

# The D4Science Infrastructure and its Open Science Way

Leonardo Candela

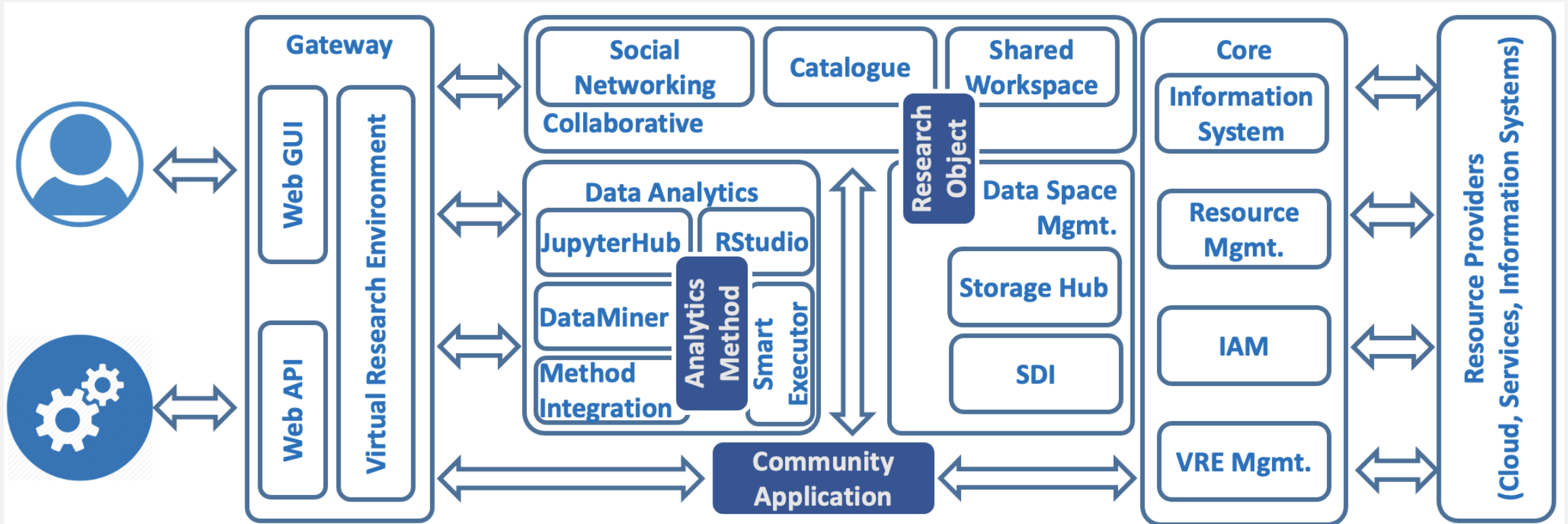
Consiglio Nazionale delle Ricerche - Istituto di Scienza e Tecnologie dell'Informazione

 <https://orcid.org/0000-0002-7279-2727>

# ID Card

- An infrastructure built by using the **system-of-systems** approach to enact the **co-creation of Virtual Research Environments** offering a **rich array of services** and promoting **collaborative open science** practices
- Developed since 2005
- Supported by **several projects** and **communities of practices**

# Architecture



# Communication

- Social Networking platform
  - Post and replies
  - Mentions & hashtags
  - Members and groups
  - User and Application posts
  - Custom notifications

The screenshot displays a social networking interface for a project. At the top, there is a navigation bar with links for 'Blue-Cloud 2026 Project Home', 'VLab Catalogue', 'Administration', and 'Members'. The main content area features a post by Rita Giuffrida from the University of Genova, dated October 04, 10:40 AM. The post text reads: 'Dear partners, we would like to share with you that the deliverable "D3.1 - First release of aggregated and harmonised EOVS datasets" has been successfully submitted to the EC portal on 30 September. This deliverable, led by IFREMER and the WBs teams, presents the type of data used by the various WBs and the work conducted so far. You can access the deliverable on both the Blue-Cloud's catalogue and Zenodo: <https://doi.org/10.5281/zenodo.13889325>'. Below the post, there are three comments: one from Simona Simoncelli thanking Rita for her reaction, one from Julie Gatti requesting to add her name, and a comment from Rita Giuffrida to Simona about changes in the catalogue. A sidebar on the right shows 'Trending Topics' including #eosc, #vre, #data\_management, #duplicates, #workbench, #eov, #pluto.jl, #wp3, #ebv, and #quality\_control. At the bottom, there is a 'Catalogue' entry for DENISE, dated October 02, 11:25 AM, with a description: 'just published the item "ThinkEngine" Please find it at <https://data.d4science.org/ctig/ResourceCatalogue/thinkengine> #answer\_set\_programming #games #knowledge\_representation #simulations'.

# Data Management and Sharing

- Shared workspace ...
  - VRE Folders
  - Private area with fine-grained sharing options
  - Per VRE view
  - History and Versioning
- ... integrated with applications

The screenshot displays the Blue-Cloud2026 workspace interface. The main window shows a file list with columns for Name, Owner, Type, Last Update, and Size. The file list includes folders like 'BSCW Project Folders', 'trash and old', 'H2020', 'iMarine Folders', 'InfraScienceGroup Folders', 'private', and 'Open Access @ IISTI', as well as files like 'IPCEI\_Bologna\_2016.04.07\_Leo\_v3.pptx' and '2017.01.15 CNR Intro@AGINFRAplus.pptx'.

