## Status of the EOSCpilot Scientific Demonstrators

Giuseppe La Rocca, 12/06/2018

Summary

[Status of the EOSCpilot Scientific Demonstrators 1](#_Toc516492986)

[Summary report from first round of SDs (Jan 2017-Dec. 2017): 2](#_Toc516492987)

[Science Demonstrator title: DPHEP 2](#_Toc516492988)

[Science Demonstrator title: TextCrowd 2](#_Toc516492989)

[Science Demonstrator title: PhotonScience 2](#_Toc516492990)

[Science Demonstrator title: PanCancer 2](#_Toc516492991)

[Science Demonstrator title: ERFI 3](#_Toc516492992)

[Summary report from second round of SDs (July 2017-June 2018): 4](#_Toc516492993)

[Science Demonstrator title: CryoEM 4](#_Toc516492994)

[Science Demonstrator title: EPOS/VERCE 4](#_Toc516492995)

[Science Demonstrator Title: LOFAR 4](#_Toc516492996)

[Science Demonstrator title: Life Science datasets 5](#_Toc516492997)

[Science Demonstrator title: Prominence 5](#_Toc516492998)

[Summary report from third round of SDs (Dec. 2017-Nov 2018): 6](#_Toc516492999)

[Science Demonstrator Title: BioImaging 6](#_Toc516493000)

[Science Demonstrator Title: Frictionless Data Exchange 6](#_Toc516493001)

[Science Demonstrator Title: VisIVO 6](#_Toc516493002)

[Science Demonstrator Title: Hydrology 7](#_Toc516493003)

[Science Demonstrator Title: VisualMedia 7](#_Toc516493004)

## 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](mailto:Jamie.Shiers@cern.ch)

WP5 contacts/Main Shepherd: John Kennedy, [jkennedy@mpcdf.mpg.de](mailto:jkennedy@mpcdf.mpg.de)

Deputy: Matthew Viljoen, [matthew.viljoen@egi.eu](mailto: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](mailto:franco.niccolucci@gmail.com)

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

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

**Working on the preparation of user story for the SD to describe how the   
researcher can use the "solution" that you have implemented so far in the context of the project. Further extending the pilot to store datasets in B2DROP. Already had a technical meeting to discuss the additional workplan. TODO: D4Science has to publish a WebDAV endpoint**

### Science Demonstrator title: Photon&Neutron

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](mailto:Volker.Guelzow@desy.de)

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

**Current Status: Pre-final report has been submitted. Final report will be available in two weeks. Michael will discuss internally the opportunity to produce a workflow for this SD.**

### Science Demonstrator title: PanCancer

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

Consortium and collaborator organisations: EMBL, EMBL-EBI

WP5 contacts/Main Shepherd: Erik van den Bergh, [evdbergh@ebi.ac.uk](mailto:evdbergh@ebi.ac.uk)

Deputy: Gergely Sipos, [gergely.sipos@egi.eu](mailto:gergely.sipos@egi.eu)

**Current Status: Pre-final report has been submitted. Running the final benchmarks to test the Onedata integration and produce the final report.**

### 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](mailto:werner.kutsch@icos-ri.eu)

WP5 contacts/Main Shepherd: Giuseppe La Rocca, [giuseppe.larocca@egi.eu](mailto:giuseppe.larocca@egi.eu)

**Current Status: ONGOING**

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

**The partner was not able to provide sufficient staff for working on this SD.**

## 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](mailto:coss@cnb.csic.es)   
WP5 contacts/Main Shepherd: Gergely Sipos, [gergely.sipos@egi.eu](mailto:gergely.sipos@egi.eu)   
Deputy: Erik van den Bergh, [evdbergh@ebi.ac.uk](mailto:evdbergh@ebi.ac.uk)

**Current Status: ONGOING**

**Almost completed the integration of workflow export/import feature into Scipion. Requested for cloud resources (especially GPGPU resources). A technical meeting will take place in the next weeks.**

### 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](mailto:A.Rietbrock@liverpool.ac.uk)

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

**Current Status: ONGOING**

**A new DCI\_BRIDGE VM image is now available in the EGI AppDB.   
Access to the EGI FedCloud infrastructure working. Further discuss next steps**

**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.**

### 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](mailto:holties@astron.nl)

WP5 contacts/Main Shepherd: Thomas Zastrow, [thomas.zastrow@mpcdf.mpg.de](mailto:thomas.zastrow@mpcdf.mpg.de)

Deputy: John Kennedy, [jkennedy@mpcdf.mpg.de](mailto:jkennedy@mpcdf.mpg.de)

**Current Status: ONGOING**

**Activities in line with the workplan. Asked for 3 months prolongation.**

### 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](mailto:jordi.rambla@crg.eu)

WP5 contacts/Main Shepherd: Erik van den Bergh, [evdbergh@ebi.ac.uk](mailto:evdbergh@ebi.ac.uk)

Deputy: Matthew Viljoen, [matthew.viljoen@egi.eu](mailto: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.

**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. Activity is currently stalled.**

### 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](mailto:shaun.de-witt@ukaea.uk)

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

**On track. Just internal issues with firewall ports, testing MPI code in cloud.**

**Working on the preparation of user story for the SD to describe how the   
researcher can use the "solution" that you have implemented so far in the context of the project.**

## 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](mailto:heriche@embl.de)

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

WP5 contacts/Main Shepherd: Erik van den Bergh, EBI [evdbergh@ebi.ac.uk](mailto: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.**

**First computations are done. Now need CPUs with more RAM. EMBL-EBI will continue to support this SD.**

### 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](mailto:petr.knoth@open.ac.uk)

WP5 contacts/Main Shepherd: Thomas Zastrow, [thomas.zastrow@rzg.mpg.de](mailto:thomas.zastrow@rzg.mpg.de)

**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.**

**Activities in line with the workplan. Petr in contact with OpenAIRE team.**

### Science Demonstrator Title: VisIVO

Goal: Data Knowledge Visual Analytics Framework for Astrophysics.  
Scientific Contact: Alessandro Costa (OACT), [alessandro.costa@oact.inaf.it](mailto: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](mailto:michael.schuh@desy.de)

**Current Status: ONGOING**

**Interested to reuse the WS-PGRADE/gUSE configuration used by the EPOS/VERCE SD. Virtual Appliance registered in the EGI AppDB. Activities in line with the workplan. Testing simulations in progress.**

### 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](mailto: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](mailto:evdbergh@ebi.ac.uk)

**Current Status: ONGOING**

**No updates from the key contact since the last 4 months.**

### 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](mailto:roberto.scopigno@isti.cnr.it)

WP5 contacts/Main Shepherd: Thomas Zastrow, [thomas.zastrow@rzg.mpg.de](mailto:thomas.zastrow@rzg.mpg.de)

**Current Status: ONGOING**

**The SD can login via D4Science or Google authentication. Implementing sharing of D4Science data. Activities in line with the workplan.**