PAMFILOS FOKIANOS

CERN ANALYSIS PRESERVATION

IMPROVE PRESERVATION AND RE-USE OF
RESEARCH RESULTS AT CERN
CERN ANALYSIS PRESERVATION

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

Development: CERN SIS-OS

Collaboration/Consulting: CERN IT, LHC experiments
Improving discovery, navigation, retrieval, and access to research resources

The project aims to build the infrastructure for persistent identifiers as a core component of open science
Welcome to the CERN Analysis Preservation Portal.

Our mission is to preserve physics analyses to facilitate their future reuse.

Do you want to know more? Check out what the portal is about.
CERN ANALYSIS PRESERVATION

END GOAL

Make the **analysis components** (metadata, files, tables, plots, likelihoods, wikis, etc) easily **reusable** - ex. in workflow engines, scripts, publication writing tools, push to other services (ex. HEPData, Inspire, Zenodo)
CERN ANALYSIS PRESERVATION

OUR GOALS

Build a **stable, flexible, collaborative** tool for physicists to **capture** and **share** their analysis assets (metadata, code, containers, workflows, etc).

Capturing all the elements needed to **understand** and **rerun** an analysis even several years later and **linking them together persistently**.

Sync with and indexing of various **experiment DBs/APIs/sources** for **searching** and **integration**.

Advanced and clean forms  
Easy, automated and efficient **file uploading**

**Powerful API** (records, files, search)
PRESEVATION ↔ RE-USABILITY

CERN Analysis Preservation

ARCHIVE

Physicist

USE

Analysis tools / Workflow engines

PRESEVRE

Ex. reana

jupyter
CERN ANALYSIS PRESERVATION

TECHNOLOGIES

INVENIO

Elasticsearch
PostgreSQL or MySQL
Python/Flask

React

RED HAT OPENShift
FAIR DATA

Established practices for “FAIR” data:

- **Findable**
  - persistent identifiers
  - rich metadata
  - indexed and searchable

- **Accessible**
  - retrievable by identifiers
  - standard protocols
  - metadata vs data accessibility

- **Interoperable**
  - knowledge representation language
  - common vocabularies
  - references to other metadata and data

- **Reusable**
  - domain-relevant attributes and community standards
  - clear licensing
  - provenance tracking
MAXIMIZING LHC OUTCOME

Maximize the LHC outcome

accessible data
LHC data preservation efforts.

accessible simulation
Delphes, BuckFast, Falcon, ...

accessible results
Experimental papers, HEPDATA, twikis...

accessible analysis information
analysis preservation / standard analysis description

Workshop on Analysis Description Languages, 6-8 May 2019, Fermilab LPC
CERN ANALYSIS PRESERVATION

FEATURES
CERN ANALYSIS PRESERVATION
ADVANCED FORMS

J/psi and psi(2S) prompt double differential cross sections in pp collisions at 7 TeV

Input Data
Please list all datasets and triggers relevant for your analysis here

Primary Datasets
/T1tb_GIGToTBTB_38xFall10/StoreResultsT1tb_GIGToTBTB_38xFall10/USER
/T1T1_2BC_200_50-8TeV_madgraphSummer12-START52_Y9_PSIM-v1/GEN

Monte Carlo Signal Datasets

Monte Carlo Background Datasets

Official JSON files

Gitlab repositories of the analysis

Docker images of the analysis

Additional Repositories

Basic scripts

Command to execute code

Files/Scripts

Combine Datacard File

Open File Manager

Statistics Questionnaire

Import

Copyright 2018 © CERN. Created & Hosted by CERN. Powered by Invenio Software.
CERN ANALYSIS PRESERVATION

Automated fetching from experiment DBs/APIs/other sources

Information from CADI database

Name
Search for Black Holes and Sphalerons

Description
Search for Black Holes and Sphalerons (full 2016 data)

Contact Person
Ka Hei Martin Kwok (BROWN-UNIV)

Twiki
https://twiki.cern.ch/twiki/bin/view/CMSSW/EXO17023

Created
03/10/2017

Paper

PAS

Publication Stage
CERN ANALYSIS PRESERVATION

JSON/YAML uploader/editor

Validates against provided JSON Schema

ex. ATLAS workflow capture
CERN ANALYSIS PRESERVATION

Running workflows

REANA Hello World Example

Create now a simple workflow that runs on REANA cluster to now a simple workflow that runs on REANA cluster

Try it now!

