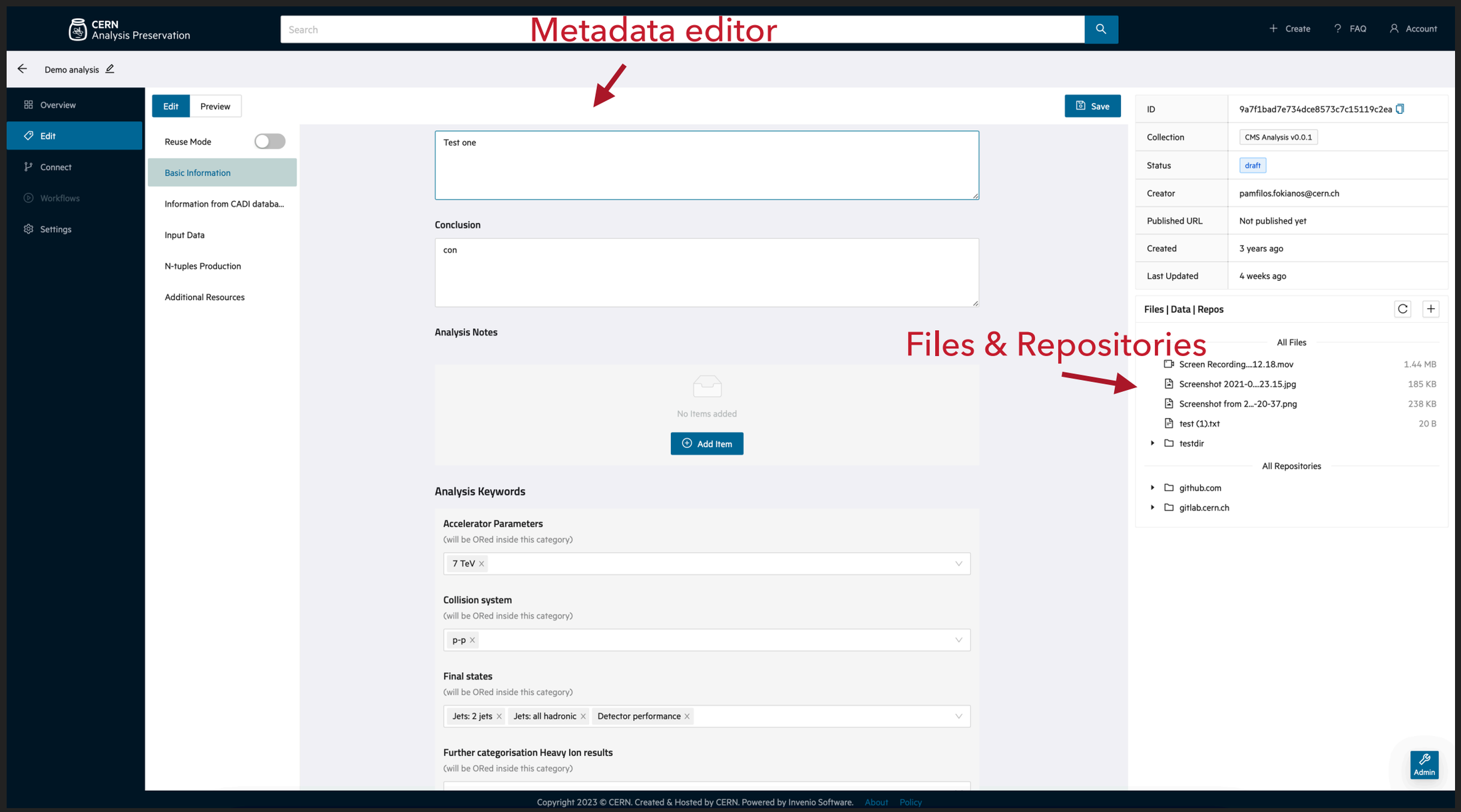
Overview of the draft



RECORD EDITOR/PREVIEW

Metadata Editor

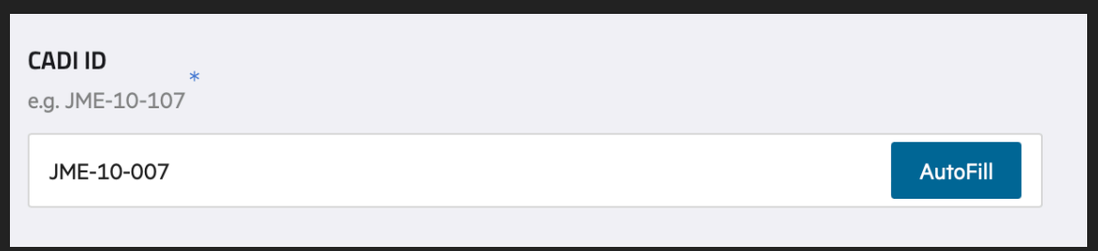
Metadata editor, built by the underlying JSON-schema. It is also responsible for data validation - very customisable

