

# EOS for Invenio digital library framework

Lars Holm Nielsen

CERN IT, Digital Repositories





## Recent uploads

November 12, 2015

Presentation

Open Access

### UN Sustainable Development Goals: What are the SDGs and what do they mean for libraries?

Bradley, Fiona

Invited speaker at the EIFL General Assembly, National Library of Latvia, 12-14 November 2015. Outlines the UN 2030 Agenda for Sustainable Development, the advocacy process to adopt the Sustainable Development Goals, and opportunities for libraries at the national level.

Uploaded on January 31, 2017.

View

January 30, 2017

Journal article

Open Access

### Guidelines for collecting vouchers and tissues intended for genomic work (Smithsonian Institution): Botany Best Practices

Funk, Vicki; Gostel, Morgan; Devine, Amanda; Kelloff, Carol; Wurdack, Kenneth; Tuccinardi, Chris; Radosavljevic, Aleks; Peters, Melinda; Coddington, Jonathan

The introduction of Next Generation Sequencing into the disciplines of plant systematics, ecology, and metagenomics, among others, has resulted in a phenomenal increase in the collecting and storing of tissue samples and their respective vouchers. This manual suggests standard practices that ...

Uploaded on January 30, 2017.

View

January 29, 2017

Dataset

Open Access

S-NRR VIIRS SDR data 2017-01-28 12:29 UTC

View

### Sep 12: Major update

Welcome to the improved Zenodo. See [what's new](#) and [known issues](#).



### Using GitHub?

Just [Log in](#) with your GitHub account and [click here](#) to start preserving your repositories.



### Zenodo in a nutshell

- **Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Communities** — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
- **Funding** — identify grants, integrated in

August 14, 2016

Dataset Open Access

# Cosmic Sculpture

Clements, David L; Sato, Suzu; Portela Fonseca, Ana

README.txt - readme file

128\_scaled.stl - STL file used for printing the monochrome version of the project as shown in Fig. 4 of the paper

cmbhollow.wrl - VRML file used for printing the hollow coloured version of the project as shown in Fig. 5 of the paper

3Dprinting.mov - movie file showing an Ultimaker2 printing the monochrome version of the project

Files		
Name	Size	
<a href="#">128_scaled.stl</a> md5:7f25ba7068ccd96a0b14236ad3bbcc67 ?	8.6 MB	<a href="#">Download</a>
<a href="#">3Dprinting.mov</a> md5:62ecdbd2bf49fa14e915ec70cb390769 ?	236.5 MB	<a href="#">Download</a>
<a href="#">cmbhollow.wrl</a> md5:2013435c029569b718386d3d3a1e47c5 ?	10.3 MB	<a href="#">Download</a>
<a href="#">README.txt</a> md5:e5149a0d9691b600eef5815c93d51e3b ?	900 Bytes	<a href="#">Download</a>

[See more details](#)

Picked up by **27** news outlets  
Blogged by **5**  
Tweeted by **120**  
On **2** Facebook pages  
Mentioned in **2** Google+ posts

**Publication date:**

August 14, 2016

**DOI:**DOI [10.5281/zenodo.60215](#)**Keyword(s):**[Cosmic Microwave Background](#) [3D Printing](#)  
[Cosmology](#) [Planck Mission](#)**Related identifiers:**

Supplement to:

[10.1088/0143-0807/38/1/015601](#)**Communities:**[Zenodo](#)**License (for files):**[Creative Commons Attribution-NonCommercial 4.0](#)

August 14, 2016

Dataset

Open Access

# Cosmic Sculpture

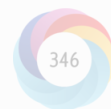
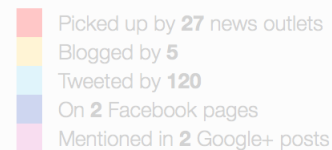
Clements, David L; Sato, Suzu; Portela Fonseca, Ana

README.txt - readme file

128\_scaled.stl - STL file use

cmbhollow.wrl - VRML file u

3Dprinting.mov - movie file

[See more details](#)

D Printing

Files

Name

128\_scaled.stl

md5:7f25ba7068ccd96a0b14

3Dprinting.mov

236.5 MB

[Download](#)

md5:62ecdbd2bf49fa14e915ec70cb390769

cmbhollow.wrl

10.3 MB

[Download](#)

md5:2013435c029569b718386d3d3a1e47c5

README.txt

900 Bytes

[Download](#)

md5:e5149a0d9691b600eef5815c93d51e3b

Supplement to:

10.1088/0143-0807/38/1/015601

**Communities:**[Zenodo](#)**License (for files):**[Creative Commons Attribution-NonCommercial 4.0](#)

# CERN Document Server

Search for over than 1.000.000 records

About Submission Help Sign in



High Energy Physics Data Repository

This new site replaces the old site at <http://hepdata.cedar.ac.uk>.

