## Status of the EOSCpilot Scientific Demonstrators

Giuseppe La Rocca, 10/04/2018

Sommario

[Status of the EOSCpilot Scientific Demonstrators 1](#_Toc511222186)

[Summary report from first round of SDs (Jan 2017-Dec. 2017): 2](#_Toc511222187)

[Science Demonstrator title: DPHEP 2](#_Toc511222188)

[Science Demonstrator title: TextCrowd 2](#_Toc511222189)

[Science Demonstrator title: PhotonScience 2](#_Toc511222190)

[Science Demonstrator title: PanCancer 2](#_Toc511222191)

[Science Demonstrator title: ERFI 3](#_Toc511222192)

[Summary report from second round of SDs (July 2017-June 2018): 4](#_Toc511222193)

[Science Demonstrator title: CryoEM 4](#_Toc511222194)

[Science Demonstrator title: EPOS/VERCE 4](#_Toc511222195)

[Science Demonstrator Title: LOFAR 5](#_Toc511222196)

[Science Demonstrator title: Life Science datasets 5](#_Toc511222197)

[Science Demonstrator title: Prominence 5](#_Toc511222198)

[Summary report from third round of SDs (Dec. 2017-Nov 2018): 7](#_Toc511222199)

[Science Demonstrator Title: BioImaging 7](#_Toc511222200)

[Science Demonstrator Title: Frictionless Data Exchange 7](#_Toc511222201)

[Science Demonstrator Title: VisIVO 7](#_Toc511222202)

[Science Demonstrator Title: Hydrology 8](#_Toc511222203)

[Science Demonstrator Title: VisualMedia 8](#_Toc511222204)

## Summary report from first round of SDs (Jan 2017-Dec. 2017):

### Science Demonstrator title: DPHEP

Goal: The overall goal of this Science Demonstrator is to implement a service for the long-term preservation and re-use of HEP data, documentation and associated software.

Consortium and collaborator organisations: CERN (CH)

Scientific Contact: Jamie Shiers (CERN), Jamie.Shiers@cern.ch

WP5 contacts/Main Shepherd: John Kennedy, jkennedy@mpcdf.mpg.de

Deputy: Matthew Viljoen, matthew.viljoen@egi.eu

**Current Status: Final report has been submitted.**

### Science Demonstrator title: TextCrowd

Goal: Support the encoding/metadata enrichment of text documents that are the main part of datasets used in Digital Humanities and Cultural Heritage research.
Consortium and collaborator organisations: PIN-University of Florence

Scientific Contact: Franco Niccolucci (PIN), franco.niccolucci@gmail.com

WP5 contacts/Main Shepherd: Kathrin Beck, kathrin.beck@mpcdf.mpg.de
Deputy: Thomas Zastrow, thomas.zastrow@mpcdf.mpg.de

**Current Status: Final report has been submitted. Interested to use B2DROP to store/share Metadata. Scheduled a meeting with the EUDAT provider on Mon 30 April.**

### Science Demonstrator title: PhotonScience

Goal: Demonstrate how cloud environments can help researchers in this scientific area to store, analyse and share experimental data.

Consortium and collaborator organisations: DESY, EMBL, ESRF, ESS, European XFEL, ILL.
Scientific Contact: Volker Gülzow (DESY), Volker.Guelzow@desy.de

WP5 contacts/Main Shepherd: Sune Rastad Bahn, Sune.RastadBahn@esss.se Deputy: Michael Schuh, michael.schuh@desy.de

**Current Status: Pre-final report has been submitted.**

**Provided settings to configure EGI AAI Check-In with the OpenStack dashboard.**

### Science Demonstrator title: PanCancer

Goal: Support the operation of the Butler application on multiple clouds.
Scientific Contact: Sergei Iakhnin (EMBL), iakhnin@embl.de

Consortium and collaborator organisations: EMBL, EMBL-EBI

WP5 contacts/Main Shepherd: Dario Vianello, Dario@ebi.ac.uk
Deputy: Gergely Sipos, gergely.sipos@egi.eu

**Current Status: Pre-final report has been submitted.
CYFRONET addressed scalability issues in the context of the HNSciCloud project.**

**Testing the new set-up in EOSCpilot.**

### Science Demonstrator title: ERFI

Goal: The overall goal of this Science Demonstrator is to demonstrate dynamics of greenhouse gases, aerosols and clouds and their roles in radiative forcing.

Consortium and collaborator organisations: ICOS ERIC, IS-ENES2 (DKRZ), IS-ENES2 (IPSL), ENVRIplus, ACTRIS.

Scientific Contact: Werner L Kutsch, werner.kutsch@icos-ri.eu

WP5 contacts/Main Shepherd: Giuseppe La Rocca, giuseppe.larocca@egi.eu

**Current Status: ONGOING**

**ICOS has to check/validate the IS-ENES datasets injected in Onedata.**

## Summary report from second round of SDs (July 2017-June 2018):

### Science Demonstrator title: CryoEM

Goal: Enhancing the Scipion application in order to link together raw data, metadata and tools used to produce a scientific workflow.

