UNIVERSITÀ degli STUDI di CATANIA

DIPARTIMENTO di FISICA e ASTRONOMIA "Ettore Majorana"

Open Science in action

Roberto Barbera (roberto.barbera@ct.infn.it)
Dipartimento di Fisica e Astronomia «E. Majorana»
Università degli Studi di Catania

Workshop Bibliosan 2019 - Roma, 3 dicembre 2019

Concetti chiave della presentazione

Rivoluzione

Cambiamento

Mutazione

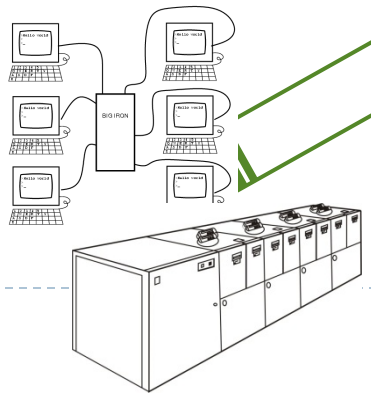
Evoluzione

(E/Ri)voluzione del calcolo scientifico

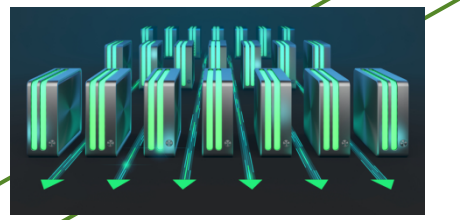
Cost of hw
Cost of networks

Time

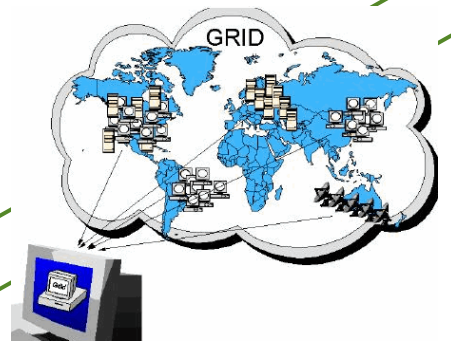
Mainframe Computing

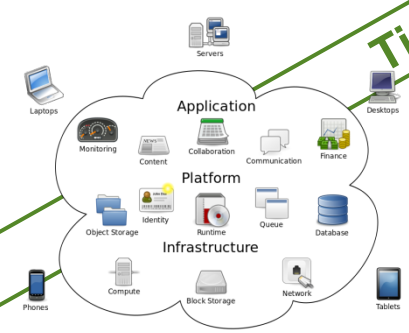


80's-90's Cluster Computing



90's-00's Grid Computing



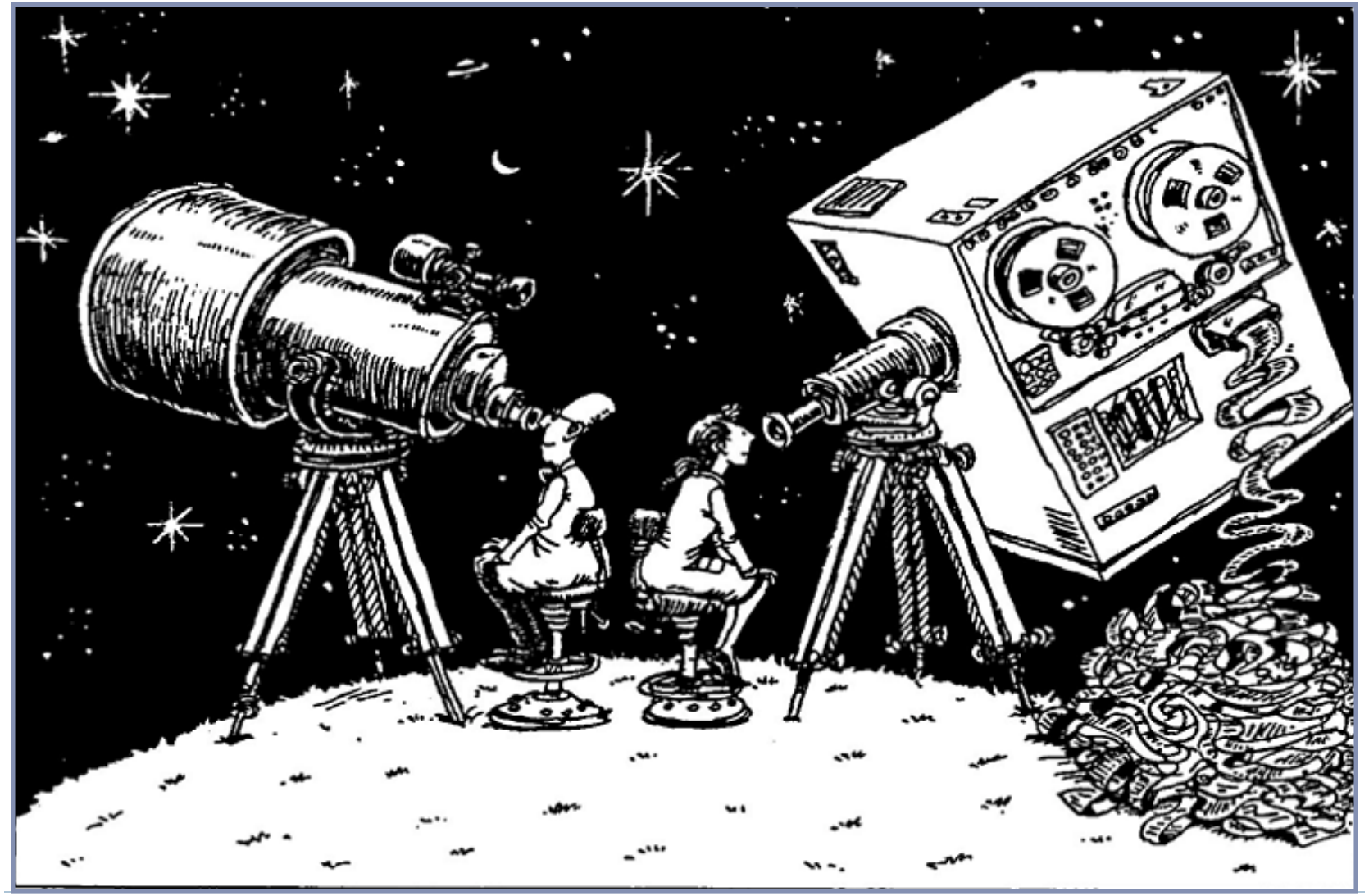


Cloud Computing

00's-10's Cloud Computing

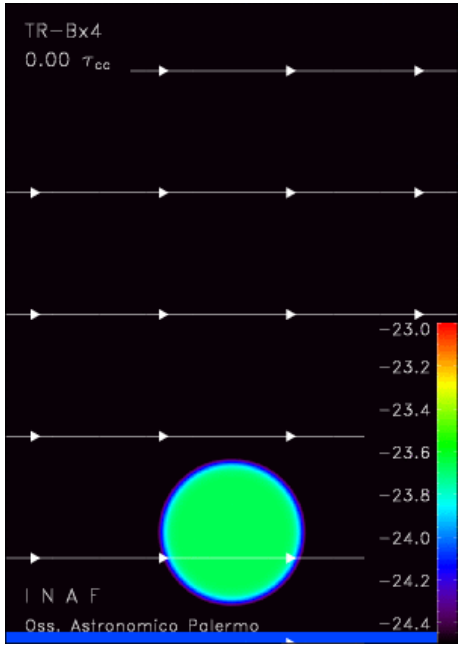
Power of COTS
WAN bandwidth

Cambiamento: ricerca «computationally intensive»

