

Science Reproducibility and Reusability with FutureGateway and a Zenodo-like repository: the PALMS experiment

R. Bruno¹, R. Barbera^{1,2}, M. Fargetta¹, R. Rotondo¹, A. Anagnostou³, S. J. Taylor³

¹ Italian National Institute of Nuclear Physics, Division of Catania, Italy

² Department of Physics and Astronomy "E. Majorana" of the University of Catania, Italy

³ Modelling & Simulation Group, Department of Computer Science, Brunel University London, UK

Abstract

Open Science (OS) is a powerful and novel paradigm to share knowledge across multidisciplinary scientific communities with the aim to improve the quality of science [1]. One of the most important OS enablers are the FAIR principles, which involves the way to Find, Access, Interoperate and Reuse research data [2]. In most of the cases published scholarship materials are not linked with computed datasets, open source software and/or virtualized computing environments and OS currently lacks of means helping to reproduce and eventually reuse cited results exploiting public or private distributed computing infrastructures (DCIs). Moreover, from the final user point of view, the best option would be the use of graphical user interfaces (GUIs) normally hosted by a Science Gateway (SG) [3] built for a specific scientific community.

The FutureGateway Framework (FGF) [4] consists of a complete software toolkit made of different parts such as: source codes, utilities, libraries and APIs capable to comfortably build reliable Science Gateways and link them to one or more DCIs avoiding any usage complexity from the final user perspective. Moreover, since SGs based on the FGF are capable to keep track of who is accessing the DCIs, not only its usage ensures OS-compliant reproducibility and reusability but also provides a possible answer in protecting or at least simply tracking people who are accessing data and this is one of the aspects that today still makes the adoption of the OS a delicate matter.

This work presents and explains how the use of the EGI's Science Software on Demand [5] (SSOD) service, built using the FutureGateway Framework in conjunction with the INFN Open Access Repository (OAR) [6], based on Zenodo [7] software, can reproduce/reuse the outputs of the agent-based Physical Activity Lifelong Modelling & Simulations (PALMS) experiment [8].

- [1] <https://www.fosteropenscience.eu/content/what-open-science-introduction>
- [2] <https://www.force11.org/group/fairgroup/fairprinciples>
- [3] <https://www.xsede.org/ecosystem/science-gateways>
- [4] <https://futuregatewayframework.github.io>
- [5] <https://fgsg.egi.eu/regegrid/web/ssod/>
- [6] <https://www.opendatarepository.it>
- [7] <https://zenodo.org/>
- [8] <https://bit.ly/2ks0jD2>



