CERN Analysis Preservation

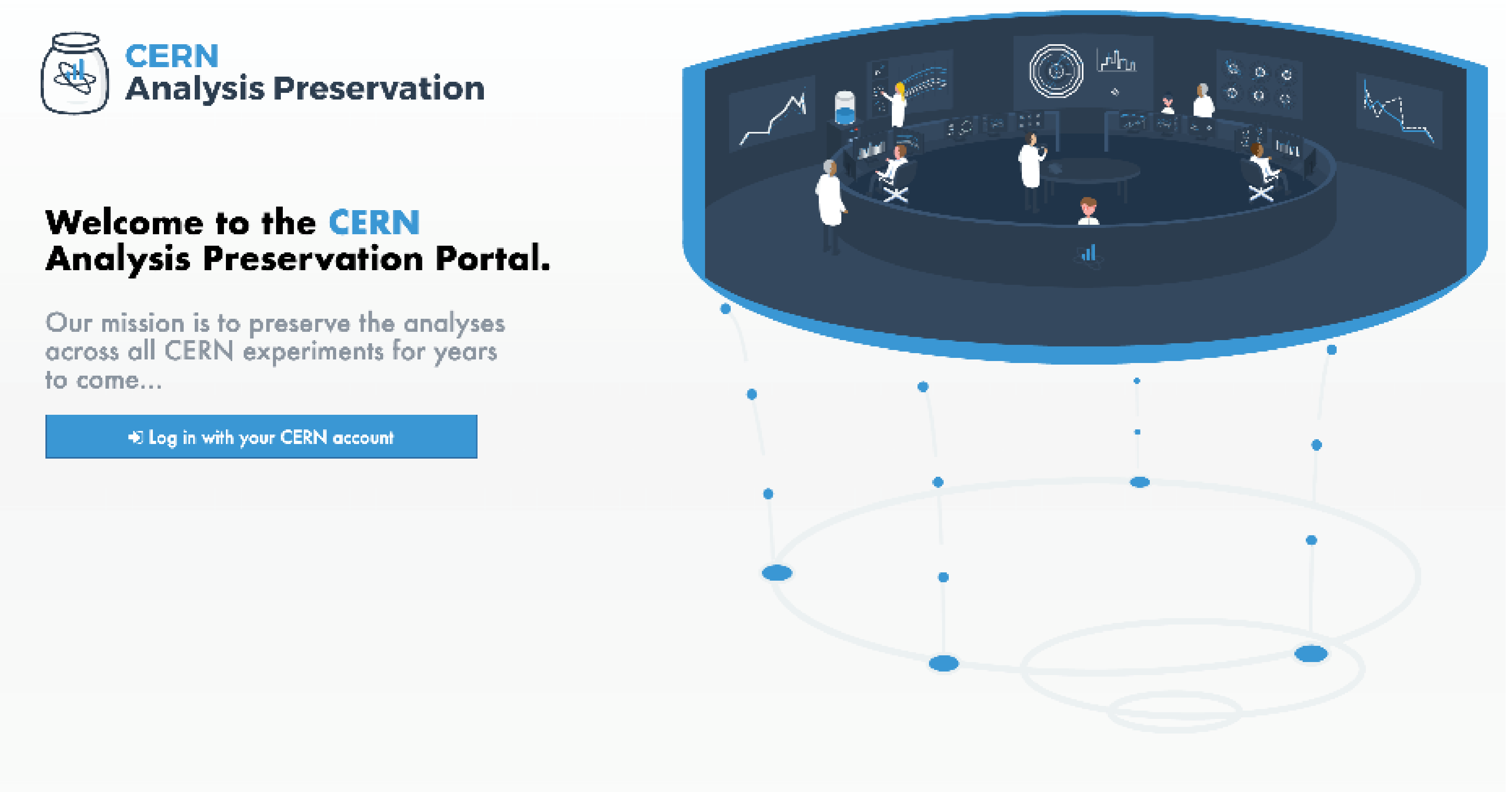
Illuminating the research workflow in High-Energy Physics to enable reproducibility