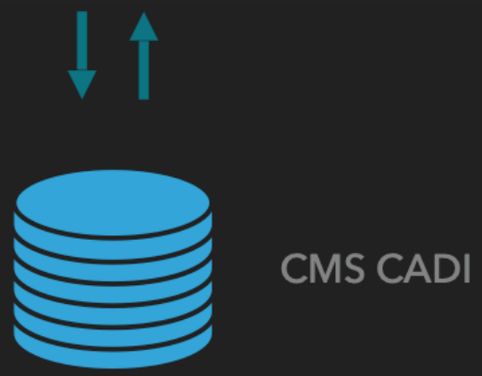


RECORD EDITOR/PREVIEW

Feature: Suggestions & Auto-completes

Automated fetching from experiment DBs/APIs/other sources





Information from CADI database

Automatically taken from CADI, based on CADI ID

Name

HCAL Commissioning in pp Collisions at sqrt{s}=7 TeV

Description

HCAL Commissioning in pp Collisions at sqrt{s}=7 TeV

Contact Person

chlebana@fnal.gov

Twiki

NO URL

Created

2010-04-21

Paper

PAS

http://cms.cern.ch:80/iCMS/analysisadmin/get?analysis=JME-10-007-pas-v2.pdf

Publication Status

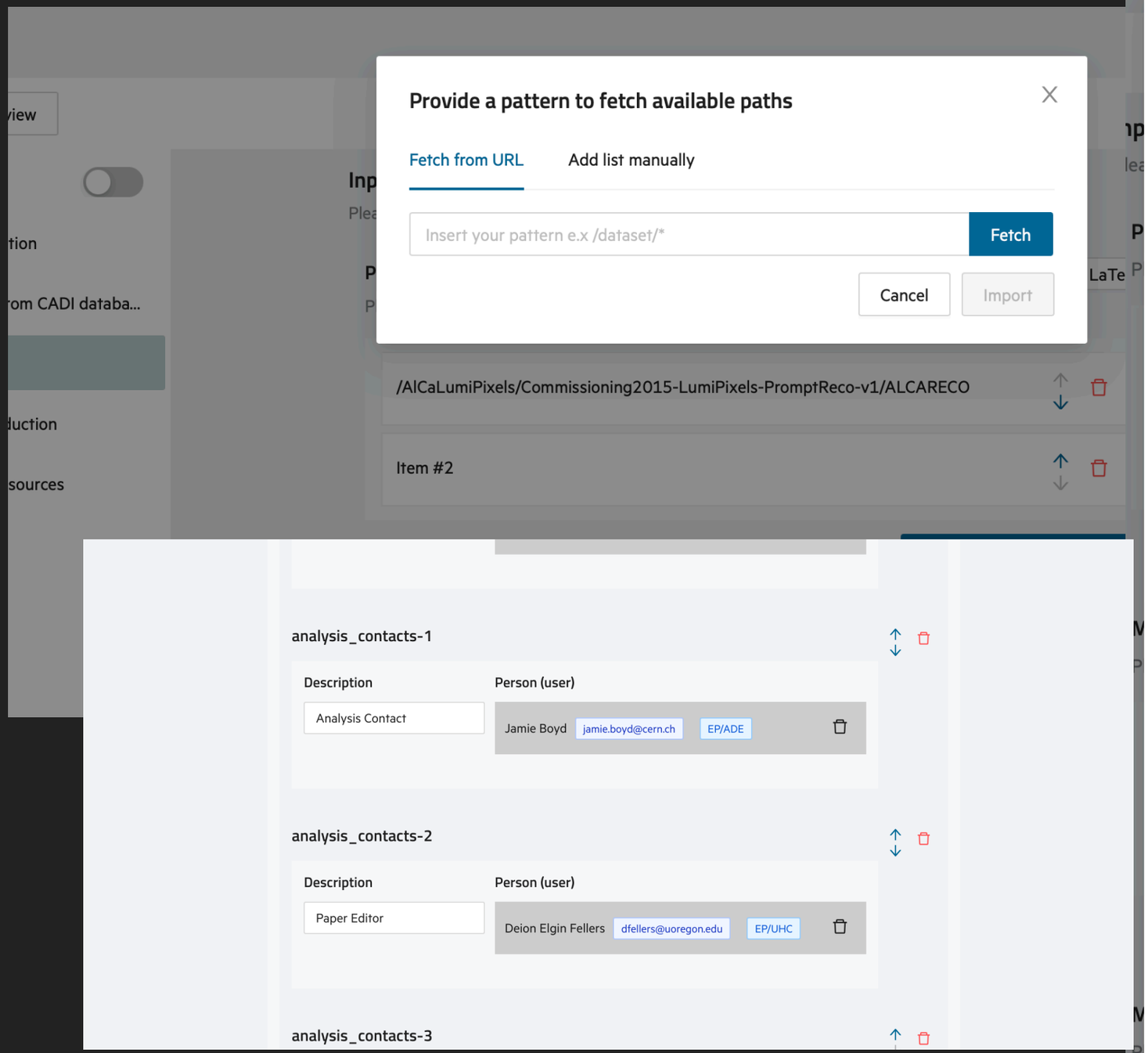
Status

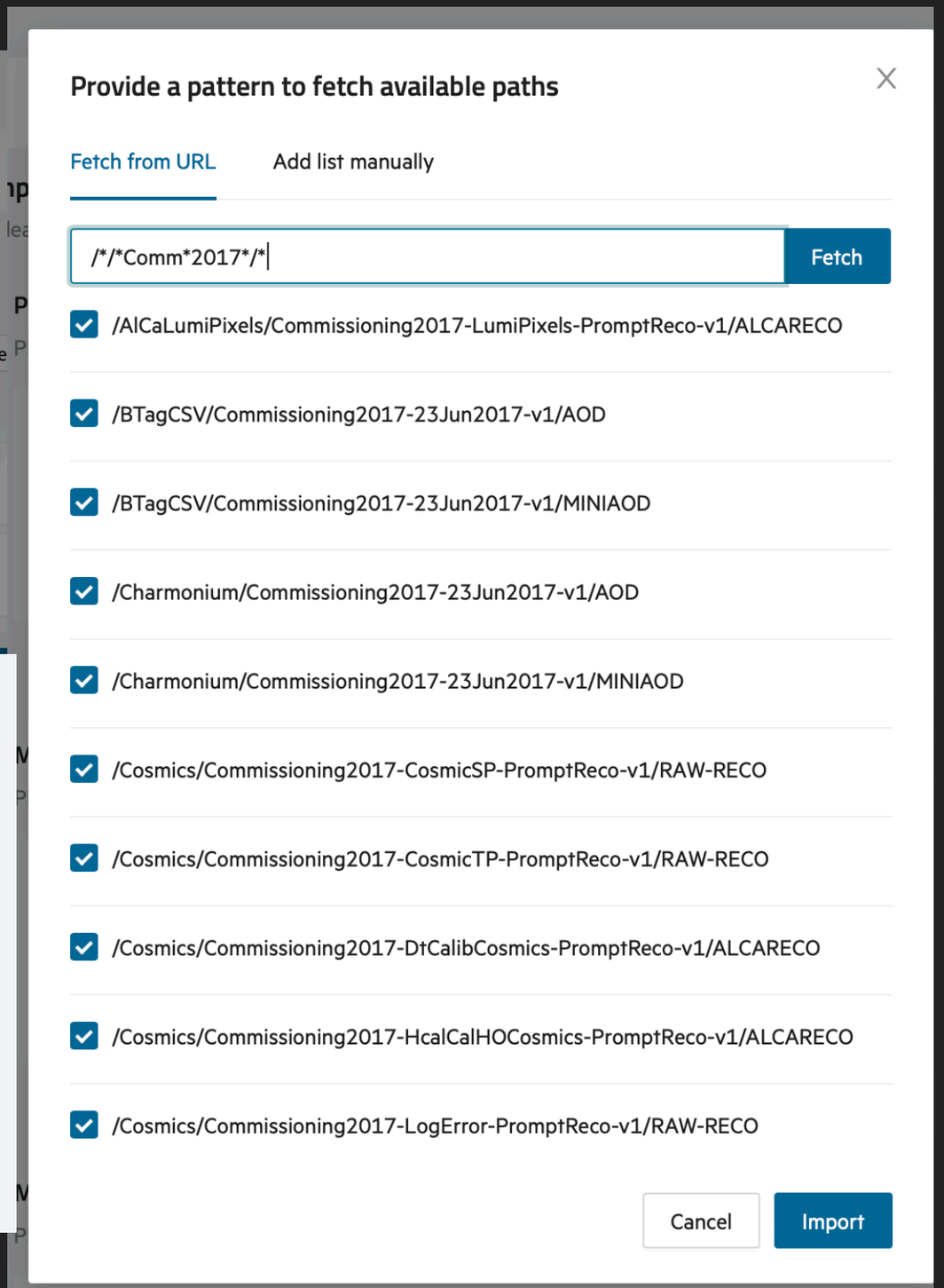
SUPERSEDED

RECORD EDITOR/PREVIEW

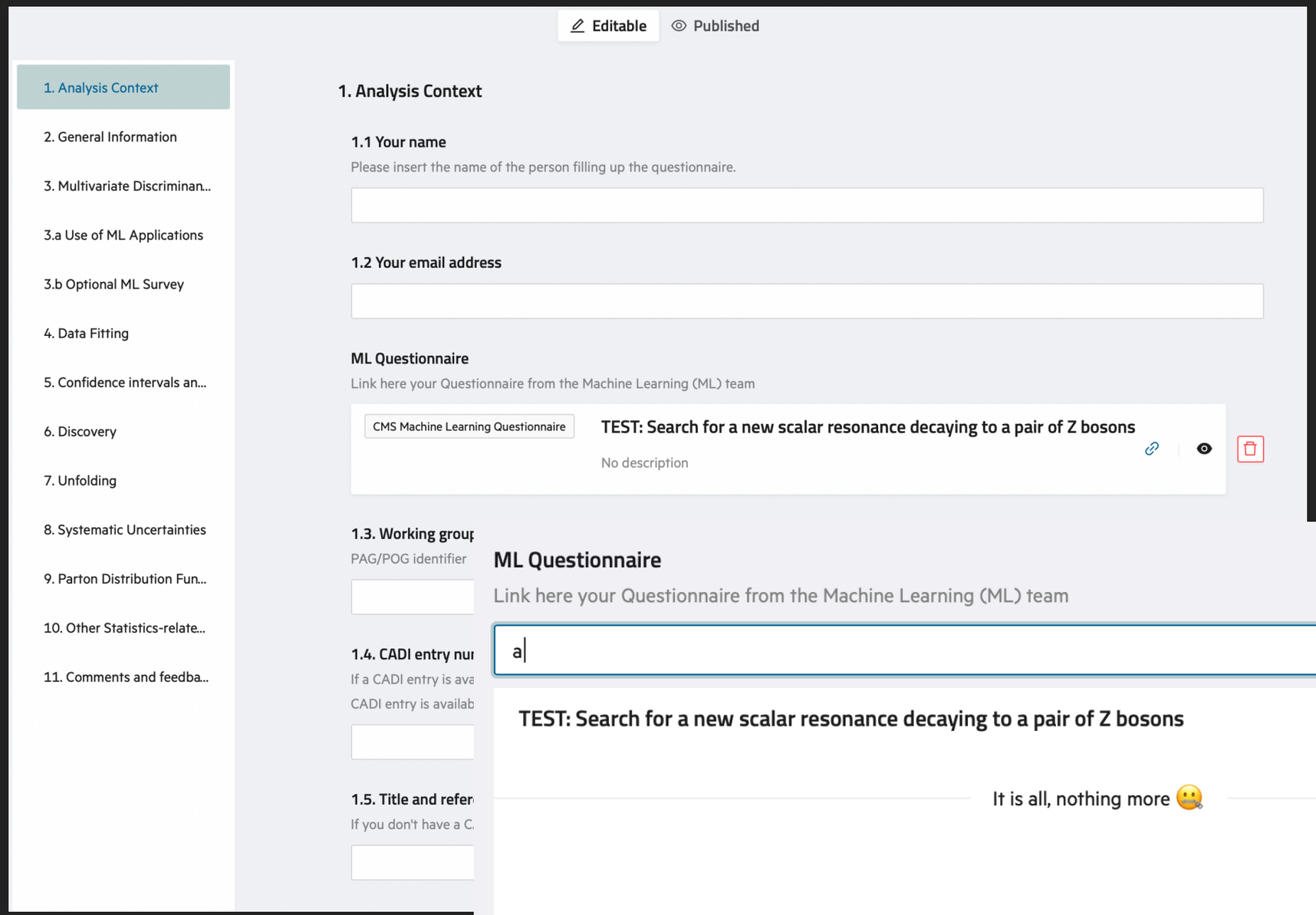
Feature: Suggestions & Auto-completes

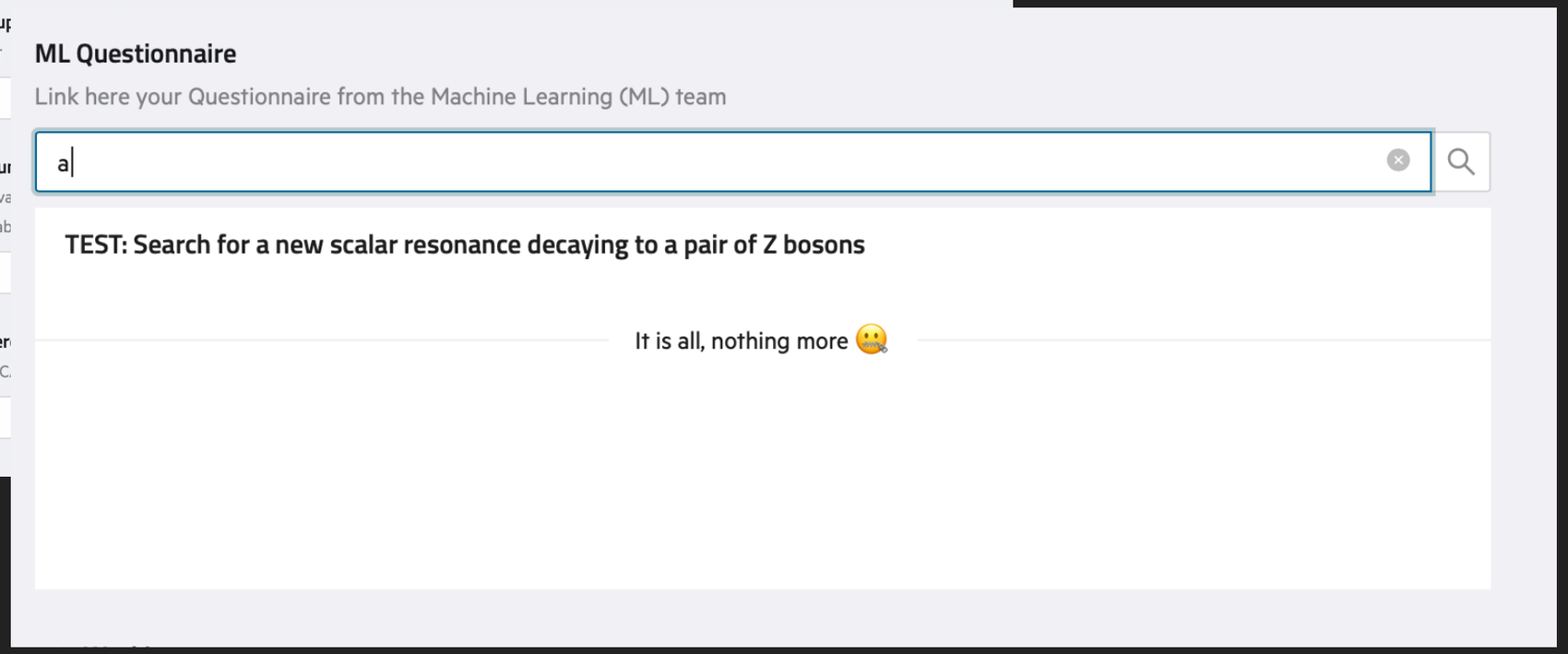
Batch imports





RECORD EDITOR/PREVIEW





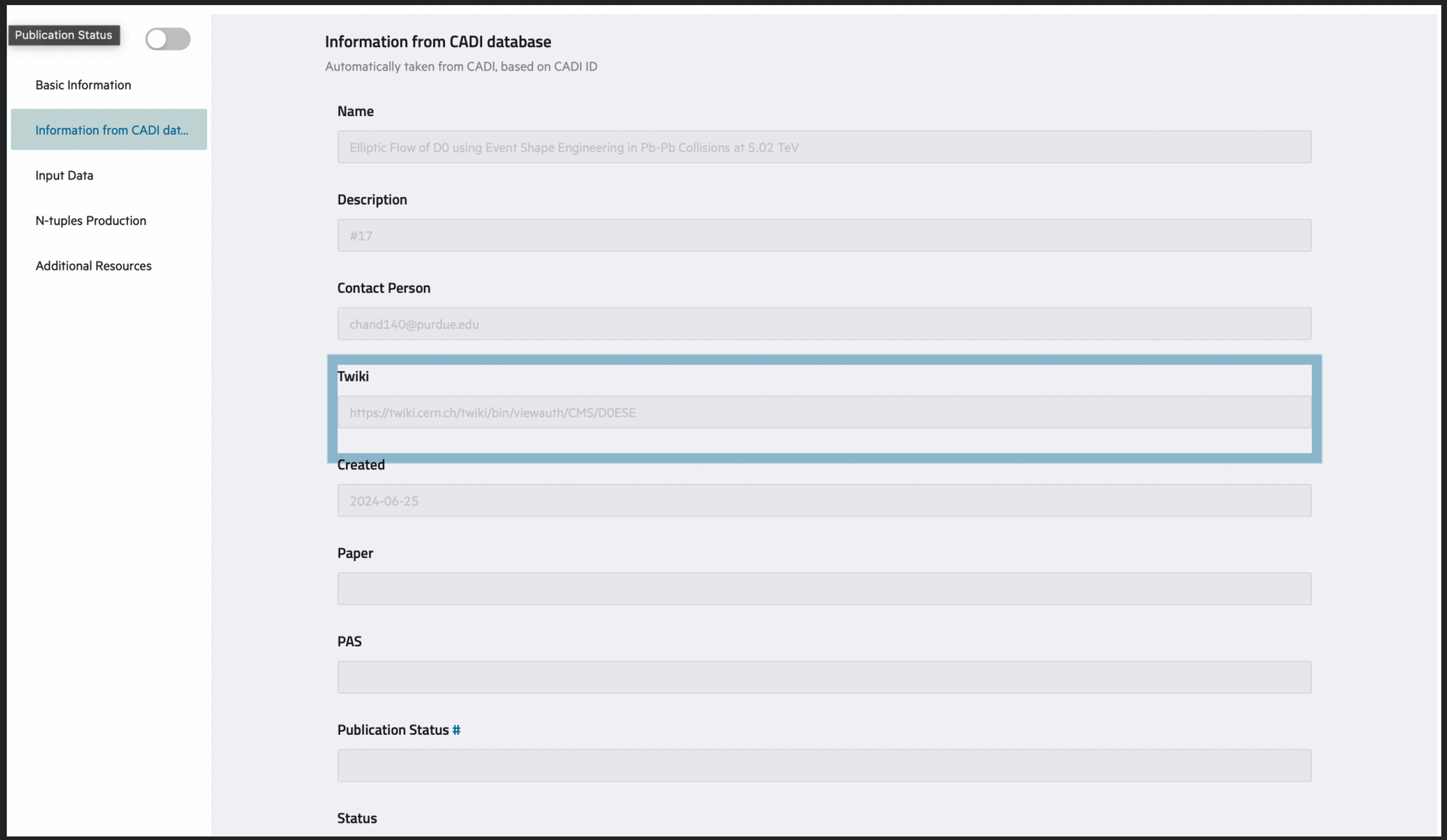
Feature: Link with other documents

- make queries with nested data
- customisable
- same/different collection documents
- goal: split/compact forms, easy submission

