

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

- Open Science
- Our proposal
- Enabling components
- Conclusions

Open Science: revolutioning Science

Open Science is the movement aiming at revolutioning 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: revolutioning Science

Open Science is the movement aiming at revolutioning 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: revolutioning Science

Open Science is the movement aiming at revolutioning 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.

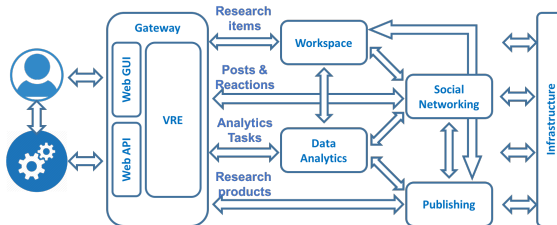


Figure : A single "open" research place

Social Networking Area - Overview

Purposes and Features

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

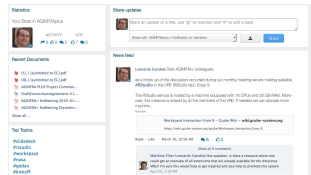


Figure : Social - Overview

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.

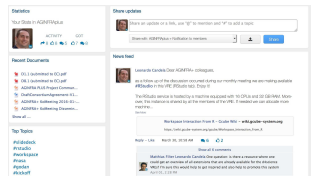


Figure : Social - Overview

Social Networking Area - Architecture

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;

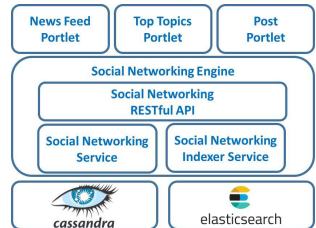
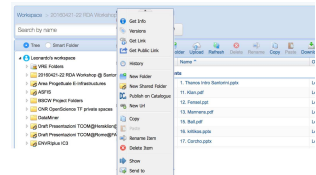


Figure : Social - Overview

Workspace Platform - 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.

Figure : Workspace - Overview

Workspace Platform - 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.

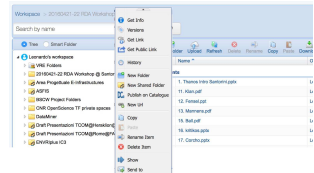


Figure : Workspace - Overview

Workspace Platform - Architecture

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;

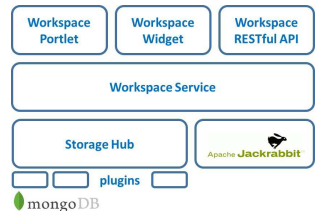


Figure : Workspace Architecture

Data Analytics Platform

Purposes and Features

- collection of ready-to-use algorithms;
- requests dinamically 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.

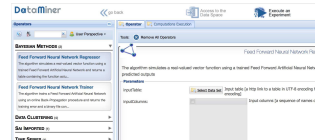


Figure : Analytics - Overview

Data Analytics Platform

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.

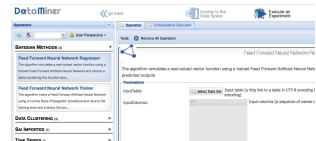


Figure : Analytics - Overview

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.

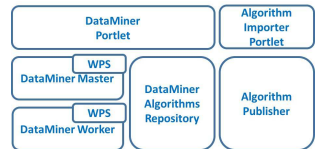


Figure : Analytics - Architecture

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.

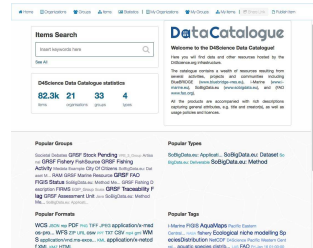


Figure : Catalogue - Overview

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.