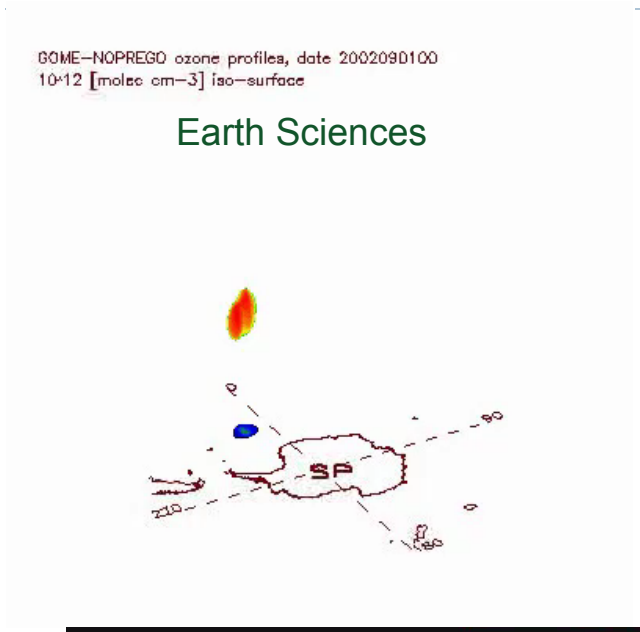


Cambiamento: ricerca «computationally intensive»