RECORD EDITOR/PREVIEW

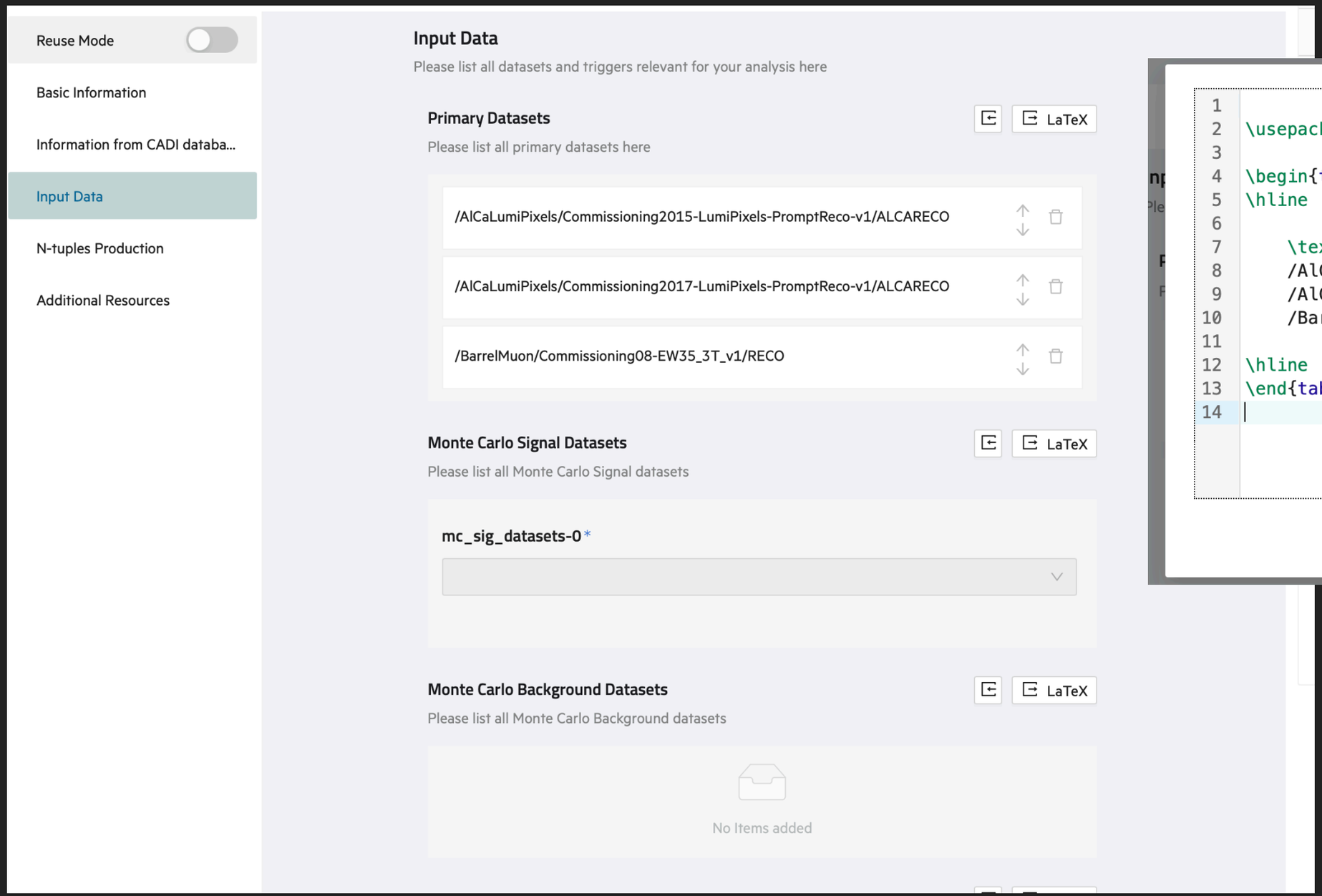
Feature: Sharable links for fields

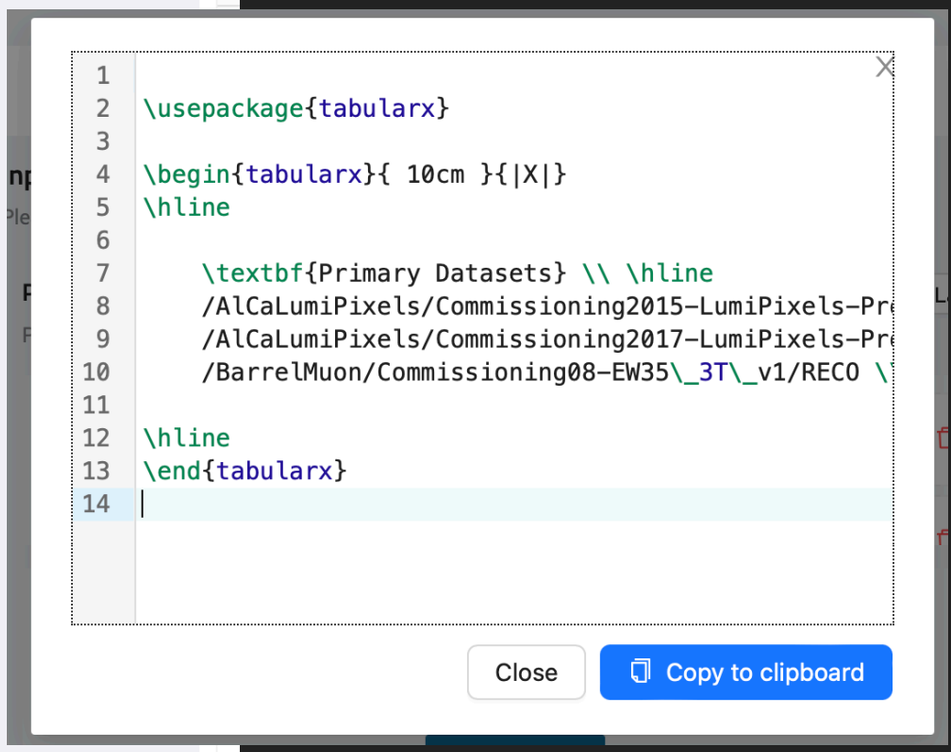
https://cap-test.cern.ch/drafts/302cc66fdd8f4d8baad79f587d99a718/edit#root::cadi_info::twiki



RECORD EDITOR/PREVIEW

Features to help in usability - e.g LaTeX exporter



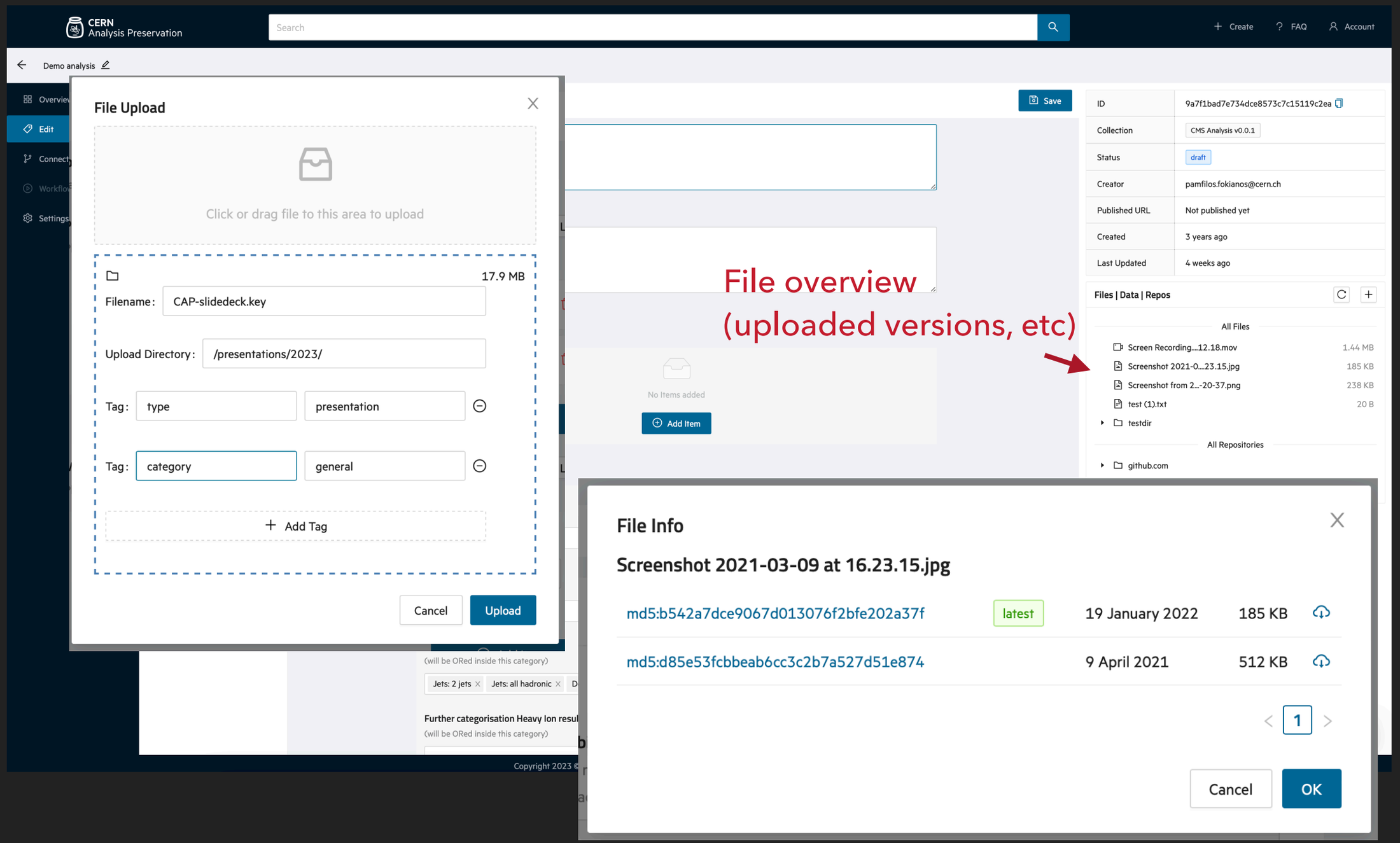


```
\usepackage{tabularx}

\begin{tabularx}{ 10cm }{|X|}
\hline

    \textbf{Primary Datasets} \\ \hline
    /AlCaLumiPixels/Commissioning2015-LumiPixels-Pr
    /AlCaLumiPixels/Commissioning2017-LumiPixels-Pr
    /BarrelMuon/Commissioning08-EW35\_3T\_v1/RECO \

\hline
\end{tabularx}
```