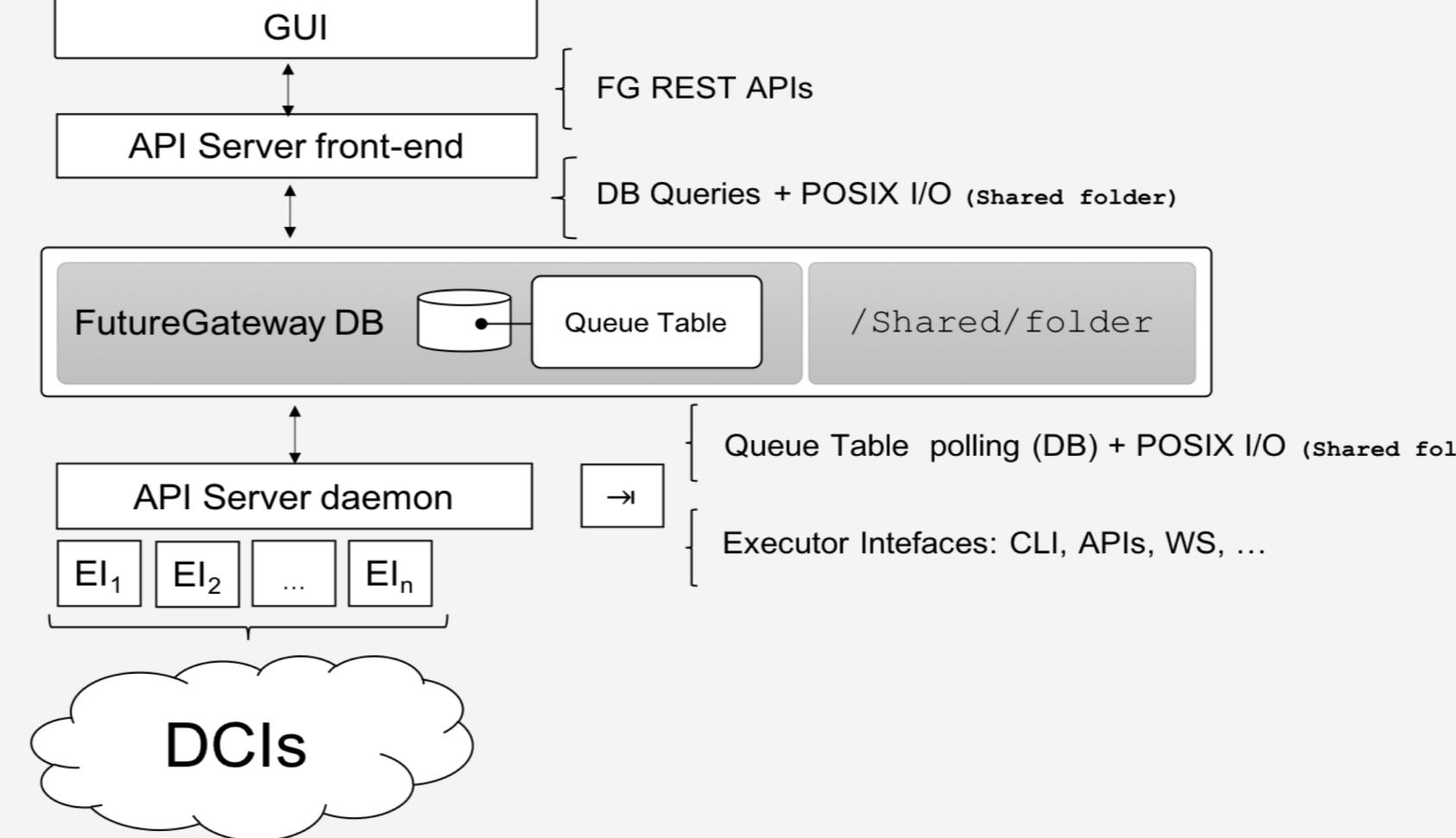
FutureGateway Framework

<https://futuregatewayframework.github.io>

The FutureGateway Framework (FGF) is a INFN software project made of services, APIs and tools, aiming to build reliable and secure Science Gateways. The FutureGateway (FG) is the core component of the framework and provides a very flexible and customisable environment to interact with any distributed computing infrastructure (DCI) just using a well defined set of REST APIs.

FG Core services consist of three components:

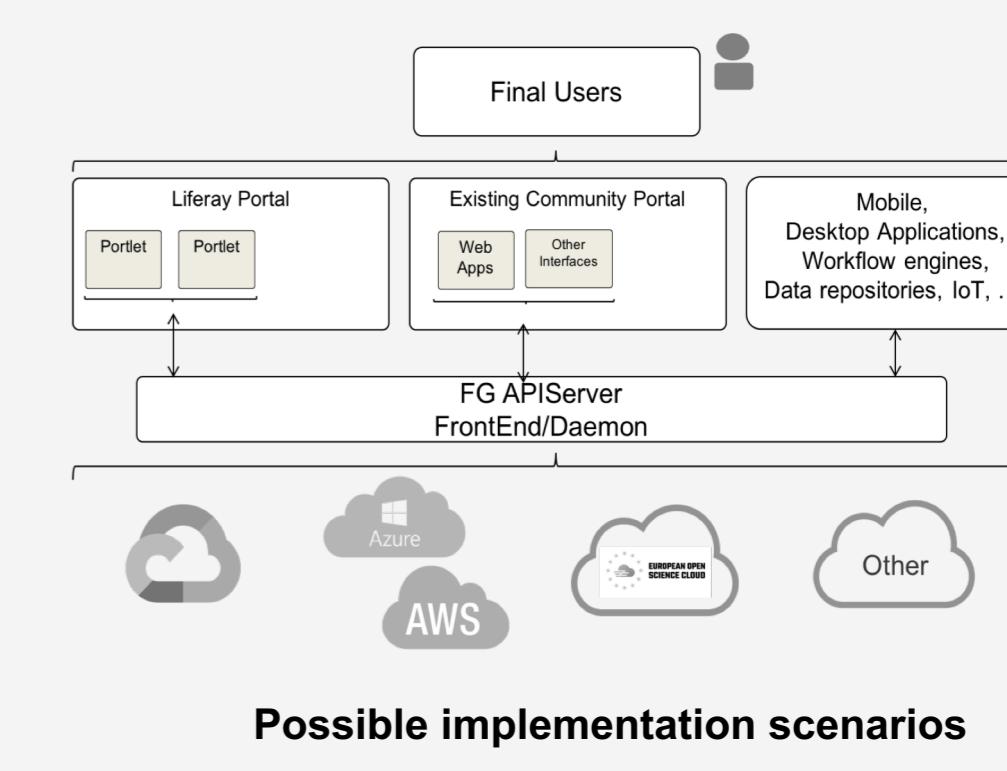
- **FutureGateway Database (fgdb)**
It stores any persistent data managed by the FutureGateway.
- **API Server front-end (fgAPIServer)**
It processes incoming REST APIs from GUIs, Workflow engines, Mobile apps, IoT services and Desktop applications.
- **API Server daemon (fgAPIServer)**
This component has the final responsibility to perform actions on top of the targeted DCI using a modular approach provided by the Executor Interfaces (EIs).



FutureGateway Architecture

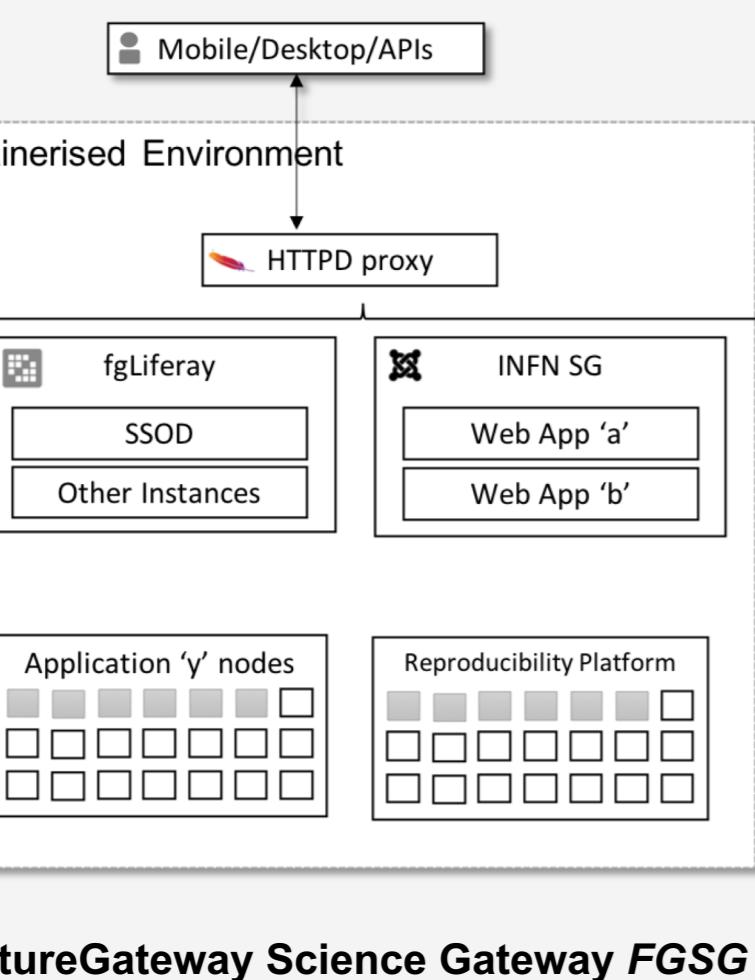
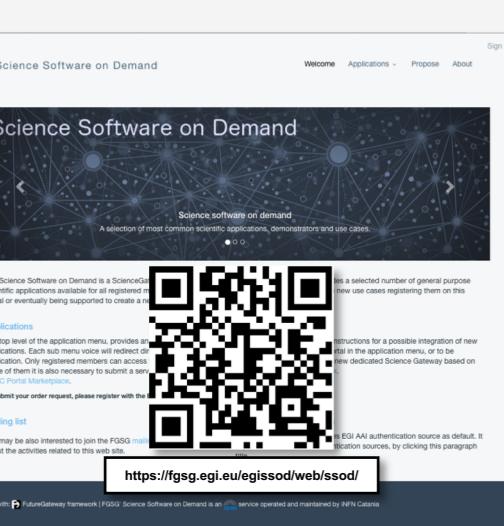
Science Gateway Instances