Consortium and collaborator organisations: CSIC (Spain)

Scientific Contact: Carlos Sanchez Sorzano (CNB), coss@cnb.csic.es
WP5 contacts/Main Shepherd: Gergely Sipos, gergely.sipos@egi.eu
Deputy: Erik van den Bergh, evdbergh@ebi.ac.uk

**Current Status: ONGOING**

**Extended and improved the demonstrator in order to make it compliant with the FAIR principles. Scipion now writes the image processing pipeline used at electron microscopy facilities in a JSON file. This JSON can be exported and imported from Scipion, and it can be deposited in public databases as EMPIAR. EMPIAR will integrate a web viewer specifically designed for this kind of files. Working on the 8 months report.**

### Science Demonstrator title: EPOS/VERCE

Goal: Produce data products such as simulated seismic waveform images, wave propagation videos, 3D volumetric meshes, sharable KMZ packages and parametric results.

Consortium and collaborator organisations: University of Liverpool (UK), KNMI (NL), INGV (IT), SCAI (DE)
Scientific Contact: Rietbrock, Andreas, A.Rietbrock@liverpool.ac.uk

WP5 contacts/Main Shepherd: Giuseppe La Rocca, giueppe.larocca@egi.eu
Deputy: Michael Schuh, michael.schuh@desy.de

**Current Status: ONGOING**

**Three cloud providers of the EGI Federation have been already identified.
The Provenance system (S-ProvFlow) has been improved in many aspects: better Rest API methods, provenance repository performances, frontend usability and modular “Dockerisation” of each component.**

**A new DCI\_BRIDGE VM image is now in the EGI AppDB and ready for testing.**

**The VERCE portal has extended the AAI framework to enable the access with the EGI AAI Check-In service using the OIDC module developed by INFN for Science Gateways based on Liferay technology.**

**Testing the integration with the EGI FedCloud infrastructure in progress.**

**IMPORTANT: The EPOS AAI Team suggested to wait a bit before to integrate the portal with the official EPOS core AAI infrastructure (based on UNITY) since it is still not ready. This is the reason why the OIDC module has been used in the context of the EOSCpilot project to speed the pilot activities. They will probably migrate to the official EPOS AAI once it is ready.**

### Science Demonstrator Title: LOFAR

Goal: Allow for the science community to locate, access, and extract science from the LOFAR archive without being an expert on data retrieval and data analysis tools. The pilot will develop services, based on existing tools such as Xenon, CWL, Docker, Virtuoso, that

allow users to initiate processing on data stored in a distributed, large-scale archive.

Consortium and collaborator organisations: ASTRON (NL), SURFsara (NL), INAF (IT), SFTC (UK), CWL (LT), ITL

Scientific Contact: Hanno Holties, holties@astron.nl

WP5 contacts/Main Shepherd: Thomas Zastrow, thomas.zastrow@mpcdf.mpg.de

Deputy: John Kennedy, jkennedy@mpcdf.mpg.de

**Current Status: ONGOING**

**Provided access to EUDAT resources B2SHARE (AAI via B2ACCESS) needs more testing. Provided accessed to PSNC/FZJ resources.**

**CWL implementation of three pipelines working.**

**Investigating FAIR services for LOFAR.**

**The 8 months report has been submitted.**

### Science Demonstrator title: Life Science datasets

Goal: Consume third party datasets stored at European Genome-phenome Archive (EGA) , consume reference datasets and update datasets.

Consortium and collaborator organisations: Centre for Genome Regulation (CGR)

Science Contact: Jordi Rambla De Argila, jordi.rambla@crg.eu

WP5 contacts/Main Shepherd: Erik van den Bergh, evdbergh@ebi.ac.uk

Deputy: Matthew Viljoen, matthew.viljoen@egi.eu

**Current Status: ONGOING**

Pipelines now working with NextFlow. They are collecting the computational requirements they may need to continue this pilot and working on the preparation of a document to describe the basic security policies needed to analyse datasets in an external resource providers.

 **The 8 months report has been submitted.**

**Collecting the security requirements and policies needed to run the workflows in other providers. Interested to use the B2FIND Metadata Catalogue. A technical meeting with the B2FIND provider has been organized.**

**ACTION: EGA Dataset has to provide the metadata schema to be used by B2FIND.**

###

### Science Demonstrator title: Prominence

Goal: Primarily Energy and Plasma Energy. Access to HPC class nodes for the Fusion Research community through a cloud interface

Consortium and collaborator organisations: CCFE (UK), Chalmers University (SE), MPIPP (DE)
Scientific Contact: Shaun de Witt, shaun.de-witt@ukaea.uk

WP5 contacts/Main Shepherd: John Kennedy, MPCDF, john.kennedy@mpcdf.mpg.de
Deputy: Frank Schluenzen, DESY, frank.schluenzen@desy.de

**Current Status: ONGOING**

Containerized MPI-based applications and successfully run them both on EGI and commercial clouds. The clusters were initially created as static SLURM clusters, using IM through the EC3 and CLUES, and later made dynamic using openVPN to create workers on multiple cloud instances using both OpenNebula and OpenStack.