Close | Copy to clipboard

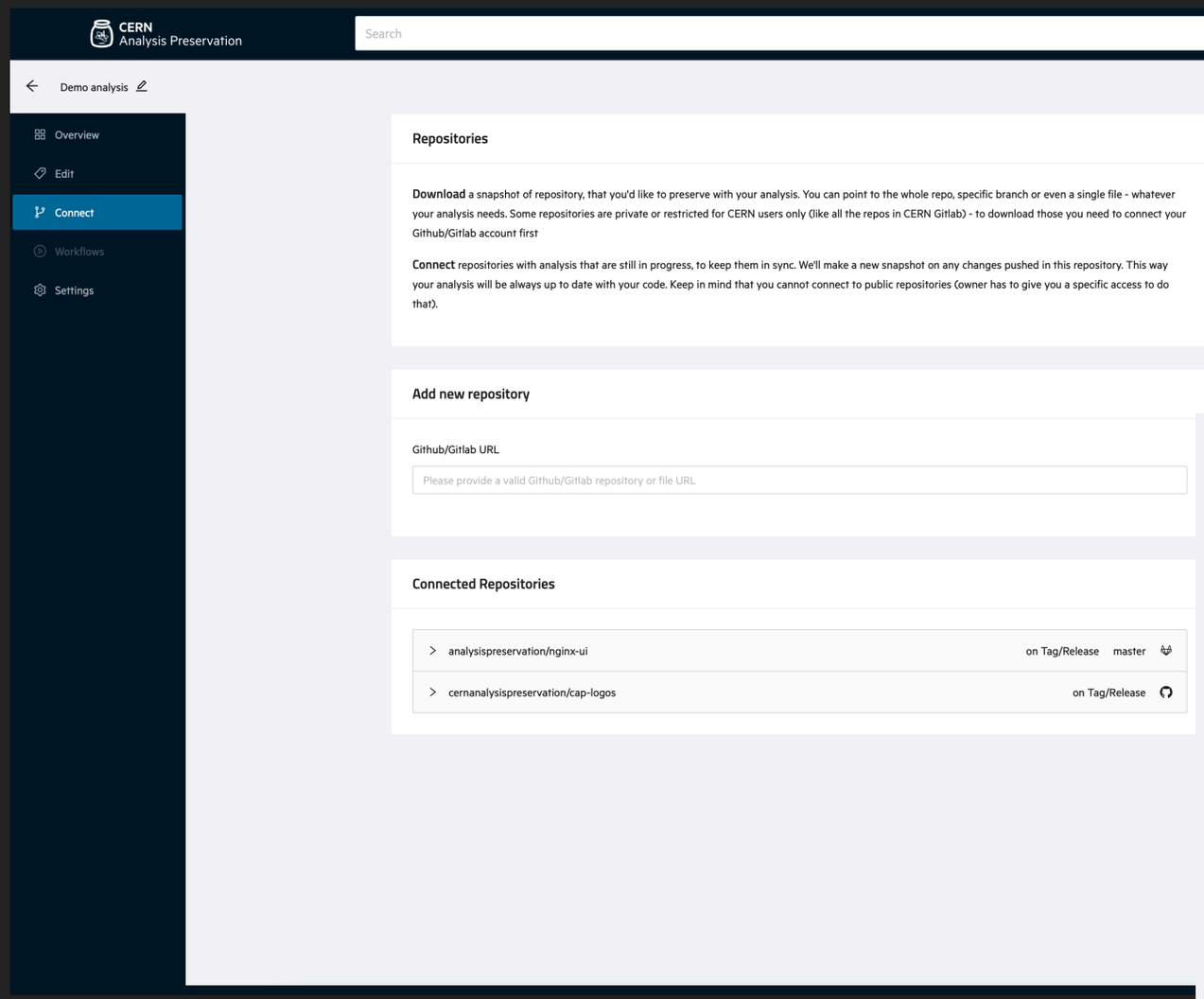
RECORD EDITOR/PREVIEW

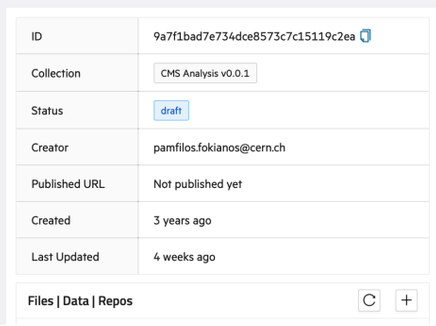
Files/Repositories: File upload - versioning, tagging

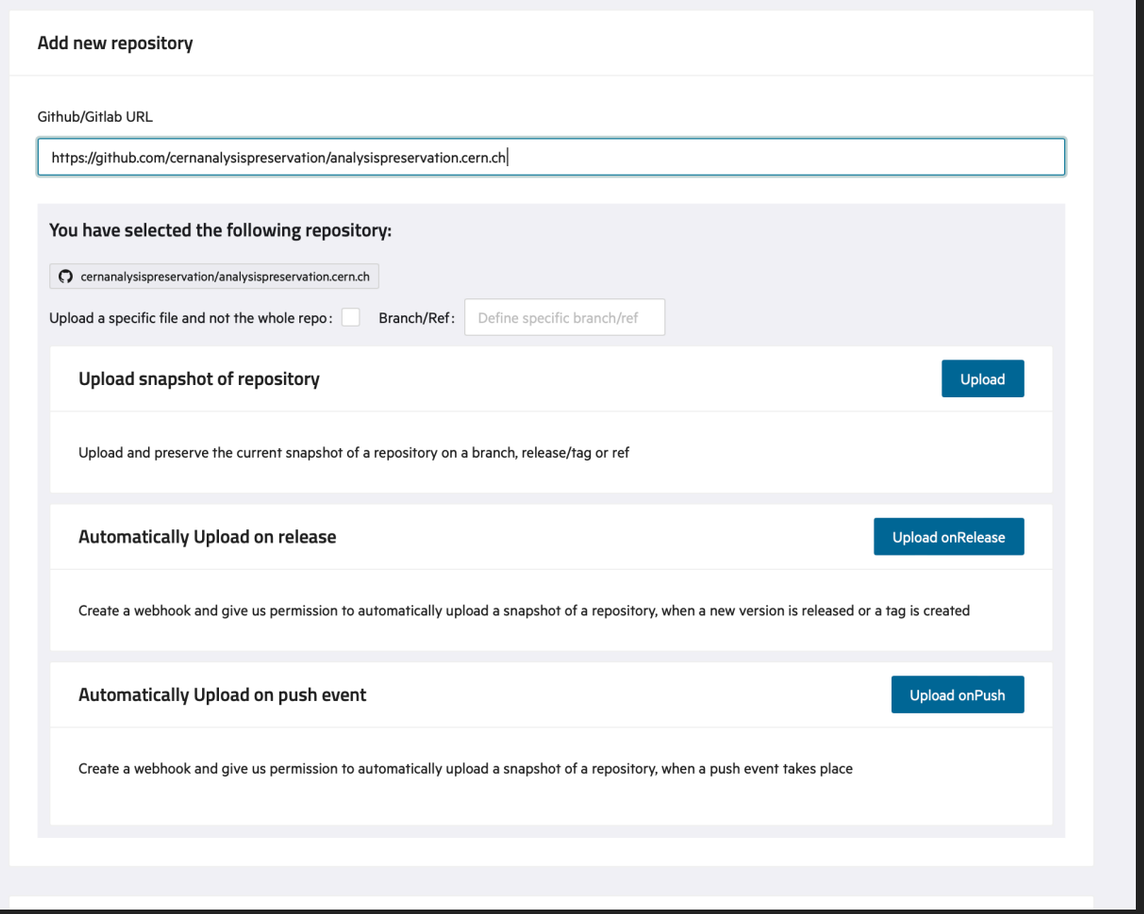


RECORD EDITOR/PREVIEW

Files/Repositories:

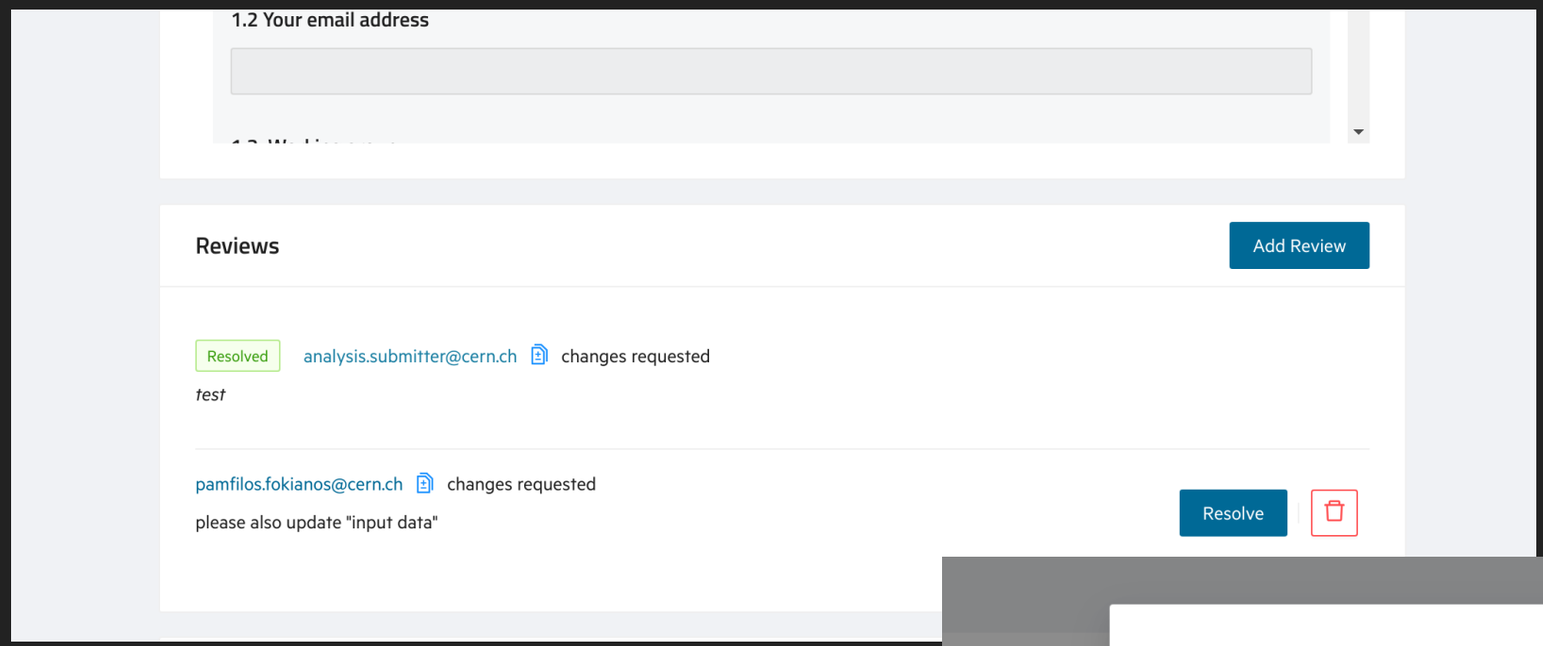


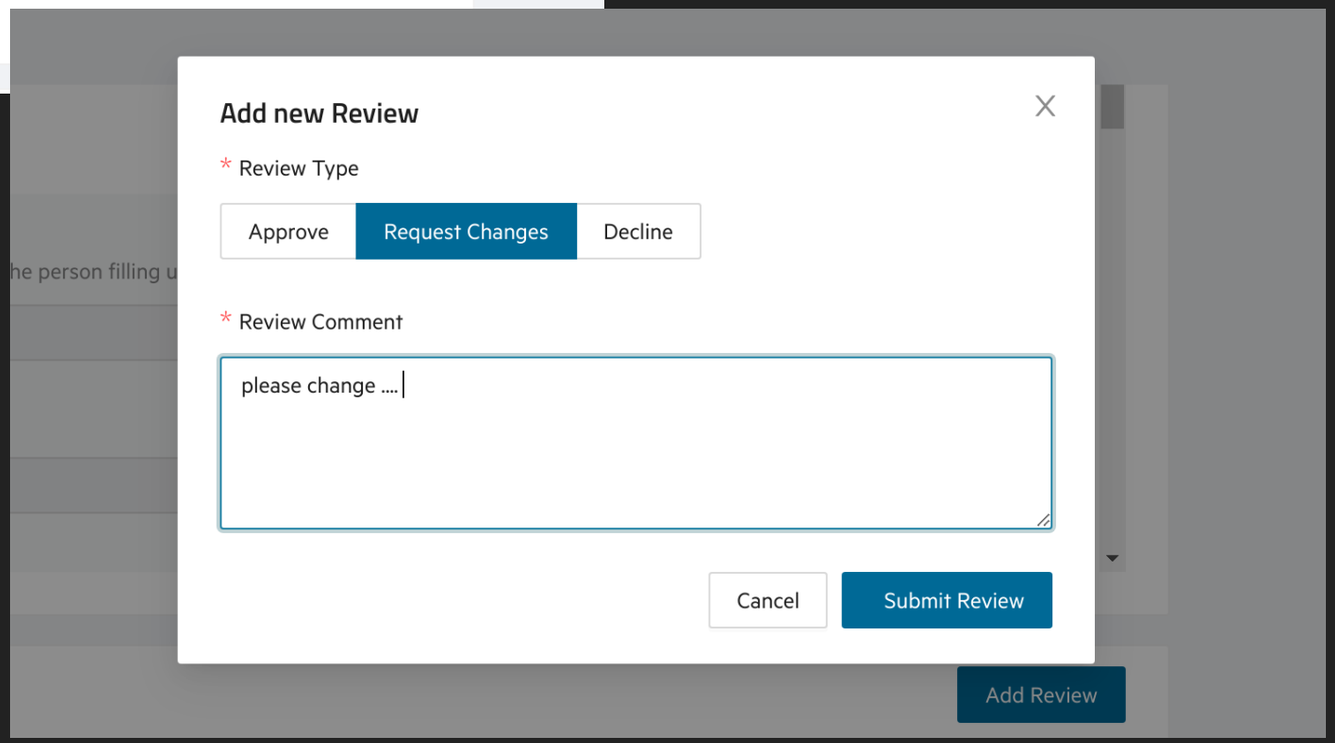




RECORD EDITOR/PREVIEW

Reviews



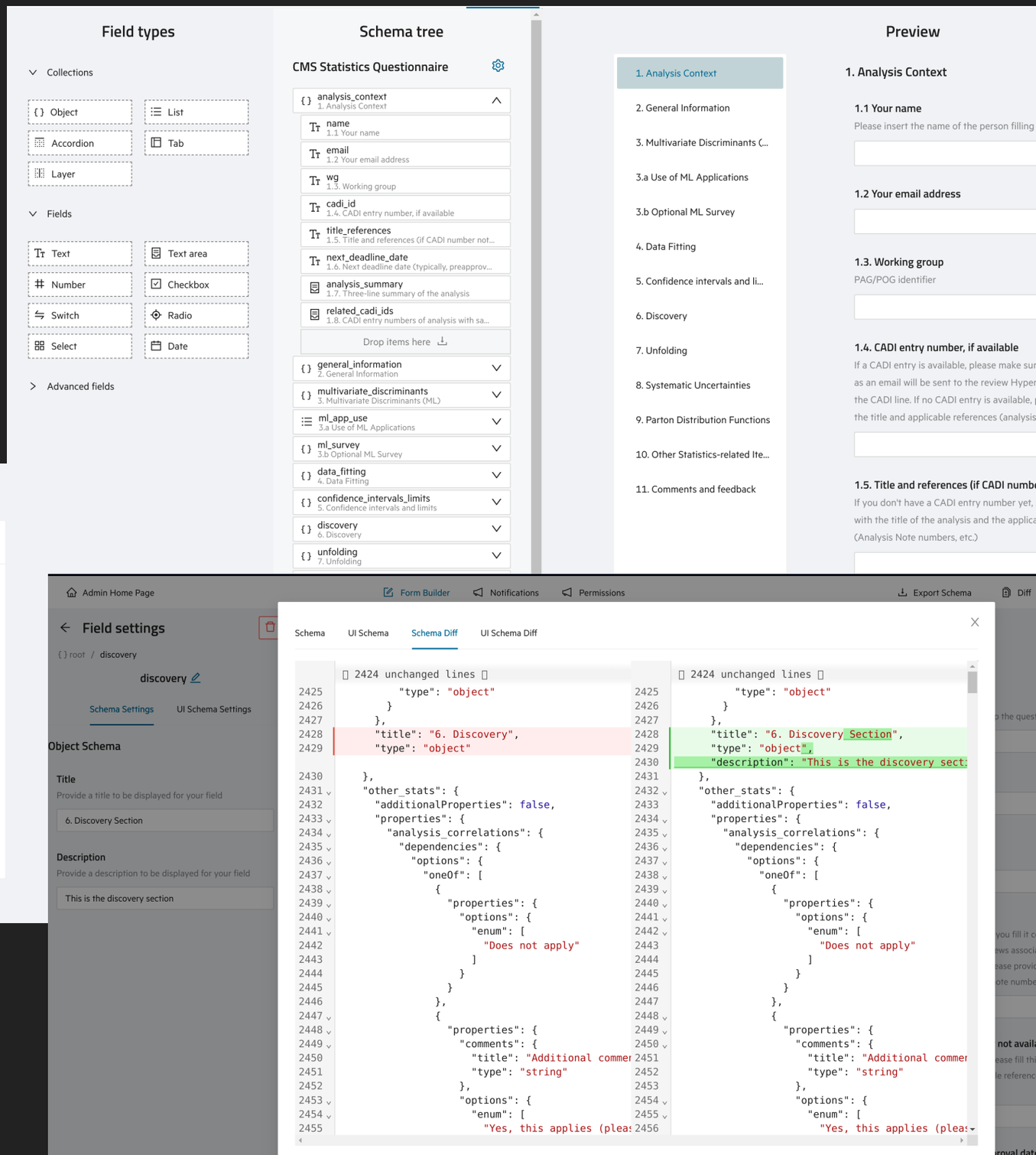


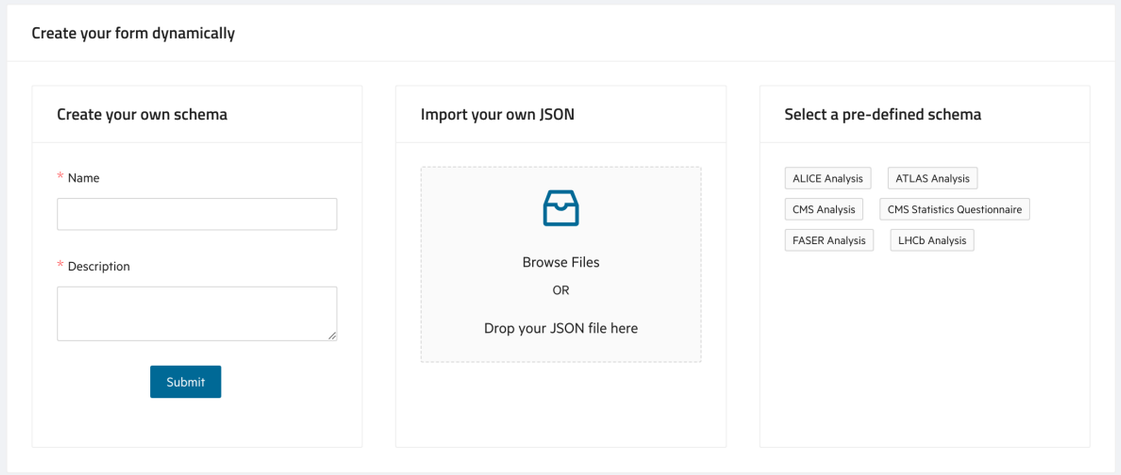
ADMIN AREA

- area for collection administrators (eg experiment coordinators)
- manage forms - content/metadata collected by the users
- manage notifications config - send customised emails when a specific action is taking place (eg