Starting from the FGF software repository it is possible to build one or many Science Gateway instances running on top of physical machines or virtualised environments such as: virtual machines or containers. The FGF provides installation procedures for each one of these different possibilities and it is also capable to run on top of different OSes.



Science Software on Demand

The Science Software on Demand (SSOD) service is a general purpose Science Gateway implemented using the FutureGateway Framework and running on top of a fully docker containerised environment, named: FutureGateway Science Gateway (FGSG).



Reproducibility platform

A section of the SSOD service is dedicated to the reproducibility platform, which exploits the FGSG environment to execute PALMS experiments starting from published outputs registered in the INFN Open Access repository (OAR). It is also possible to execute PALMS submitting alternative input datasets, reusing published data as well.

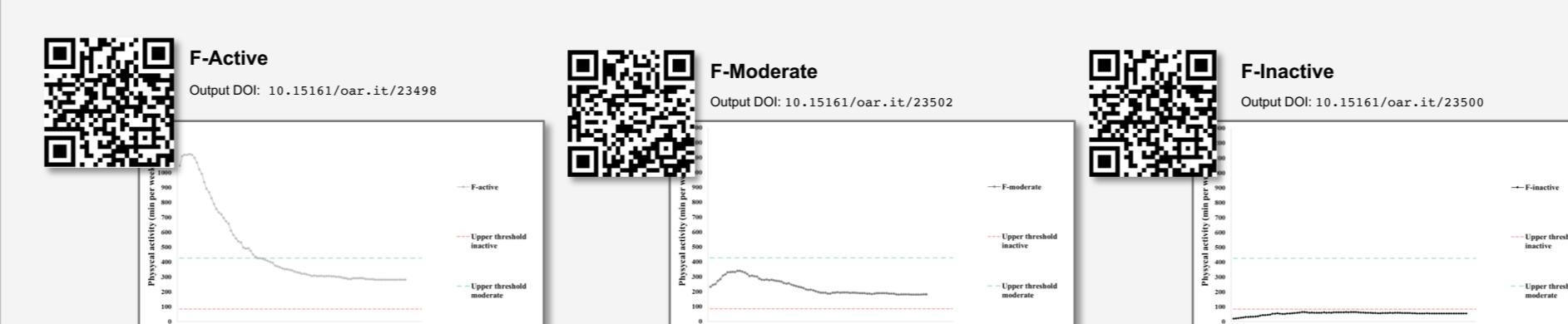
The reproducibility platform in FGSG is performed by docker compose, instantiating two containers: the first for the REPAST execution and the second to host and publish computed output files.