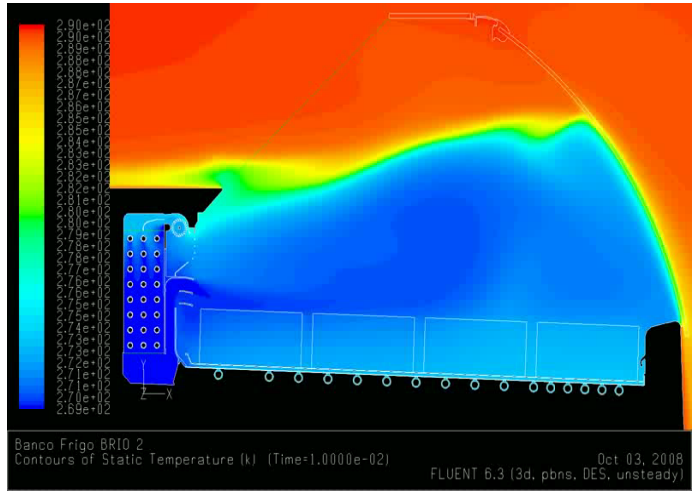
Astrophysics



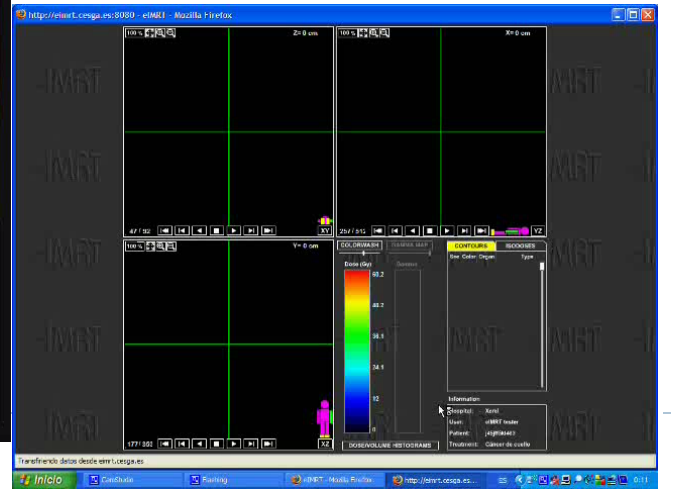
Earth Sciences

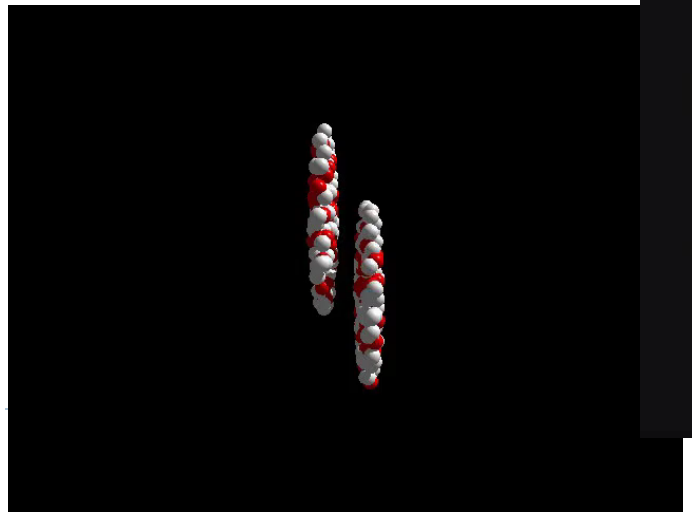