when a entry is published, when a review happens, on file upload, etc)
- control collection permissions - admins can give read, write, review, publish rights to users and egroups, without having to go through the CAP team

ADMIN AREA – METADATA/FORMS

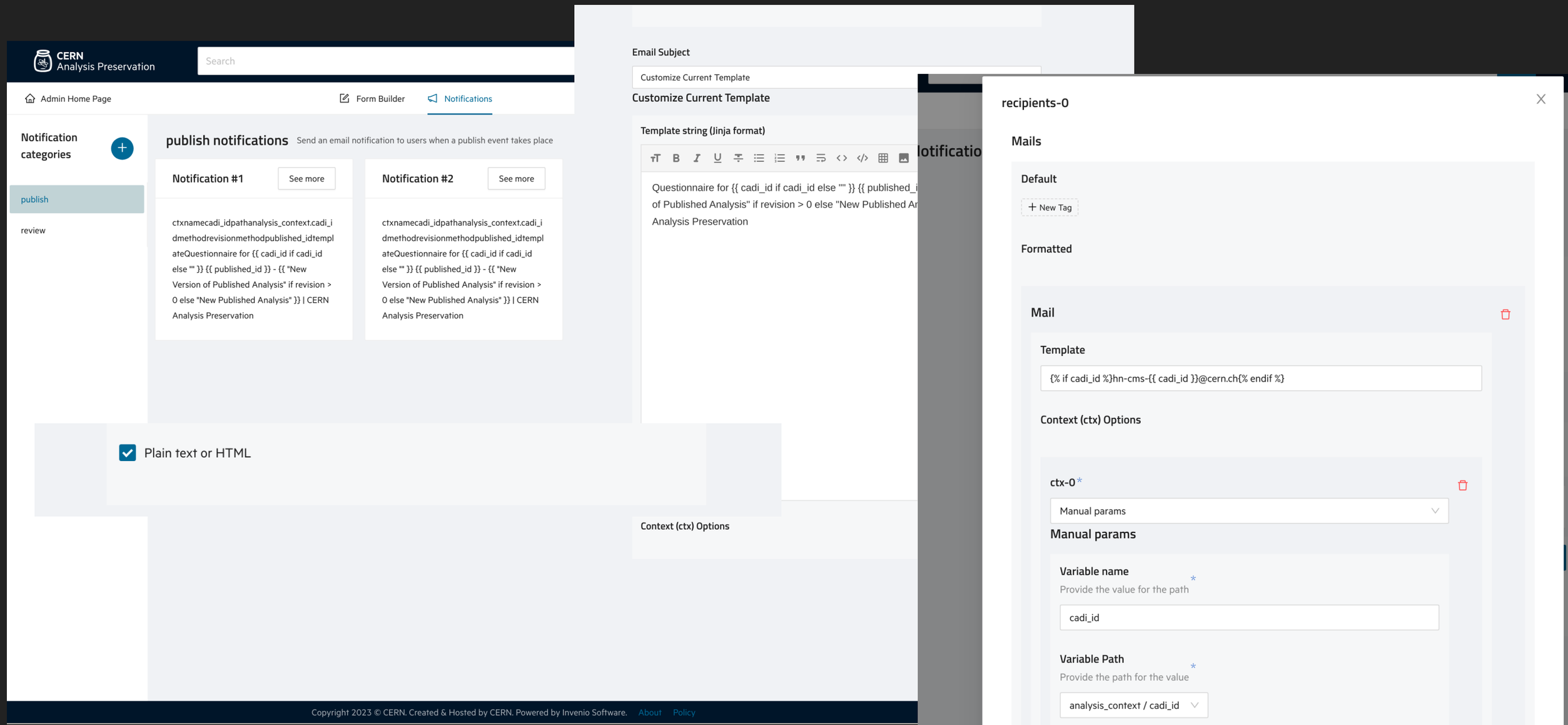
- area for collection administrators (eg experiment coordinators) to
- create/update their forms - content/metadata collected by the users