PALMS

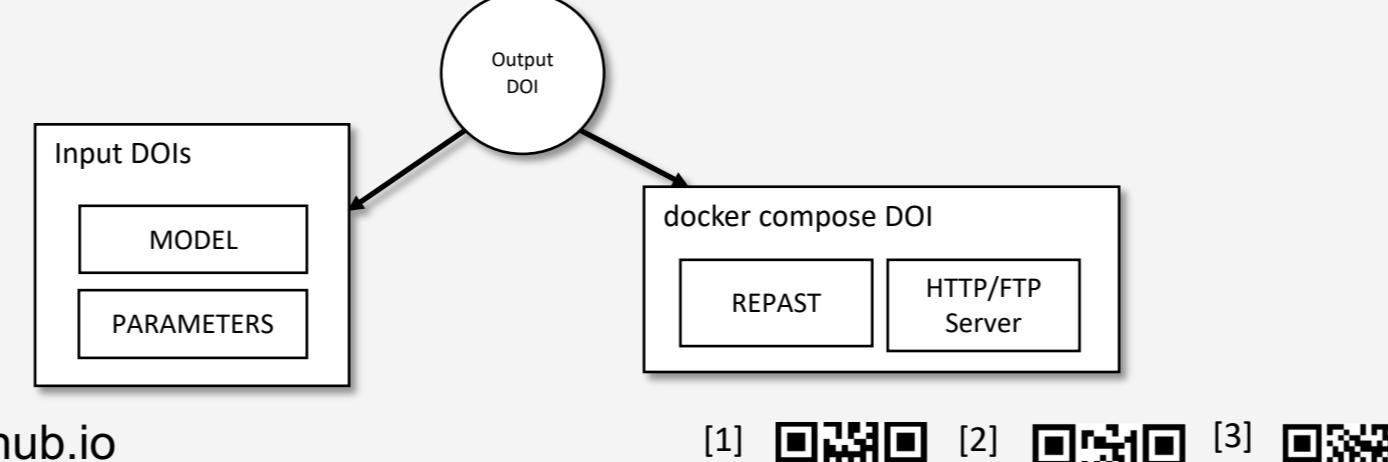
Physical Activity Lifelong Modelling & Simulations (PALMS) is an agent based micro-simulation that predicts the lifelong physical activity behaviour of a population taking into account individual characteristics and their effect on physical activity over time. The model produces individual and aggregated quantitative outputs for quality of life and health conditions related costs.

PALMS uses REPAST [1] simulation platform to produce its outputs installed inside a container image file publicly available in docker hub [2]: osabuoun/repast [3]. To execute a single simulation two input files have to be specified: the model that configures the simulation in REPAST for the PALMS specific case and the parameter file, which contains the specific input dataset needed for PALMS execution.

Output Examples



Output files are published in the INFN Open Access Repository (OAR) and they are also linked with software (docker image) and the relative input parameter files using their own DOIs as unique references.



- [1] <https://repast.github.io>
- [2] <https://hub.docker.com>
- [3] <https://hub.docker.com/r/osabuoun/repast>

INFN Open Access Repository

Research organisations are moving to new models of sharing publications and data among communities in order to overcome limitations of current publishing systems: free and open access, data and publication associations, etc. INFN and other organisations, both public and private, have signed a global initiative launched by Science Europe, named Plan S [1], aimed at moving the state funded research works in open repositories or journals available to all. The INFN Open Access Repository, uses Zenodo[2] Digital Assets Management System (DAMS), hosted by a virtualisation environment running on a Kubernetes cluster. Inside this repository, PALMS datasets, model files and the required software have been published.

