

Open Data, Open Source, Open Science

Tim Smith in-lieu of WG member: Jose Benito Gonzalez

at the Open Science Strategy Working Group 2021-07-08

Mission

- IT/Collaboration Devices and Applications Group

- Provide tools to power the practice of Open Science

- Collaboration, communication, collaborative code development / article editing, reviewing

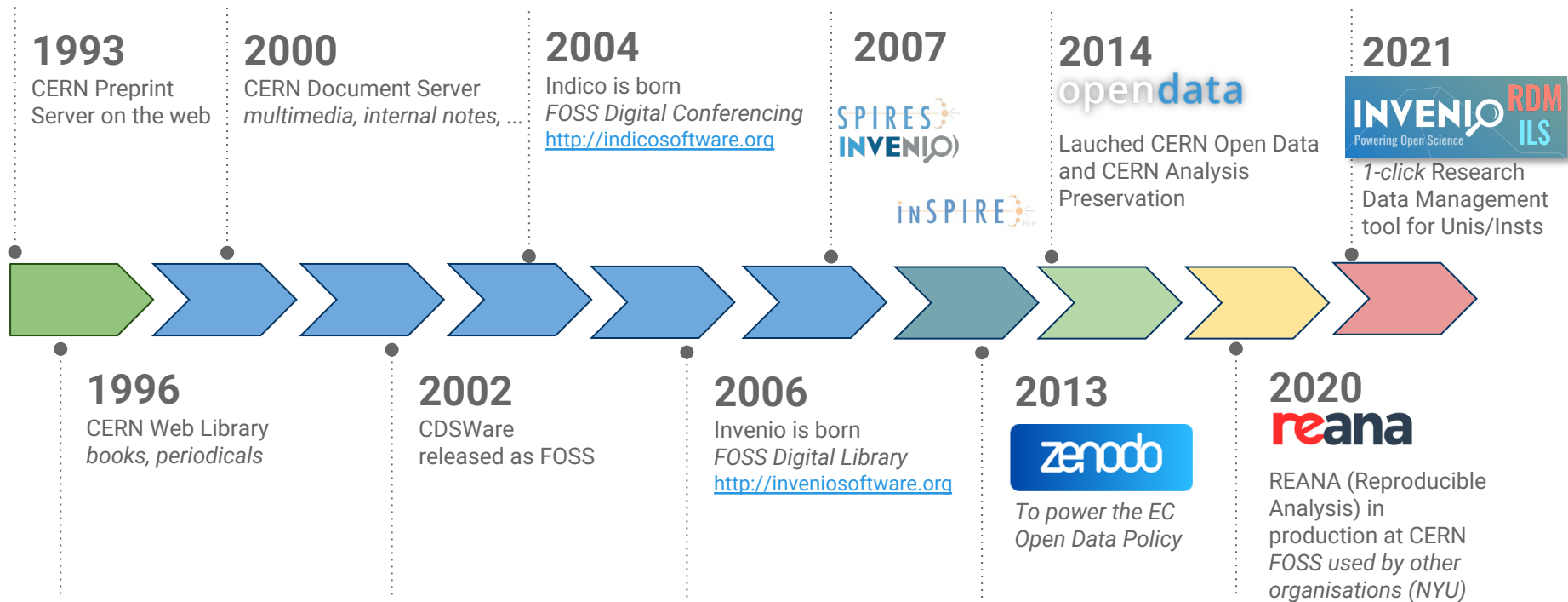


- IT/CDA/Digital Repositories Section

- Provide tools to capture, preserve, disseminate the products of Open Science



Innovating in Openness



OpSc Pillar 1: Publish & Share

CERN Document Server
Access articles, reports and multimedia content in PDF

Search 538,436 records for: [input] [Search] Search Tips Advanced Search

- Articles & Preprints (435,916)
 - CERN Articles & Preprints (333,246)
 - CERN Published Articles (13,817) Preprints (138,070)
 - Theses (5,719) Reports (8,740) CERN Notes (8,123)
 - Committee Documents (3,646)
- Books & Proceedings (6,172)
 - Books (1,128) Proceedings (770)
 - Design Reports (156)
- Lectures & Events (55,980)
 - Conference Announcements (144)
 - Academic Training Lectures (89)
 - Student Lectures (147)
 - Table seminars and other events (10,071)
 - E-learning modules (6)
- Periodicals & Progress Reports (176)
 - Periodicals (76) Progress Reports (76)

