Enacting Open Science by gCube

M. Assante ¹, L. Candela ¹, D. Castelli ¹, G. Coro ¹, F. Mangiacrapa ¹, P. Pagano ¹, C. Perciante ¹, ²

¹ Institute of Information Science and Technologies (ISTI) – CNR – Pisa, Italy
² Department of Information Engineering – University of Pisa – Pisa, Italy

9th International Workshop on Science Gateways
19-21 June 2017
Outline

- Open Science
- Our proposal
- Enabling components
- Conclusions
Open Science: revolusioning Science

Open Science is the movement aiming at revolusioning science and make it better.

How?
- there is no “one size fits all” definition;
- different people have different ideas;
- choices made by organisations/people define what for them it actually is.

Desired results
- better interpretation, understanding and reproducibility;
- enhanced transparency and scientific fraud detection;
- research costs reduction;
- fair scientific reward;
- better identification/assessment within the “tsunami of scientific literature”.
Open Science is the movement aiming at revolutionizing science and make it better.

How?
- there is no “one size fits all” definition;
- different people have different ideas;
- choices made by organisations/people define what for them it actually is.

Desired results
- better interpretation, understanding and reproducibility;
- enhanced transparency and scientific fraud detection;
- research costs reduction;
- fair scientific reward;
- better identification/assessment within the “tsunami of scientific literature”.
Open Science is the movement aiming at revolutionizing science and make it better.

How?
- there is no “one size fits all” definition;
- different people have different ideas;
- choices made by organisations/people define what for them it actually is.

Desired results
- better interpretation, understanding and reproducibility;
- enhanced transparency and scientific fraud detection;
- research costs reduction;
- fair scientific reward;
- better identification/assessment within the “tsunami of scientific literature”.
Our proposal: gCube-system

A software framework conceived to enable **Virtual Research Environments**, i.e. web-based working environments equipped with a set services to support research activities of communities and make transparency, openness and reproducibility the norm.

**Figure**: A single “open” research place
Social Networking Area - Overview

Purposes and Features

- area for communication among VRE’s members;
- discussions via posting;
- reply/like/mentioning/top-topics are supported.

Features for Open Science

- every item is equipped with a unique identifier;
- discussions are really transparent and open;
- actions taken by users are carefully captured and documented.
Social Networking Area - Overview

Purposes and Features
- area for communication among VRE’s members;
- discussions via posting;
- reply/like/mentioning/top-topics are supported.

Features for Open Science
- every item is equipped with a unique identifier;
- discussions are really transparent and open;
- actions taken by users are carefully captured and documented.
Architecture

- Cassandra and ElasticSearch for storing/searching data;
- portlets for writing posts, retrieving them, adding comments, liking and top topics management;
- RESTful APIs for programmatic/external access;

Figure: Social - Overview
Purposes and Features

- organise user’s material;
- access to shared material;
- open-ended set of items equipped with extensible metadata.

Features for Open Science

- every item is equipped with a unique identifier;
- versioning is supported;
- extensible metadata via key/pair mechanism;
- tightly integrated with Social Area and Publishing Area.
Purpose and Features
- organise user's material;
- access to shared material;
- open-ended set of items equipped with extensible metadata.

Features for Open Science
- every item is equipped with a unique identifier;
- versioning is supported;
- extensible metadata via key/pair mechanism;
- tightly integrated with Social Area and Publishing Area.

Figure: Workspace - Overview
Architecture

- Apache Jackrabbit for storing nodes hierarchy (metadata);
- Item payload is stored on different storage solutions (MongoDB, GeoServer/THREDDS, RDB, ...);
- Access via GUI (Portlet, Widget) and RESTful APIs;

Figure: Workspace Architecture
Data Analytics Platform

Purposes and Features

- collection of ready-to-use algorithms;
- requests dynamically distributed;
- compliant with standard web-based protocol;
- extensible via Algorithm Publisher/Importer services;

Features for Open Science

- processes with unique identifiers;
- processes described/exposed by OGC standard;
- algorithms written in different programming languages;
- detailed record provenance (PROV-O);
- integration with the Workspace Area.