Search on 8409 publications and 68252 data tables.

Search for a paper, author, experiment

e.g. reaction  $PP \rightarrow LQ LQ X$  title has "p"

Data from



ALICE

View Data

Recently Updated

Images

opendata  
CERN

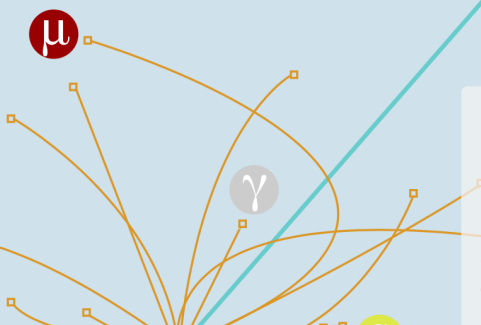
ABOUT SEARCH EDUCATION RESEARCH

## Education

Visualise events, check reconstructed data, run tools or build your own!

## Research

Get the genuine working environments, virtual machines and datasets to start your research



inspirelabs

LITERATURE AUTHORS DATA CONFERENCES JOBS INSTITUTIONS



INSPIRE Labs provides a sneak preview of new features and designs currently under development.

Try it out, and please use the feedback button to let us know what you think.

Search 0 articles

Search literature

How to search

quark multiplicity  
5048  
for-ellis -refersto.a  
all | authors:andru

# History (Invenio v1/v2)

- Invenio 1.x:
  - File ingestion would happily copy files 3 times.
- AFS:
  - 100GB volumes -> app aware of volume layout.
  - AFS overload -> kernel hang + dead service
- Every other month:
  - *“Dear Jan, could you please increase quota with 1TB.”*



# What we wanted (Invenio v3)...

- Support large file uploads over HTTP (100GB+)
- No knowledge of “storage system internals”
- Efficient use of disk space (e.g. when versioning large datasets with few changes).
- Resilience against unavailability of storage system.

# First EOS experience (S3)

```
fp = open('my-local-file')  
eos_s3.setcontents("my-local-file", fp)  
eos_s3.makepublic("my-local-file")
```

# First EOS experience (S3)

```
fp = open('my-local-file')  
eos_s3.setcontents("my-local-file", fp)  
eos_s3.makepublic("my-local-file")
```

File deleted!!!!

# XRootD + file system abstraction

- Invenio: Python/Flask application
- XRootD Python bindings + PyFilesystem
- = **XRootDPyFS**

```
>>> import xrootdpyfs
>>> from fs.opener import opener
>>> fs, path = opener.parse("root://localhost//tmp/")
>>> fs.listdir("xrootdpyfs")
[u'test.txt']
```

# Happy days :-)

- Awesome stuff you can do with EOS/XRootD:
  - offload checksum calculation
  - random seek
  - liberal disk quotas ;-)

# Large file uploads over HTTP

- Idea:
  - Client: Split file and upload files in concurrent HTTP requests.
  - EOS: Write to same file to avoid having to merge parts.

# Large file uploads over HTTP

- **FAIL**
  - Client: Split file and upload files in concurrent HTTP requests.
  - EOS: Write to same file to avoid having to merge parts.
- Instead: Binary stream in single HTTP request

# Python app server

- Gunicorn WSGI server:
  - Read loop: 8kb - **FAIL**
- UWSGI python app server:
  - Very low timeouts!!!



# Other gotchas

- Authentication:
  - ~~k5reauth~~ / k5start (supervisord)
- EOS booking size:
  - Browser XHR requests: not all browsers providing Content-Length.
- XRootD python bindings in a Python virtualenv.

# Planning

- Use EOS geo replication features (i.e. ask Luca).
- Offload file serving to EOS HTTP gateway.
- CASTOR integration (i.e. use eosarchive)

# Wish list

- XRootD python bindings:
  - Python 3 support
  - Pythonic install (i.e. `pip install pyxrootd`)

# Thanks EOS!



**Thomas Robitaille**

@astrofrog

Following



Just uploaded 33Gb of data to [@ZENODO\\_ORG](#) in 20 minutes. Mind blown!