**The 8 months report has been submitted.**

**Used the IM to facilitate the deployment of the SLURM clusters in the cloud infrastructure. Access to the EGI FedCloud infrastructure is successfully: (133 vCPU cores at CESNET-MetaCloud, 8 vCPU cores at IN2P3-IRES and CESGA).**

## Summary report from third round of SDs (Dec. 2017-Nov 2018):

### Science Demonstrator Title: BioImaging

Goal: The overall goal of this science demonstrator is to perform comprehensive machine learning analyses on these datasets, with the ultimate goal of identifying functional connections between genes and/or small molecules that target them based on image-based phenotypes.

Scientific Contact: Jean-Karim Heriche, heriche@embl.de

Consortium and collaborator organisations: University of Dundee (UK), EMBL (DE), EMBL-EBI (UK) related to Euro-Bioimaging.

WP5 contacts/Main Shepherd: Dario Vianello, Dario@ebi.ac.uk
Deputy: Erik van den Bergh, EBI evdbergh@ebi.ac.uk

**Current Status: ONGOING**

**Configuring a tenant on the Embassy Cloud at EBI to run the workflow.**

**Interested in some high-level tools to orchestrate VMs deployment. Possible interest in the INDIGO orchestrator solution.**

### Science Demonstrator Title: Frictionless Data Exchange

Goal: The proposed SD will work to pilot a demonstrator service for fast and highly scalable exchange of data across repositories storing research datasets, manuscripts and scientific software. The data exchange in the demonstrator will be based on the ResourceSync protocol.

Consortium and collaborator organisations: ISTI-CNR (IT), PIN (IT), MIBACT – ICCU (IT), Athena Research Center (GR), MPG (DE), CNRS (FR).

Scientific Contact: Petr Knoth, petr.knoth@open.ac.uk

WP5 contacts/Main Shepherd: Thomas Zastrow, thomas.zastrow@rzg.mpg.de
Deputy: Dario Vianello (EBI), dario@ebi.ac.uk

**Current Status: ONGOING**

**Design and implementation of a scalable client solution for accessing CORE metadata and full texts. Initial definition of the experimental design. Currently they are not using TextCrowd dataset yet. They need to do some preliminary tests.**

### Science Demonstrator Title: VisIVO

Goal: Data Knowledge Visual Analytics Framework for Astrophysics.
Scientific Contact: Alessandro Costa (OACT), alessandro.costa@oact.inaf.it

Consortium and collaborator organisations: INAF (IT) with international engagements in SCI-BUS, ER-flow, VIALACTEA, INDIGO DataCloud, ASTERICS, AENEAS, AARC2
WP5 contacts/Main Shepherd: Michael Schuh, michael.schuh@desy.de Deputy: Dario Vianello, EBI dario@ebi.ac.uk

**Current Status: ONGOING**

**Interested to reuse the WS-PGRADE/gUSE configuration used by the EPOS/VERCE SD. Interested to use the B2FIND solution. Set up a doodle to schedule a meeting with the provider.**

### Science Demonstrator Title: Hydrology

Goal: Switching on the EOSC for Reproducible Computational Hydrology by FAIR-ifying eWaterCycle and SWITCH-ON

Scientific Contact: Rolf Hut, r.w.hut@tudelft.nl

Consortium and collaborator organisations: Delft University of Technology (NL), Netherlands eScience Center (NL), SMHI (SE), EMBnet, SURFSara (NL), EGI, CYFRONET (PL), Bavarian Academy of Sciences (DE)

WP5 contacts/Main Shepherd: Erik van den Bergh, EBI evdbergh@ebi.ac.uk

Deputy: Dario Vianello, EBI dario@ebi.ac.uk

**Current Status: ONGOING**

**Testing pipeline. CWL will be used as interoperable language for the workflows.
OneData is explored as data repository.**

### Science Demonstrator Title: VisualMedia

Goal: a service for sharing and visualizing visual media files on the web

Consortium and collaborator organisations: ISTI-CNR (IT), PIN (IT), MIBACT – ICCU (IT), Athena Research Center (GR), MPG (DE), CNRS (FR).

Scientific Contact: Roberto Scopigno (CNR), roberto.scopigno@isti.cnr.it

WP5 contacts/Main Shepherd: Thomas Zastrow, thomas.zastrow@rzg.mpg.de
Deputy: Dario Vianello (EBI), dario@ebi.ac.uk

**Current Status: ONGOING**

**Working with CNR team in order to run the workflow under the D4Science Infrastructure. This activity involves three steps: Authentication, Storage and Scalability.**