Two windows are overlaid on the main interface:

- Accounting history of: stat\_algo.project**: This window shows a table of operations. The operations are categorized into 'Read (2 Items)', 'Shared (2 Items)', 'Unshared (1 Item)', and 'Updated (17 Items)'. The 'Updated' operations show that 'stat\_algo.project' was updated by Baptiste Grenier.
- Versions of: Detect\_outlayer\_from\_dirty\_dataset.ipynb**: This window shows a table of file versions. The table has columns for Version Id, Created, and Current Version. The versions range from 1.141 to 1.147, with version 1.147 being the current version.

Version Id	Created	Current Version
1.141	Thu Oct 24 10:45:45 GMT+200 2024	No
1.142	Mon Oct 28 10:01:01 GMT+100 2024	No
1.143	Mon Oct 28 10:50:12 GMT+100 2024	No
1.144	Thu Oct 31 10:21:03 GMT+100 2024	No
1.145	Thu Oct 31 11:57:52 GMT+100 2024	No
1.146	Thu Oct 31 11:59:54 GMT+100 2024	No
1.147	Thu Oct 31 12:01:25 GMT+100 2024	Yes

# Data Analytics

- Many platforms ...
  - DataMiner & CCP
    - Executable methods for humans and bots (OGC WPS & Processes)
    - Reproducibility-ready
  - Jupyter Lab & RStudio & Galaxy
- ... common patterns
  - Community-driven customization
  - Workspace for data, code, «computations» and results
  - Catalogue for publishing

The image displays three overlapping screenshots of data analytics platforms:

- Top Screenshot (Analytics Engine (CCP)):** Shows a web interface for running methods. The main content area displays 'SimpleImageClassifier' with a description: 'A simple image classifier with parametrizable url to input picture compatible with the D4Science infrastructure'. It includes an 'Inputs' field and a 'Runtime' section. A sidebar on the left shows a 'Methods List' with categories like 'Uncategorised' and 'Image Classifier'.
- Middle Screenshot (RStudio):** Shows the RStudio web interface. The main window displays the R console with the text 'R version 4.2.3 (2023-01-15) -- "Shortstork-Bronie"'. The top menu bar includes 'File', 'Edit', 'Code', 'View', 'Plots', 'Session', 'Build', 'Debug', 'Profile', 'Tools', and 'Help'.
- Bottom Screenshot (Galaxy):** Shows the Galaxy web interface. The main content area displays 'WEKEO - Harmonised Data Access API' with the text: 'The Harmonised Data Access API (HDA API) allows uniform access to the whole WEKEO catalogue, including subsetting and downloading functionalities.' The interface includes a 'Tools' sidebar on the left and a 'History' panel on the right.

# Publishing by a Catalogue

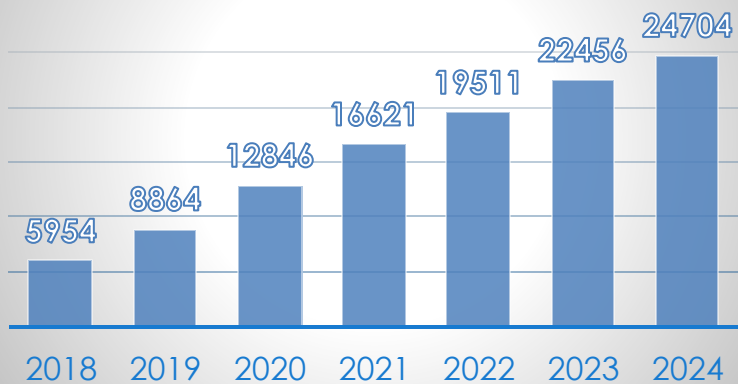
- Publishing almost **any research products** ...
  - Community-defined type-specific metadata
  - Open-ended
  - Custom workflow
- ... by making it **FAIR**
  - PURL per item
  - Rich metadata
  - Human and machine-oriented exploitation (gCat REST API, DCAT, OAI-PMH)

The image shows a digital catalogue interface. On the left, there are filters for 'Organisations' and 'Types'. The main area displays search results for '682 items found', ordered by 'Relevance'. Two items are visible: 'NetMe' (Application) and 'Annotazione semantica di delibere comunali' (Experiment). A 'Publish Item' modal window is overlaid on the right, showing a three-step process: 1. Edit Common Metadata, 2. Edit Item Specific Metadata & Publish, and 3. Add Resources. Below the modal, there are two social media-style posts from 'Catalogue' dated October 07, 12:10 PM and October 02, 11:25 AM, each mentioning a newly published item with a URL and hashtags.



# Key Performance Indicators

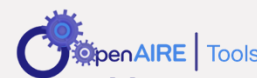
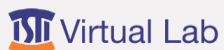
## Users



24 Gateways



217 Active VREs





# Lesson learned

- VRE-as-a-Service is key for many communities of practice
- Co-creation and openness are key for community uptake
- There is «no one-size fits all» solution for Open Science

# THANKS FOR YOUR ATTENTION



[www.d4science.org](http://www.d4science.org)

L. Candela, D. Castelli and P. Pagano. **The D4Science Experience on Virtual Research Environment Development**. *Computing in Science & Engineering*, vol. 25, no. 2, pp. 12-19, March-April 2023, doi:[10.1109/MCSE.2023.3290433](https://doi.org/10.1109/MCSE.2023.3290433)

M. Assante, L. Candela, D. Castelli, et al. **Virtual research environments co-creation: The D4Science experience**. *Concurrency Computat Pract Exper.* 2023; 35(18):e6925. doi:[10.1002/cpe.6925](https://doi.org/10.1002/cpe.6925)

M. Assante, L. Candela, D. Castelli et al. **Enacting open science by D4Science**. *Future Generation Computer Systems*, Vol 101, 2019, pp. 555-563, doi:[10.1016/j.future.2019.05.063](https://doi.org/10.1016/j.future.2019.05.063)