Institutional
>100 publication workflows

opendata CERN

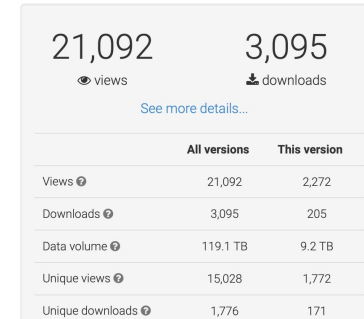
CMS releases heavy-ion data from 2010 and 2011
2020-12-21 by CMS Collaboration

The OPERA Collaboration releases new data samples: neutrino-induced charmed hadron production and electron neutrino interactions
2020-05-20 by OPERA Collaboration

Examples of basic analysis techniques for neutral meson and photon data from the PHENIX detector
David, Gabor; Potekhin, Maxim; PHENIX collaboration

Thematic

Usage Statistics:



Gene Ontology Data Archive

Carbon, Seth; Mungall, Chris



The CodeMeta Project

zenodo

FAIR Data Advanced Use Cases: from principles to practice in the Netherlands

13,762 | 718

OpenAIRE

Versioning:

Versions

- Version 3 10.5281/zenodo.999150 Sep 29, 2017
- Version 2 10.5281/zenodo.803266 Jun 6, 2017
- Version 1 10.5281/zenodo.35177 Dec 10, 2015

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.620228. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

GitHub Guides

Making Your Code Citable
10 minute read

Digital Object Identifiers (DOIs) are the backbone of the academic reference and metrics system. If you're a researcher writing software, this guide will show you how to make the work you share on GitHub citable by archiving one of your GitHub repositories and assigning a DOI with the data archiving tool Zenodo.

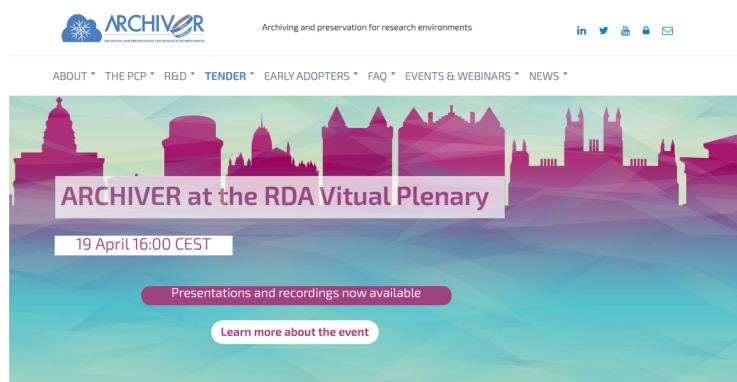
Intro
Choosing Your Repo
Login to Zenodo
Check Repo Settings
Create a New Release
Minting a DOI
Finishing up

Multi-Disciplinary

OpSc Pillar 2: Preservation & Long-term Access

Digital Memory project:

Create long-term preservation platform based on FOSS Archivematica



Participating in Archiver EU project



OpSc Pillar 3: Reproducibility

reana

Home Examples Get Started Documentation News Contact

reana

Reproducible research data analysis platform

Flexible

Run many computational workflow engines.

Scalable

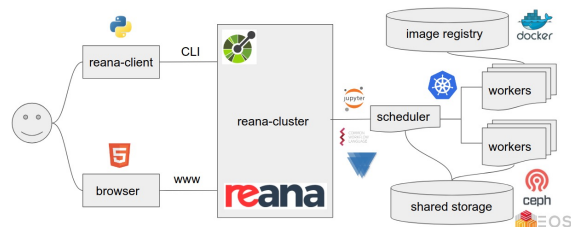
Support for remote compute clouds.