Sünje Dallmeier-Tiessen, **Artemis Lavasa**, Tibor Šimko, Javier Delgado Fernández, Pamfilos Fokianos, Robin Dasler, Anxhela Dani, Annemarie Mattmann, Ioannis Tsanaktsidis, Anna Trzcinska, Diego Rodriguez Rodriguez

… and many other contributors at CERN and elsewhere

CERN

CERN Analysis Preservation



CERN
Analysis Preservation

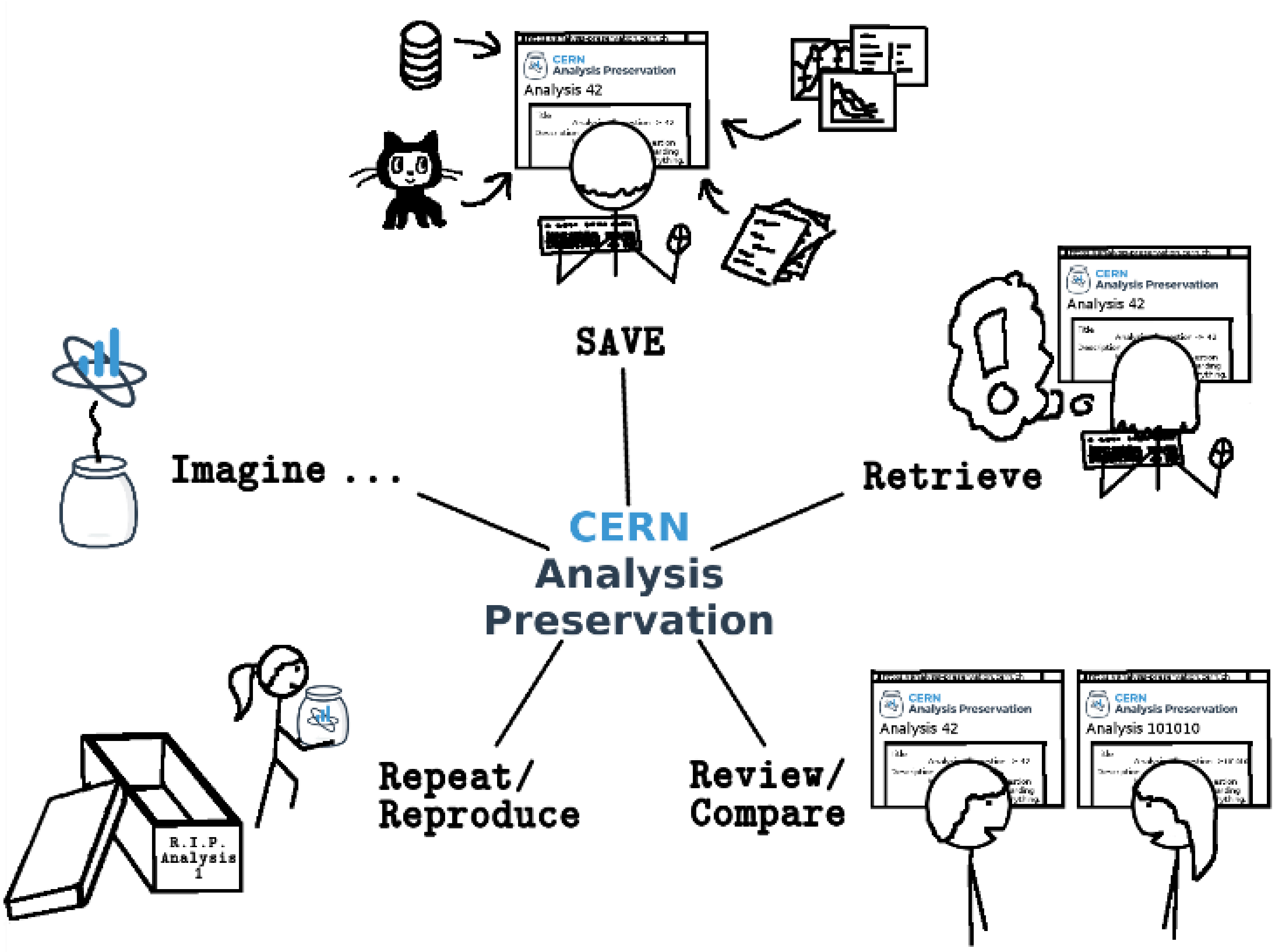
Welcome to the CERN Analysis Preservation Portal.

