Big Data archiving and publication
The Zenodo experience

ICEDIG Opening Conference – 2018/03/06
Long Term

Rate: 50 PB/y  150 PB/y  600 PB/y

Total: 240 PB  EB  ZB
Open Data as a Service

2009
Open Access Mandate

2012
Open Data Pilot
Findable & Interoperable

Grant DB: FP7, H2020 …WT (UK), FCT (Portugal), ARC & NHMRC (Australia)

Repository: Institute National Subject

JSON

JavaScript Object Notation

{ }

} } JSON

schema

@TimSmithCH
Communities

LORY - Lucerne Open Repository

Recent uploads

@TimSmithCH

Projects, Subjects, Institutes, Nations, Conferences, …
Open Scholarship: Share ≠ Publish ≠ Preserve

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.

PreTip: This tutorial is aimed at researchers who want to cite GitHub repositories in academic literature. Provided you’ve already set up a GitHub repository, this tutorial can be completed without installing any special software. If you haven’t yet created a project on GitHub, start first by uploading your work to a repository.

Decouple and recouple - Edit

DOI: 10.5281/zenodo.8735

75% of world SW DOIs

@TimSmithCH
Publication moment is too late!
Capture before the knowledge evaporates
While data is live

Accessible

Closed access

Embargoed access

Open access

Restricted access

Share unpublished data with reviewers
Share sensitive data with controlled list
Building system to **instantiate** preserved analysis **on the cloud**

Supporting **multiple**…
- computing clouds
- running environments
- resource orchestration
- workflow engines
- shared storage systems

Features

9 funders
660k grants
TRL9 Service

- Data: 151 TB (38 TB logical)
- Files: 1.2 M
- Published Records: 378 k
- Largest File: 236 GB
- Largest Dataset: 682 GB
- 30 nodes
- QA staging cluster
- Spam using ML
Biodiversity Literature Repository

- 438 GB (1.2%)
- Records: 208 k (55%)
• Research Methods
• Research Packages
• Things ?
EU IST FP7 ULOOP Project Publications

Recent uploads

D5.2: Standardisation Report
Zhou, Q.;
This document corresponds to deliverable 5.2 of the ULOOP project. The document provides a description of the standardization related activities conducted during the first two years in Task 5.2, including an overview of the relevant standardization bodies, tool and platform and differences and similarities.
Uploaded on February 16, 2016

D3.6: ULOOP Mobility Aspects Specification and Refined Software
Zhou, Q.;
This document corresponds to deliverable 3.6 of the ULOOP project. Deliverable D3.6 includes two parts: software which implements mobility aspects function, and this document, which provides a description of the specification and software developed in ULOOP 1.3 Mobility Aspects, as well as coining.
Uploaded on February 16, 2016

D4.1: Pilot Setup Report
Yildiz, Muerse:
This document represents deliverable D4.1 of the EU FP7 ICT project ULOOP (User-centric Wireless Local Loop, grant number 257418). The document is intended to provide a roadmap and instruction sets for successful deployment of
Thank You! Questions?

Tim.Smith@cern.ch
0000-0002-1567-7116
http://cern.ch/tim.smith

http://zenodo.org
http://www.openaire.eu
http://invenio-software.org
Lars.Nielsen@cern.ch