Reusable

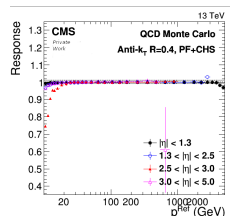
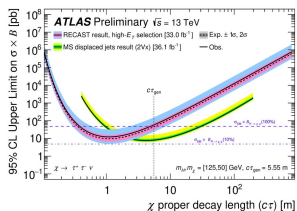
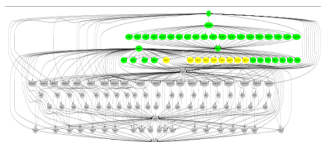
Containerise once, reuse elsewhere. Cloud-native.

Free

Free Software. MIT licence. Made with ❤️ at CERN.



Running containerised workflows on remote compute clouds



Computational recipes: notebooks & workflows

REANA mini-workshop 2021-02-24

Wednesday 24 Feb 2021, 14:00 → 20:00 Europe/Zurich

Description

A mini-workshop amongst REANA users and developers. Sharing tips and tricks on identifying what needs improvement. The mini-workshop is split into several topical blocks. The organization is very informal, we are strongly looking at providing a point in time and space to meet and work on some identified concrete topics in an 'open office' manner. Please feel free to drop by and leave and drop in again anytime throughout the event! We should have regular coffee breaks throughout the afternoon, but we shall always try to start the formal session at an indicated hour.

Live discussion: [@https://mattermost.usc.edu/CERN](https://mattermost.usc.edu/CERN) [@https://github.com/reana](https://github.com/reana)

Videoconference Rooms

- REANA mini-workshop 2021-02-24

14:00 - 15:00 CMS Jet Energy Correction case study

A walk through the CMS Jet Energy Correction case study, showing the problem domain, the code organization, the workflow specifications, the usage of REANA, discussing lessons learnt, what worked well, what needs improvement. Followed by an open floor debate.

Speaker: Adriano Emiliotti (LNF-INFN) (LNF-INFN)

15:00 - 16:00 Using HTCondor compute backend

A walk through the HTCondor compute backend, followed by identifying what needs to be done to improve REANA \leftrightarrow HTCondor bridge.

Speaker: Saurabh Mishra (CERN) (CERN)

16:00 - 17:00 ATLAS RECAST

A walk through the ATLAS RECAST conceptualization example, followed by discussing differences when using computational workflows on RECAST and REANA, and a hands-on session to identify each system compatibility issues.

Speaker: Daniela MacCioni (CERN) (CERN)

17:00 - 18:00 Using Python API

A walk through REANA server options for validating container images and workflow specifications, to catch possible problem early. Discussing most common issues with an image registry container images, forcing certain OS and GID. Followed by a contribution regarding the validation checks.

Speaker: REANA Team (REANA)

18:00 - 19:00 Container and workflow validation

A walk through REANA server options for validating container images and workflow specifications, to catch possible problem early. Discussing most common issues with an image registry container images, forcing certain OS and GID. Followed by a contribution regarding the validation checks.

Speaker: REANA Team (REANA)

19:00 - 20:00 Open floor

Raise any REANA-related topics of your interest! Open floor for discussion and non-facilitators amongst REANA users and developers.

Current / Anticipated Challenges

- COD: match central / experiment resources for metadata / documentation
- Zenodo: keep up with expectations – development resources
- Zenodo: mechanism to accept payments for large datasets
- InvenioRDM: maintaining vibrant FOSS community
- REANA: productionisation and scale-out



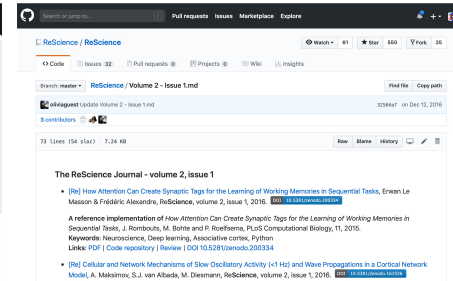
The Future

*Updating our understanding
of scientific impact in light of
open scholarship*



The Leiden Manifesto
for research metrics

- Decomposing / exposing scientific pipelines
- Cite / value / credit all contributions
- Open peer review / Modular peer review



- We are building strong independent FAIR services for archival, access, sharing and reproducing research
- We will work on interoperability and ease of use by researchers