Our mission is to preserve the analyses across all CERN experiments for years to come...

Aim

- Preserve information about and tools for analyses created by the four LHC collaborations
  - Information about primary and reduced datasets
  - Information about underlying OS platforms and user software used to study it
  - Configuration parameters
  - High-level physics information (e.g. physics object selection, cuts and vetos)
  - Any necessary documentation and discussions recorded alongside the process
- Reproduce an analysis even many years after its initial publication → extend impact of preserved analyses through validation and recasting services

User stories



Three pillars

1. Describe

Knowledge modelling
JSON Schema

2. Capture

Push: deposit API
Pull: background ingestion

3. Reuse

Runnable instructions
Instantiate on OpenStack

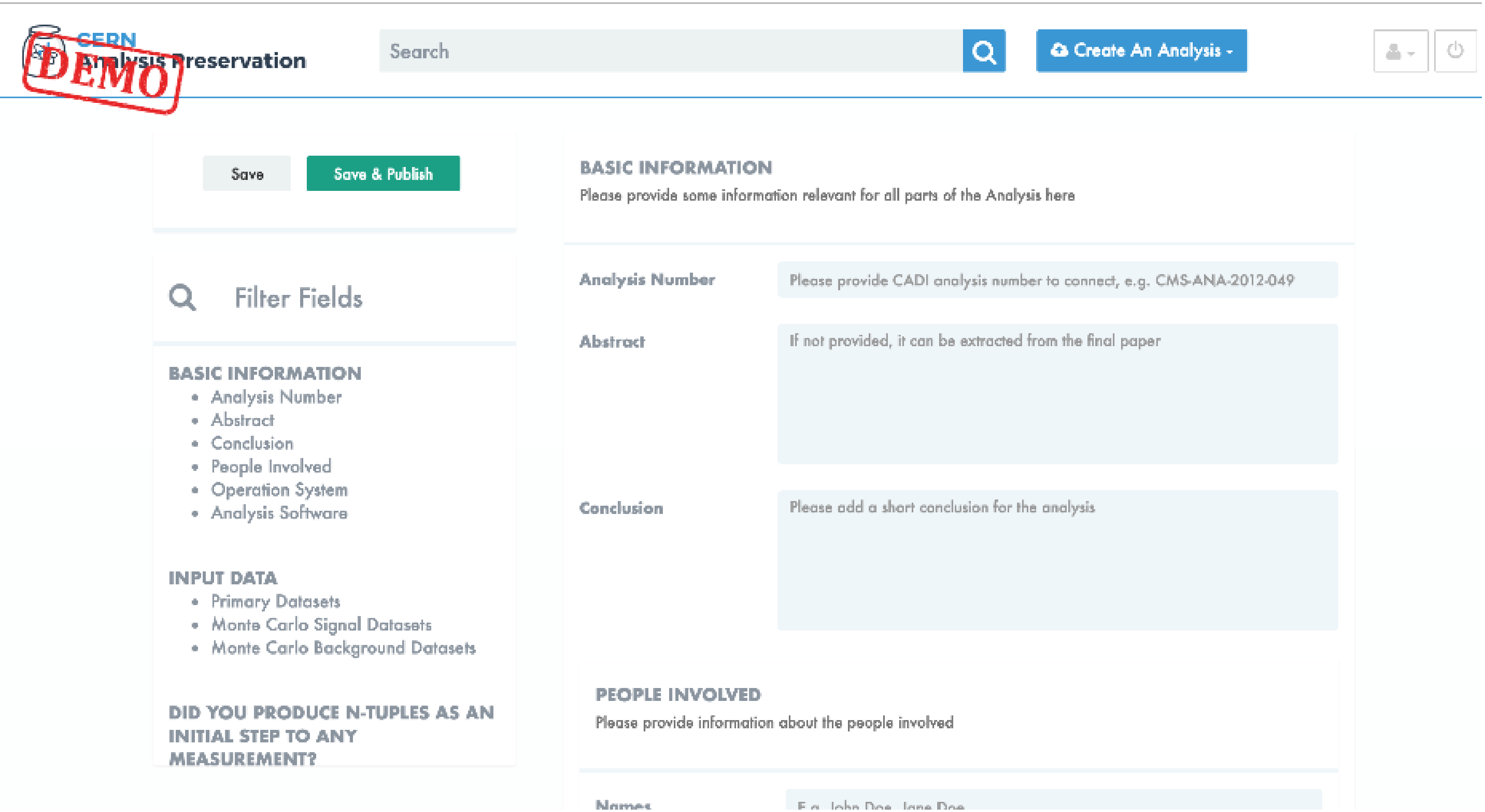