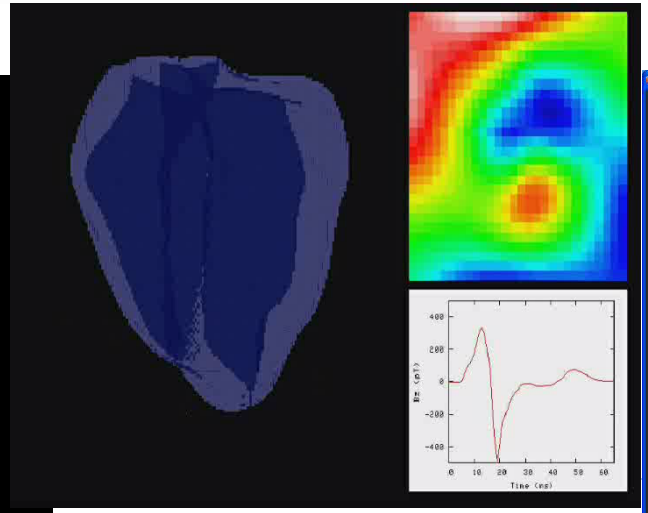
Engineering - CFD



Life Sciences

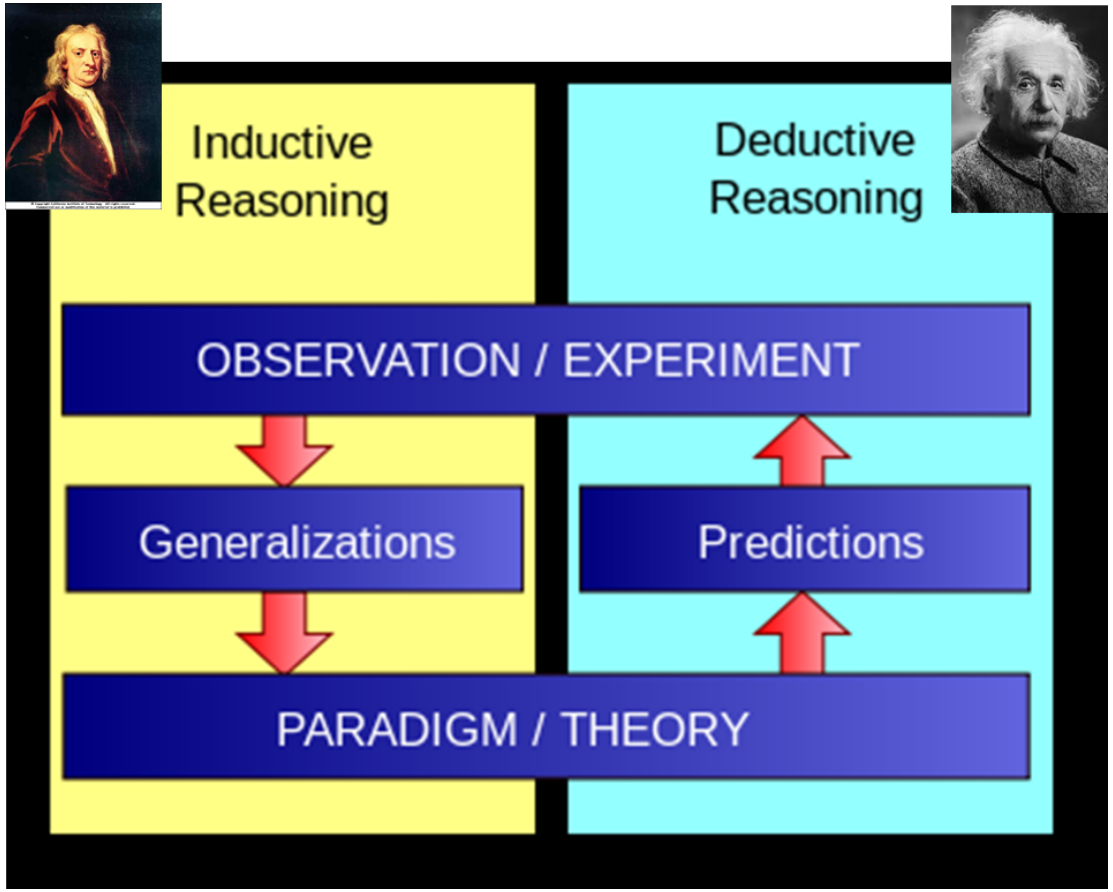






High Energy Physics

Cambiamento: ricerca «computationally intensive»