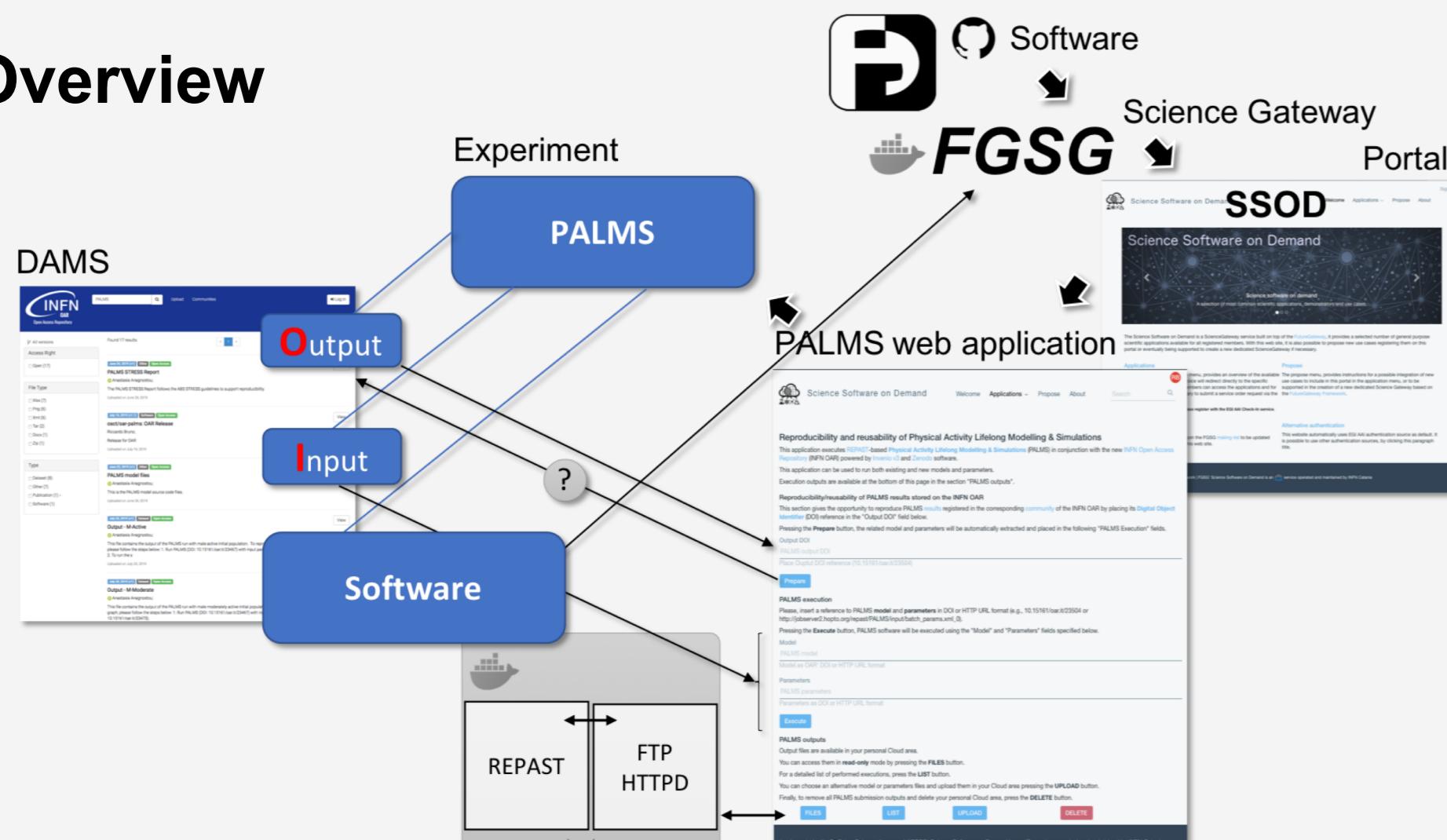
The repository is currently open for testing by all INFN staff and associated researchers and people from other organisations are free to evaluate it.

INFN OAR supports federated authentication such as, eduGAIN, GridIP and OAuth2 to support social media authentication (Google, LinkedIn, etc.).