Three pillars

1. Describe

JSON Schema

- is also commonly used by many of the collaboration databases
- accommodates the complex metadata in the best way possible

Capturing the research outputs



CERN Analysis Preservation

DEMO

Search

Save

Save & Publish

Filter Fields

BASIC INFORMATION

- Analysis Number
- Abstract
- Conclusion
- People Involved
- Operation System
- Analysis Software

INPUT DATA

- Primary Datasets
- Monte Carlo Signal Datasets
- Monte Carlo Background Datasets

DID YOU PRODUCE N-TUPLES AS AN INITIAL STEP TO ANY MEASUREMENT?

BASIC INFORMATION

Please provide some information relevant for all parts of the Analysis here

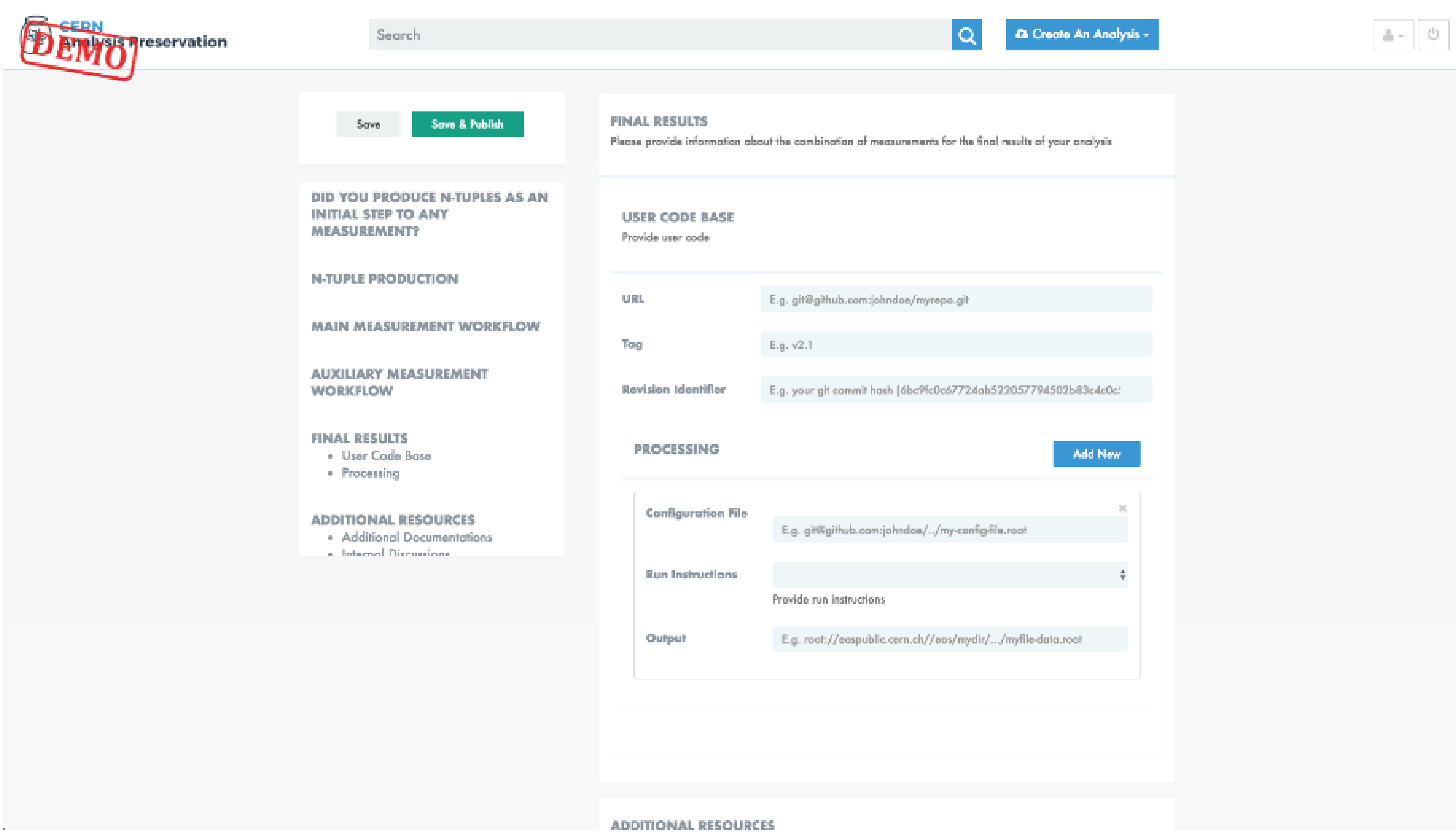
Analysis Number	Please provide CADI analysis number to connect, e.g. CMS-ANA-2012-049
Abstract	If not provided, it can be extracted from the final paper
Conclusion	Please add a short conclusion for the analysis