Figure: Analytics - Overview
Purposes and Features

- collection of ready-to-use algorithms;
- requests dynamically distributed;
- compliant with standard web-based protocol;
- extensible via Algorithm Publisher/Importer services;

Features for Open Science

- processes with unique identifiers;
- processes described/exposed by OGC standard;
- algorithms written in different programming languages;
- detailed record provenance (PROV-O);
- integration with the Workspace Area.

Figure: Analytics - Overview
Data Analytics Platform

Architecture

- DataMiner (DM) portlet for executing processes;
- DM Master accepts requests for executing processes;
- DM Worker executes processes assigned by a DM Master;
- open set of algorithms hosted by the DM Algorithms Repository;
- injection of new algorithms is allowed through Algorithm Importer portlet and the Algorithm Publisher service.

Figure: Analytics - Architecture
Purposes and Features

- catalogue of artefacts with search/browse functionalities;
- openness with respect to the typologies of products published;
- item has type, open ended set of metadata and optional resource(s).

Features for Open Science

- item has a persistent, unique identifier;
- payload(s) stored in persistent storage area;
- items equipped with license;
- inform new item availability via social;
- customisation of typologies of products and metadata.
Data Publishing Platform - Overview

Purposes and Features
- catalogue of artefacts with search/browse functionalities;
- openness with respect to the typologies of products published;
- item has type, open ended set of metadata and optional resource(s).

Features for Open Science
- item has a persistent, unique identifier;
- payload(s) stored in persistent storage area;
- items equipped with license;
- inform new item availability via social;
- customisation of typologies of products and metadata.

Figure: Catalogue - Overview
Architecture

- CKAN technology as core service;
- catalogue Service realises the business logic;
- catalogue Portlet for navigation and Widget for publication;
- catalogue RESTful APIs for external access and publication;
- payloads are stored in the Workspace/Storage Hub area.

Figure: Catalogue - Architecture
Open Science in action with gCube

Figure: Boosting Science with gCube collaborative environment
Some of Open Science’s desired results can be achieved using gCube VREs and services

- FAIR principles compliant;
- Open, extensible and customisable;
- Offered as-a-service.

Communities of different research areas and different projects successfully use VREs for their activities.
More Information

- About gCube-system: www.gcube-system.org
- About D4Science Infrastructure: www.d4science.org
Contact Information

Costantino Perciante
National Research Council of Italy (CNR-ISTI)
costantino.perciante@{isti.cnr.it, ing.unipi.it}
Questions
Extras (I) - Social Area Overview

Figure: Overview
Extras (II) - Social Area Architecture

Figure: Architecture
Extras (III) - Catalogue Area Overview

Figure: Overview
Extras (IV) - Catalogue Area Architecture

Catalogue Portlet
Publishing Widget
Catalogue RESTful API
Catalogue Service
Catalogue Item Types
Storage Hub
Workspace

Figure: Architecture
Extras (V) - Analytics Area Overview

Figure: Overview
Extras (VI) - Analytics Area Architecture

Figure: Architecture
Extras (VII) - Workspace Area Overview

**Figure**: Overview

- **Workspace Portlet**
- **Workspace Widget**
- **Workspace RESTful API**
- **Workspace Service**
- **Storage Hub**
- **mongoDB**
- **plugins**

Costantino Perciante, ISTI-CNR

Enacting Open Science by gCube
Extras (VIII) - Workspace Area Architecture

Figure: Architecture

Workspace Portlet  Workspace Widget  Workspace RESTful API

Workspace Service

Storage Hub  plugins

mongoDB  Apache Jackrabbit

Costantino Perciante, ISTI-CNR  Enacting Open Science by gCube