REANA Hello World Example

Engine: reana
Name in REANA: wine-seagull
Run: #1 (wine-seagull:1)

Workflow Info

Inputs

ADD FILE

codet/helloworld.py

Specifications

Output

Logs

CHEP 2019 Posters - REANA - R17 P409 - link
INTEGRATIONS

connect services account for seamless imports/exports

easy/automated integration with code
push/pull capabilities
CERN ANALYSIS PRESERVATION

CAP CLIENT

Seamless and easy contributions for users work

From services, scripted or with CI

Push / pull metadata, code, files

$ pip install cap-client

$ export CAP_SERVER_URL=https://analysispreservation.cern.ch/

$ export CAP_ACCESS_TOKEN=<generated access token from server>

$ cap-client files list --pid/-p <existing pid>

```
[
    {
      "checksum": "md5:f0428126e7cf7b0d4af7091c68ae2a9f",
      "filename": "file.json",
      "filesize": 25,
      "id": "25852e50-be6d-47a5-897b-1f3df015fac7"
    },
    {
      "checksum": "md5:926fb9c44251d70614ee42d34c5365b6",
      "filename": "Analysis_Notes_07112019.pdf",
      "filesize": 160898,
      "id": "89743c9b-106d-4235-8e96-23a164c7b1f4"
    }
]
```

$ cap-client me

```
{
  "collaborations": [
    "ATLAS",
  ],
  "id": 1,
  "email": "atlas-user@cern.ch"
}
```
### Advanced Search Tips

To perform a free text search, simply enter a text string. This will search all the fields for the given term.

If you look for a value in a specific field, prefix the value with the name or alias of this field.

```
object:electron
```

To search for an exact phrase, you will need to enclose the entire phrase in quotation marks.

```
researcher:"John Doe"
```

You can use wildcards in your queries (**” for a single character and***” for multiple ones). Keep in mind that none of these can be used as first character in your search.

```
dataset:MicroMegas
```

You can build more complex search criteria using the Boolean operators **AND**, **OR**, and **NOT**.

```
dataset:"MicroMegas" AND trigger:"HLT_Mu12IsoRec OR HLT_Photon135"
```

To point to nested fields in your analysis, use **." operator or one of available aliases.

```
researcher reviewer analyze status keyword dataset trigger object
```

You can find more search tips using Lucene query syntax [here](#).
CERN ANALYSIS PRESERVATION

Visualize data
CONTENT BUILDER

Create your own models and start preserving
EXAMPLES & DOCUMENTATION

REST API (OpenAPIv3 specs) and `cap-client` documentation

CERN Analysis Preservation REST API (1.0.0)

Download OpenAPI specification: Download

E-mail: analysis-preservation-support@cern.ch | URL: https://analysispreservation.cern.ch/about | License: Apache 2.0 | Terms of Service

CERN Analysis Preservation offers a REST API to access the service independently from the web interface. This allows users to automate specific tasks or create their own data interface, using simple HTTP calls.

Click here if you are looking for our development documentation.

Introduction

CERN Analysis Preservation provides:
- resource URLs.
- HTTP response codes to indicate:
  - accepting and returning JSON in the response.

You can use your favorite HTTP libraries currently available in Python. Every endpoint is easily accessible for automation.

Authentication

If you want to gain access to CERN Analysis Preservation, you need to authenticate with the API. The CAP offers the following forms of authentication:
- API Key: `access_token`

Access Token

Your access token will allow you to use the same permissions as specified in your application. Do not share your personal access tokens with others.

Deposits

The Deposit resource is used for uploading and editing records on CAP.

Create a new draft

To create a new draft you need to pass a correct analysis type (e.g., `ana_type`) OR a correct JSON schema.

JSON Schemas and analysis types (`ana_type`) depend on the access rights you have and mainly correspond to the CERN collaborations you are part of (ALICE, ATLAS, CMS, LHCb).

To get the list of available schemas you can call the API:

```
POST /deposits
```

Request samples

```
Payload
```

Content type: `application/json`

```
Example:
```

```
  simple
```

Response samples

```
 responded
```

Content type: `application/json`

```
Example:
```

```
  
```
  
```

Directions:

- [View Content](#content)
- [View Source](#source)
- [Report a Bug](#bug)
- [Share](#share)

Download OpenAPI specification: Download

E-mail: analysis-preservation-support@cern.ch | URL: https://analysispreservation.cern.ch/about | License: Apache 2.0 | Terms of Service

CERN Analysis Preservation offers a REST API to access the service independently from the web interface. This allows users to automate specific tasks or create their own data interface, using simple HTTP calls.

Click here if you are looking for our development documentation.
CERN ANALYSIS PRESERVATION

WHO IS IT FOR?

Individual Physicists

Experiments/Collaborations/WGs

Institutions
CERN ANALYSIS PRESERVATION
COLLABORATIONS, STATUS AND FUTURE
Welcome to the CERN Analysis Preservation Portal.

Our mission is to preserve physics analyses to facilitate their future reuse.

Do you want to know more? Check out what the service is about.

Web: [https://analysispreservation.cern.ch](https://analysispreservation.cern.ch)
Github: [https://github.com/cernanalysispreservation](https://github.com/cernanalysispreservation)
Email: analysis-preservation-support@cern.ch
Twitter: @analysispreserv
PORTALS, FRAMEWORKS AND TOOLS

INVENIO

Web: http://inveniosoftware.org
Github: http://github.com/inveniosoftware
Email: info@inveniosoftware.org
Twitter: @inveniosoftware

reana

Web: http://www.reanahub.io
Github: http://github.com/reanahub
Email: info@reanahub.io
Twitter: @reanahub
THANK YOU