PEOPLE INVOLVED

Please provide information about the people involved

Names	E.g. John Doe, Jane Doe

Capturing the research outputs



DID YOU PRODUCE N-TUPLES AS AN INITIAL STEP TO ANY MEASUREMENT?

N-TUPLE PRODUCTION

MAIN MEASUREMENT WORKFLOW

AUXILIARY MEASUREMENT WORKFLOW

FINAL RESULTS
- User Code Base
- Processing

ADDITIONAL RESOURCES
- Additional Documentations
- Internal Discussions

FINAL RESULTS

Please provide information about the combination of measurements for the final results of your analysis

USER CODE BASE

Provide user code

URL: E.g. git@github.com:johndoe/myrepo.git

Tag: E.g. v2.1

Revision Identifier: E.g. your git commit hash [6bc9fc0c67724ab522057794502b83c4c0c!

PROCESSING

Configuration File: E.g. git@github.com:johndoe/../my-config-file.root

Run Instructions:

Provide run instructions

Output: E.g. root://eospublic.cern.ch//eos/mydir/.../myfile-data.root

ADDITIONAL RESOURCES

Some metadata fields for a primary dataset

- Title
- Description
- License
- Persistent identifier
- Date issued
- Date modified
- Date available
- Dataset id
- Data type

- Run number
- Number of events
- Number of lumis
- Number of files
- Number of blocks
- Triggers
- Trigger selection
- Run period
- Trigger efficiency
- Event selection
- Event filter

JSON Schemas: https://github.com/cernanalysispreservation/analysis-preservation.cern.ch/tree/master/cap/jsonschemas

Challenge

- So far there is no solution for formally describing this type of experimental results
- We use locally invented, highly specialised metadata fields = not standardised
- It is necessary to describe and be able to search all the elements of an analysis and the knowledge around it, not just high-level information

Thank you

Find us:
https://github.com/cernanalysispreservation
analysis-preservation-development@cern.ch

artemis.lavasa@cern.ch

orcid.org/0000-0001-5633-2459