- **Examples of IR:**
  - Classical Mechanics
  - Newton's Gravitation Theory
- **Examples of DR:**
  - General Relativity
  - Standard Model of Particle Physics

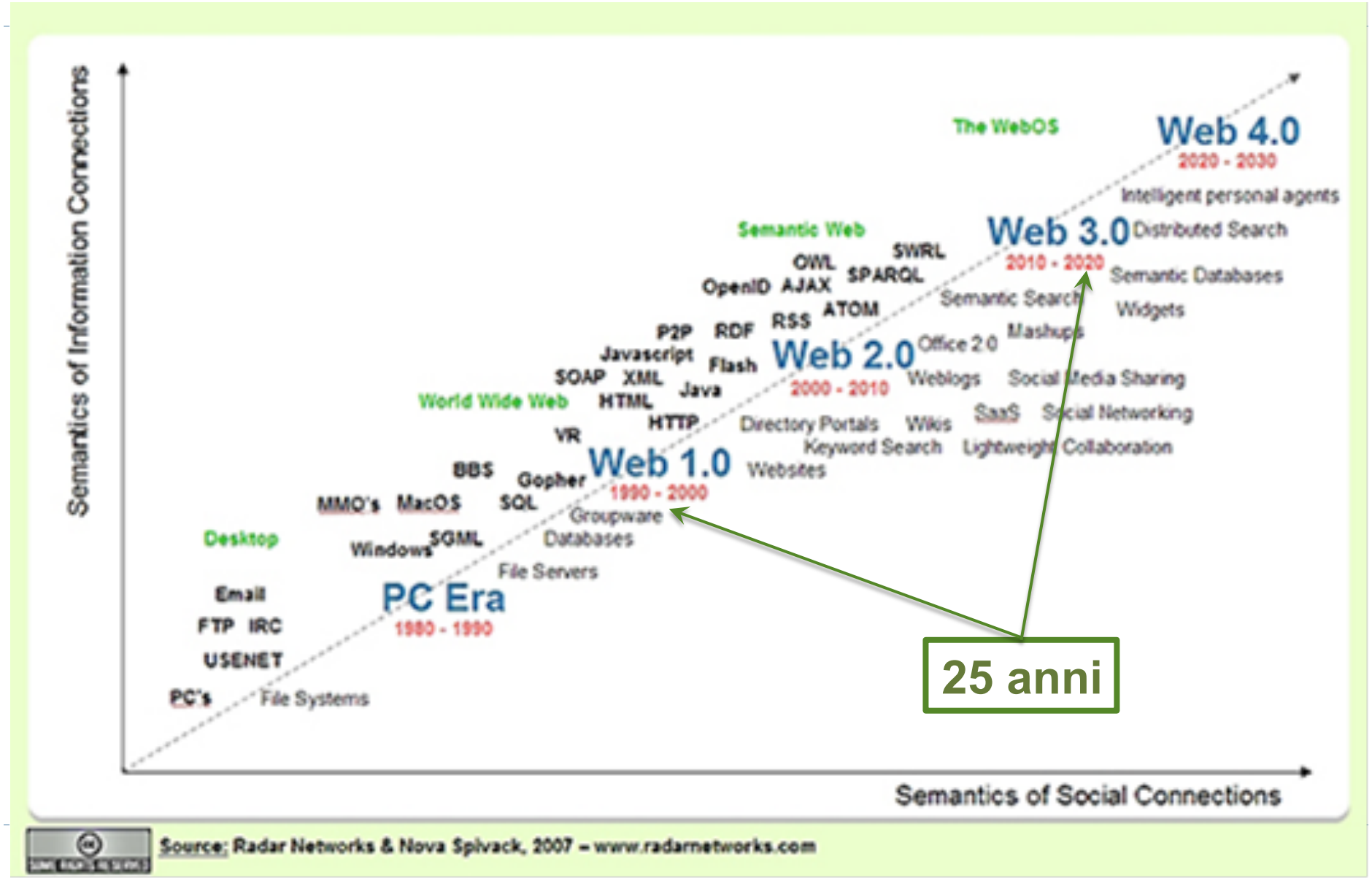
Computer simulations reconcile the inductive and deductive reasonings of the Scientific Method