- [1] <https://www.coalition-s.org>
- [2] <https://github.com/zenodo/zenodo>



Overview



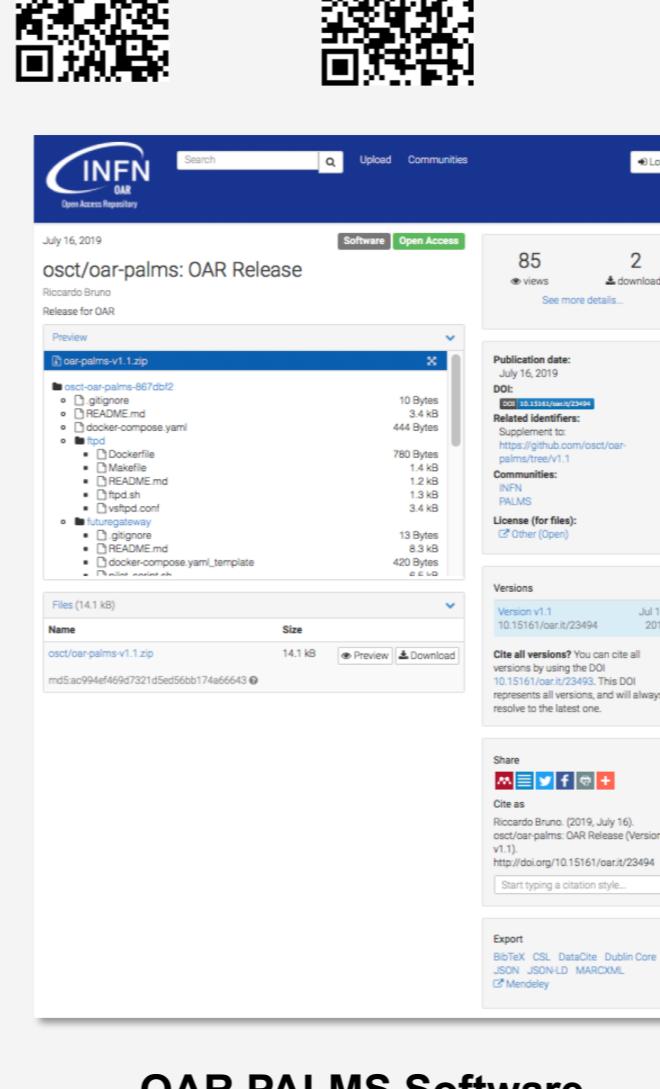
Conclusions

Due to its simplicity, the PALMS experiment has been an important use case to test the Reproducibility and Reusability platform exploiting services hosted in the FGSG environment. This model can be easily cloned in similar use cases or even expanded to even more complex usage scenarios such as those involving virtual clusters, private clouds, etc.

The FutureGateway is a mature product appreciated in those cases where desktop applications as well as mobile and web interfaces need to access DCIs. The same holds for the interaction with workflow engines and IoT. We are expecting the same success also in the context of the Open Science where Reproducibility and the Reusability matters. Investigations are in progress to extend the work done with the PALMS experiments to a general and widely usable solution.

PALMS R&R Software

The necessary software to reproduce and reuse PALMS experiment is publicly available on GitHub and it contains all the necessary files to execute PALMS simulations wherever a docker virtualisation platform is available. The software repository explains how to build docker images and run a PALMS simulation. The two containers necessary to perform PALMS simulations provide the REPAST simulation agent software and the FTP/HTTP server, respectively. The latest version of the software is published on the INFN OAR.



Projects and Institutions

- [1] INFN Institute of Nuclear Physics: <http://home.infn.it/>
- [2] INFN division of Catania: <http://www.ct.infn.it/>
- [3] Department of Physics and Astronomy "E. Majorana", Catania, IT
- [4] Brunel University London: <https://www.brunel.ac.uk/>
- [5] EGI.eu: <https://www.egi.eu/>
- [6] EOSC-hub: <https://www.eosc-hub.eu/>
- [7] INDIGO-DataCloud: <https://www.indigo-datacloud.eu/>