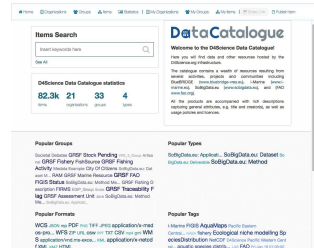


Figure : Catalogue - Overview

Data Publishing Platform - Architecture

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

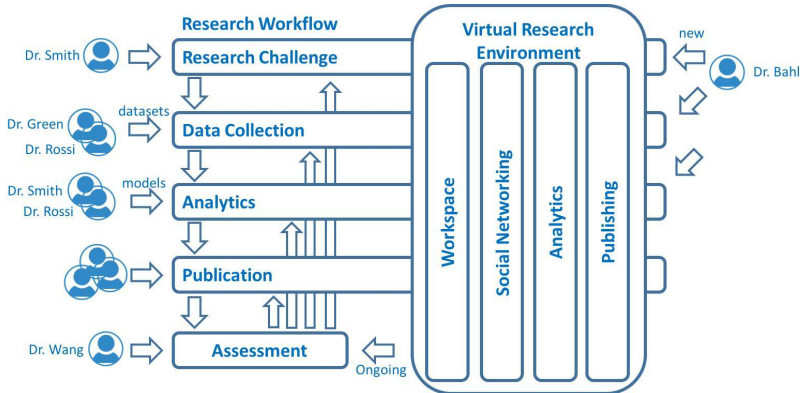
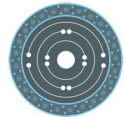


Figure : Boosting Science with gCube collaborative environment

Conclusions

- 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

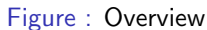
Costantino Perciante

National Research Council of Italy (CNR-ISTI)

costantino.perciante@{isti.cnr.it, ing.unipi.it}

Questions





Extras (II) - Social Area Architecture

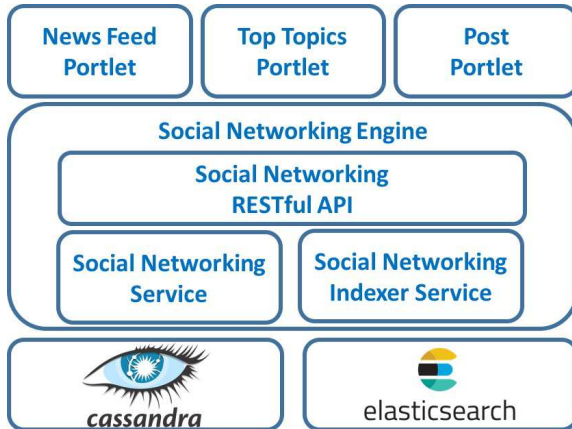


Figure : Architecture

Extras (III) - Catalogue Area Overview

The screenshot displays the D4Science Data Catalogue web interface. At the top, a navigation bar includes links for Home, Organizations, Groups, Items, Statistics, My Organizations, My Groups, My Items, Share Link, and Publish Item. The main content area is divided into several sections:

- Items Search:** A search bar with the placeholder text "Insert keywords here" and a magnifying glass icon. Below it is a "See All" link.
- D4Science Data Catalogue statistics:** A section showing four statistics: 82.3k items, 21 organisations, 33 groups, and 4 types.
- Welcome to the D4Science Data Catalogue!** A welcome message stating that users will find data and other resources hosted by the D4Science.org infrastructure. It lists several activities, projects, and communities including BlueBRIDGE, i-Marine, SoBigData.eu, and FAO. It also mentions that all products are accompanied with rich descriptions capturing general attributes, e.g. title and creator(s), as well as usage policies and licences.
- Popular Groups:** A list of popular groups including Societal Debates, GRSF Stock Pending, VRE_3_Group, Artisanal, GRSF Fishery FishSource, GRSF Fishing Activity Metadata Example City Of Citizens, SoBigData.eu: Dataset M..., RAM GRSF Marine Resource, GRSF FAO, FIGIS Status, SoBigData.eu: Method Me..., GRSF Fishing Description FIRMS EGP_Group Scala, GRSF Traceability F lag GRSF Assessment Unit Java, SoBigData.eu: Method Me..., SoBigData.eu: Appraisal...
- Popular Formats:** A list of popular formats including WCS, JSON, rep, PDF, PNG, TIFF, JPEG, application/x-msd, os-pro..., WFS, ZIP, URL, csw, PPT, TXT, CSV, mp4, gml, WM, S application/vnd.ms-excel..., XML application/x-netcdf, f XML, km2, HTML.
- Popular Types:** A list of popular types including SoBigData.eu: Application, SoBigData.eu: Dataset, SoBigData.eu: Deliverable, and SoBigData.eu: Method.
- Popular Tags:** A list of popular tags including i-Marine, FIGIS, AquaMaps, Pacific Eastern Central..., NASA fishery Ecological niche modelling Species Distribution NetCDF, D4Science Pacific Western Central..., aquatic species distribution, LRS, FAO, Jan 16 01:00:00.