https://plato.stanford.edu/entries/scientific-method/

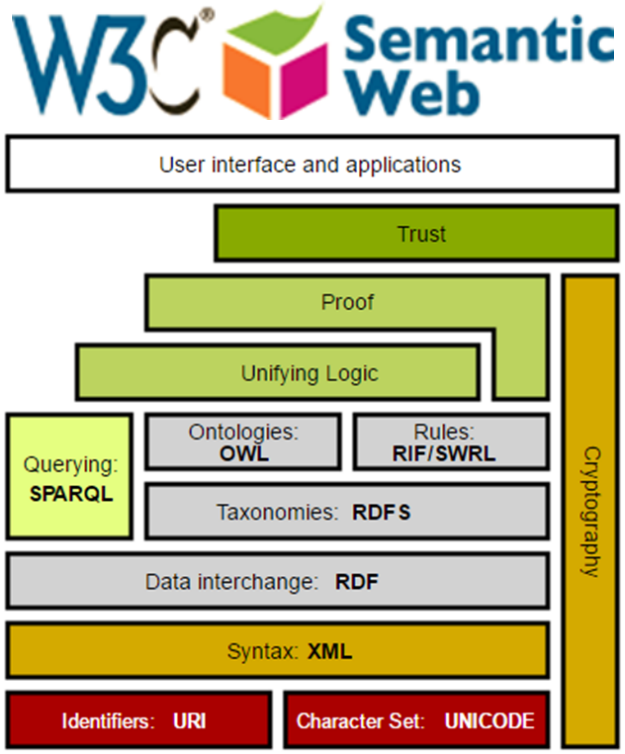
Evoluzione: il concetto di «e-Science»