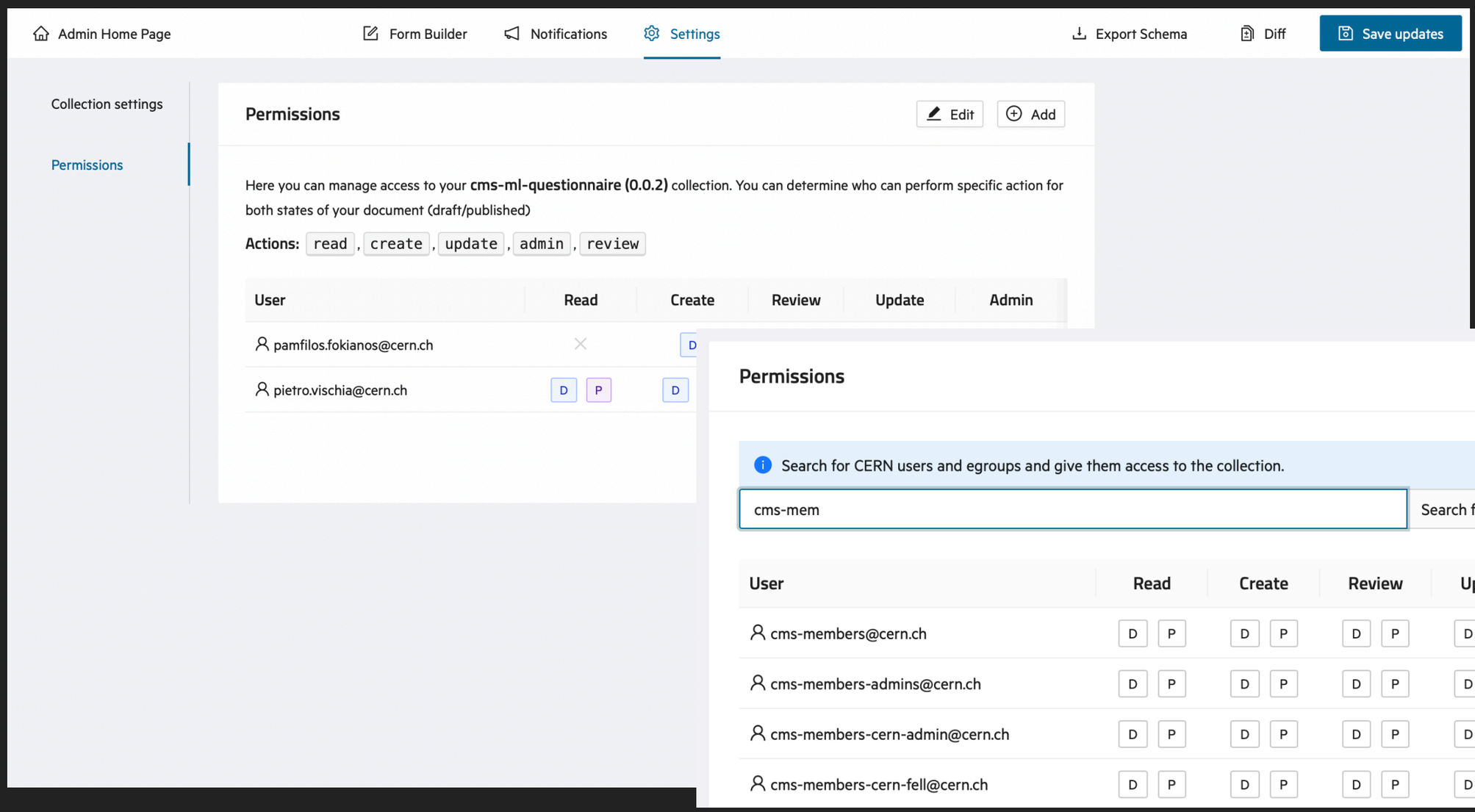
ADMIN AREA – NOTIFICATION

- Create/update notifications - send customised emails when a specific action is taking place (eg when a entry is published, when a review happens, on file upload, etc)



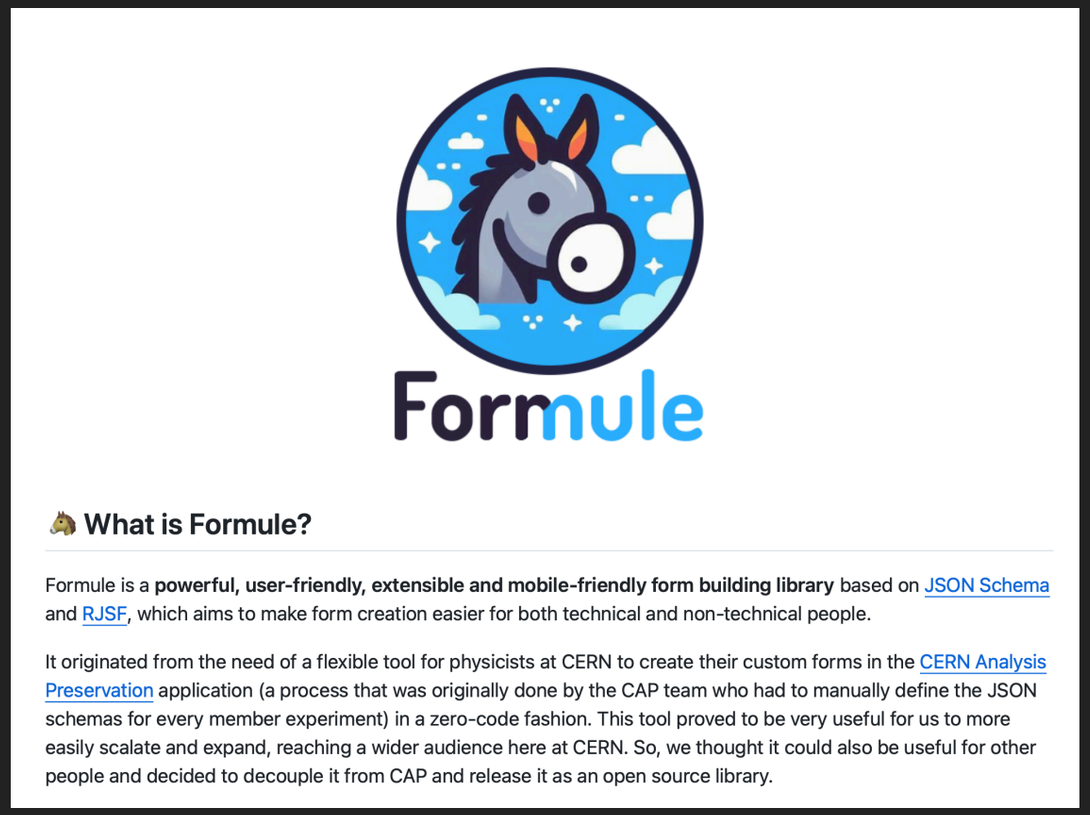
ADMIN AREA – PERMISSIONS

- Control collection permissions - admins can give read, write, review, publish rights to users and e-groups, without having to go through the CAP team



REACT-FORMULE

- JSON-schema builder, form builder, form rendered, data validation
- Open source
- Standalone tool
- First npm release happened ~ Q2 ‘24
- To be used as replacement for our main code-base in CAP



https://github.com/cern-sis/react-formule

DEMO: https://cern-sis.github.io/react-formule/

WHAT'S NEXT?

- Better metadata/file retention (back up) - tapes with use of SIP (Submission Information package)
- Promote "admin area" to coordinators, preservation heads, admins etc
- Onboard more experiments and smaller WG
- Document/file previewers - editors
- Add search configurations to admin (aggregations, filters, etc)

LANDSCAPE PROJECT

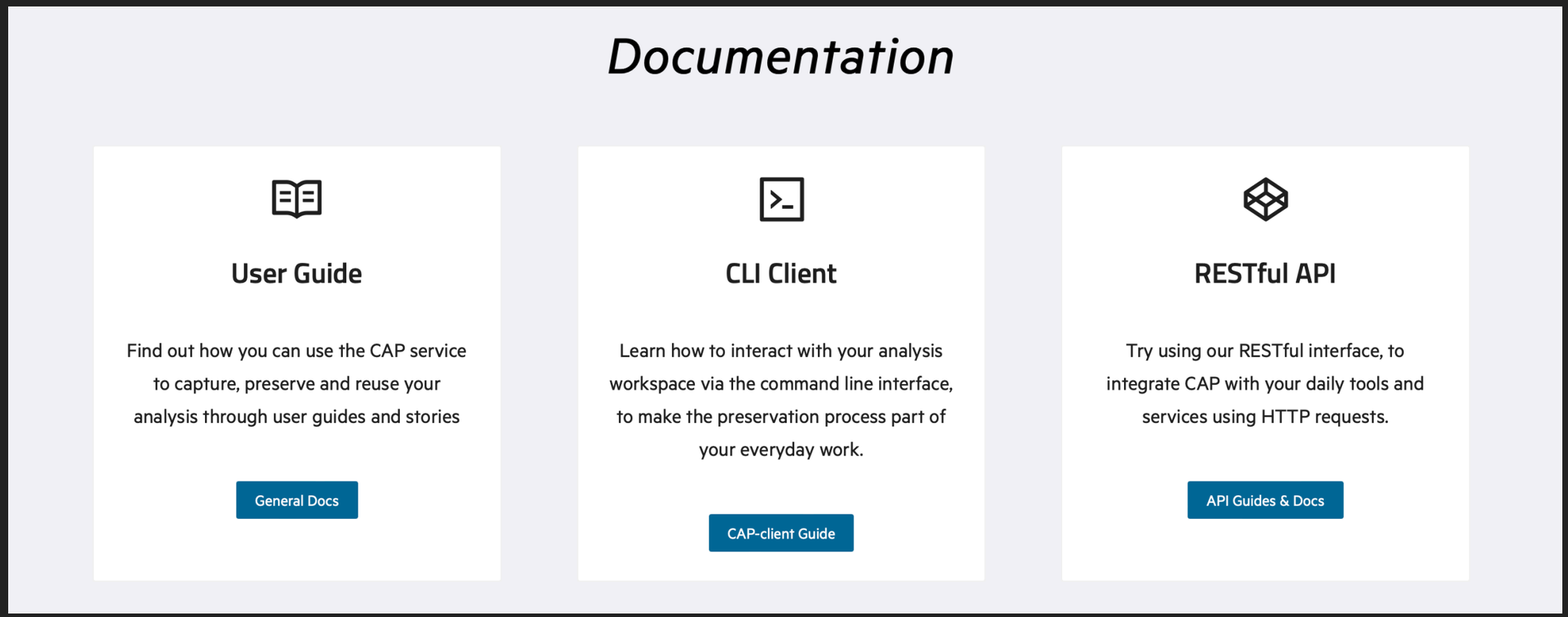
- Aim for CAP to be used by small experiments
- Collect/ landscape project features/ requirements
- Authors collection - example use case
  - example author collections (searching, export to CSV, XMLs, etc)
- Other CAP features to be further developed:
  - Review process

CERN
Analysis Preservation

Web: https://analysispreservation.cern.ch
Github: https://github.com/cernanalysispreservation
Email: analysis-preservation-support@cern.ch

CERN ANALYSIS PRESERVATION

Documentation section



https://analysispreservation.cern.ch/docs/general/
https://analysispreservation.cern.ch/docs/api/
https://analysispreservation.cern.ch/docs/cap-client/

CERN

THANK YOU