Figure : Overview

Extras (IV) - Catalogue Area Architecture



Figure : Architecture

Extras (V) - Analytics Area Overview

The screenshot displays the Dataminer web interface. At the top, there is a navigation bar with the 'Dataminer' logo, a 'go back' button, and icons for 'Access to the Data Space' and 'Execute an Experiment'. Below this, a sidebar on the left lists various operators under categories like 'BAYESIAN METHODS (2)', 'DATA CLUSTERING (4)', 'SAI IMPORTED (1)', and 'TIME SERIES (4)'. The 'Feed Forward Neural Network Regressor' operator is selected. The main panel shows the operator's configuration, including a description: 'The algorithm simulates a real-valued vector function using a trained Feed Forward Artificial Neural Network predicted outputs'. Under the 'Parameters' section, there are fields for 'inputTable:' (with a 'Select Data Set' button and a description: 'Input table [a http link to a table in UTF-8 encoding / encoding]') and 'inputColumns:' (with a description: 'Input columns [a sequence of names c]').

Figure : Overview

Extras (VI) - Analytics Area Architecture

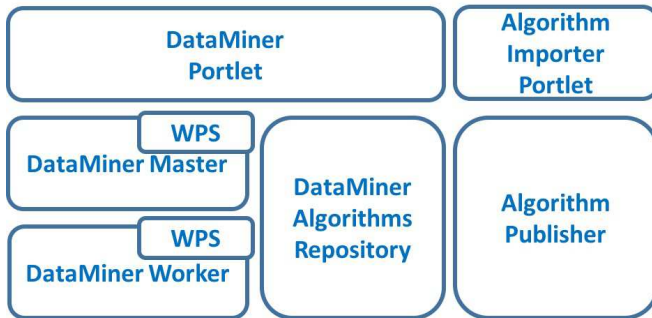


Figure : Architecture

Extras (VII) - Workspace Area Overview

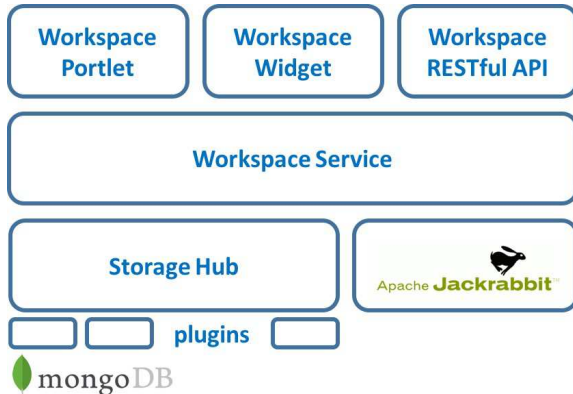


Figure : Overview

Extras (VIII) - Workspace Area Architecture

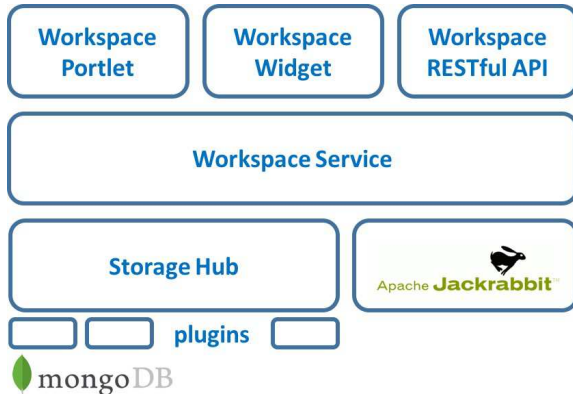


Figure : Architecture