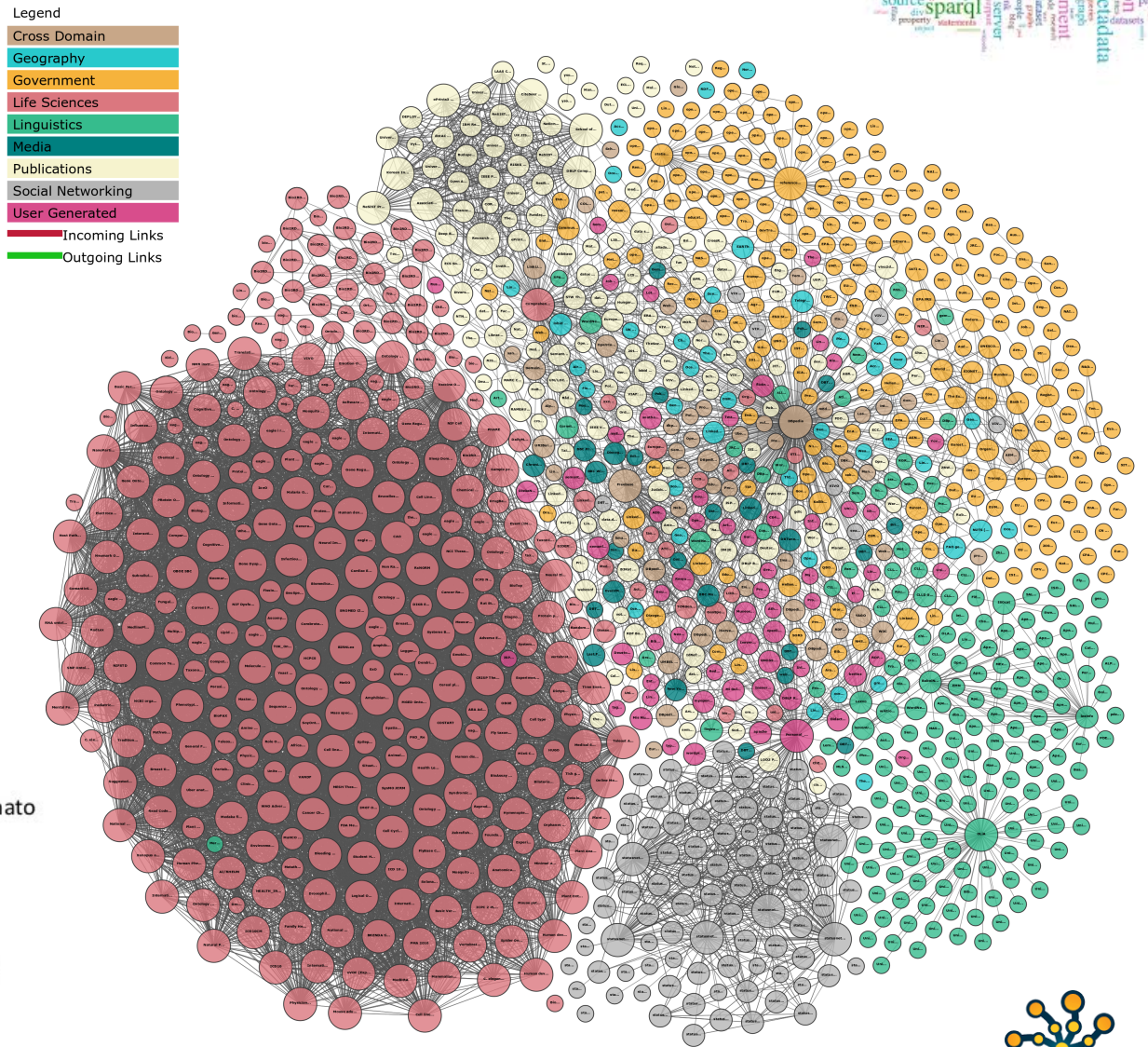
- Virtual Research Communities
- e-Infrastructure
- Applications
- Data
- Instruments/sensors

(Ri/E)voluzione del web



La «rivoluzione» del Linked Open Data (LOD)





★ OL — Dato disponibile sul Web in un qualsiasi formato (anche PDF) rilasciato con licenza Open

★★ OL RE — Leggibile dal calcolatore. Dati strutturati in formati proprietari (es. Excel)

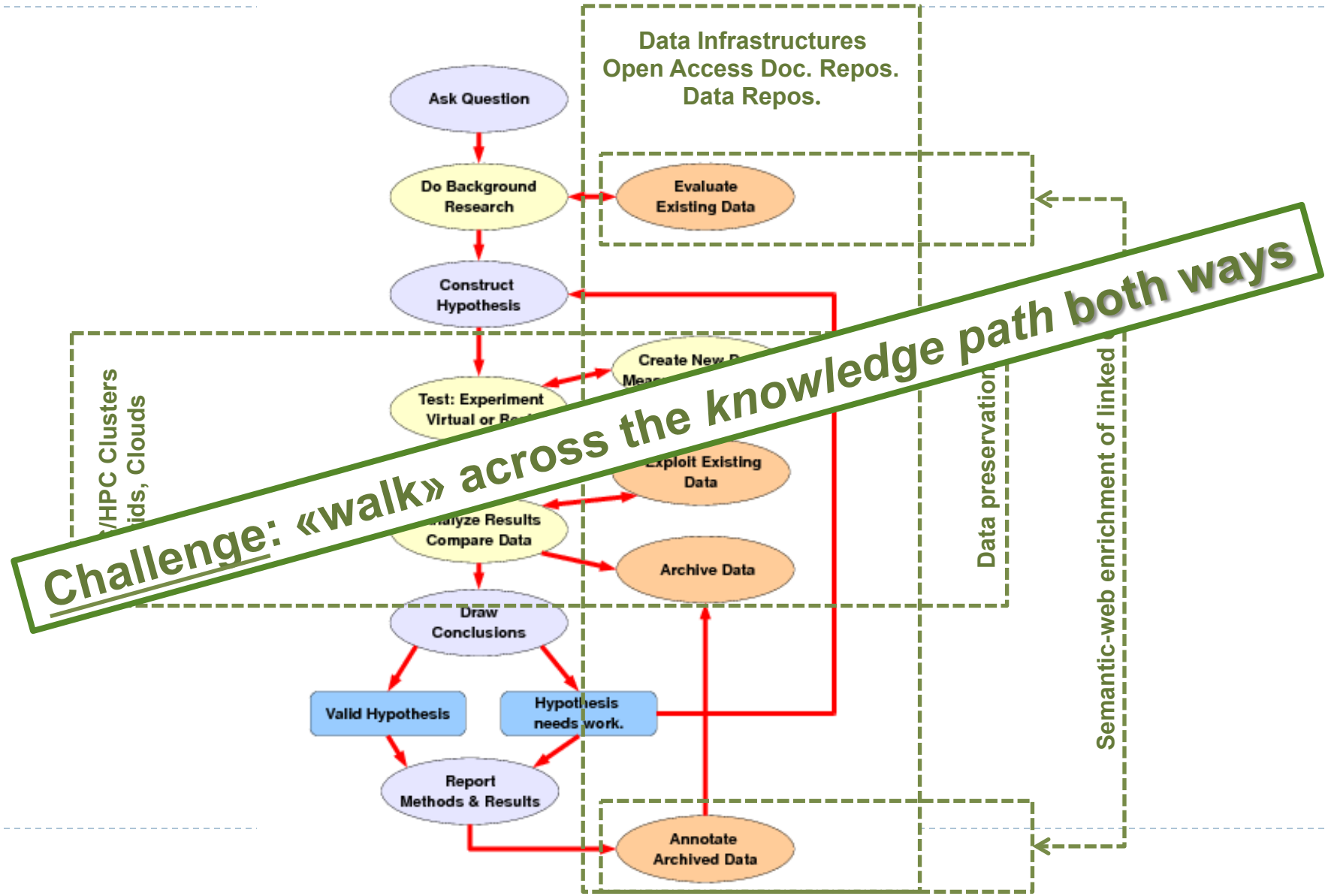
★★★ OL RE OF — Come gli Open Data a 2 stelle ma in formato non proprietario (es. XML)

★★★★ OL RE OF URI — Come I livelli precedenti ma segue gli standard W3C (RDF e Sparql)

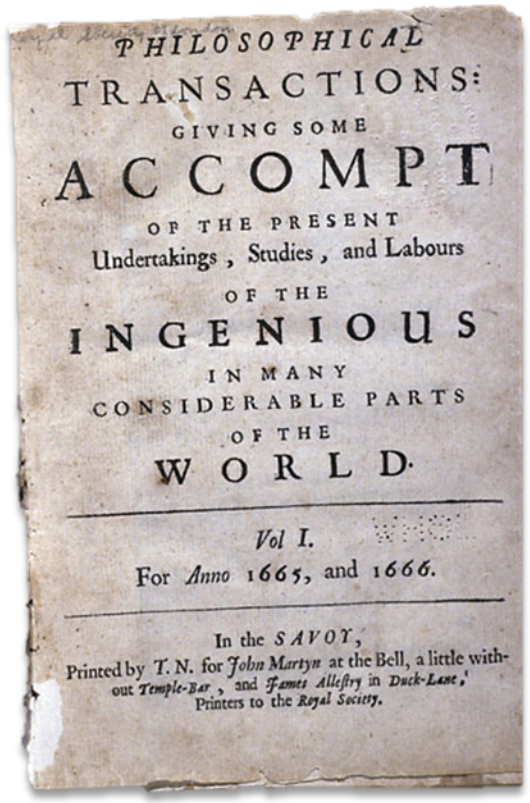
★★★★★ OL RE OF URI LD — Come i livelli precedenti ma i dati sono collegati (Linked Data)

DBpedia

Le e-Infrastrutture, il Web 3.0 ed il Metodo Scientifico

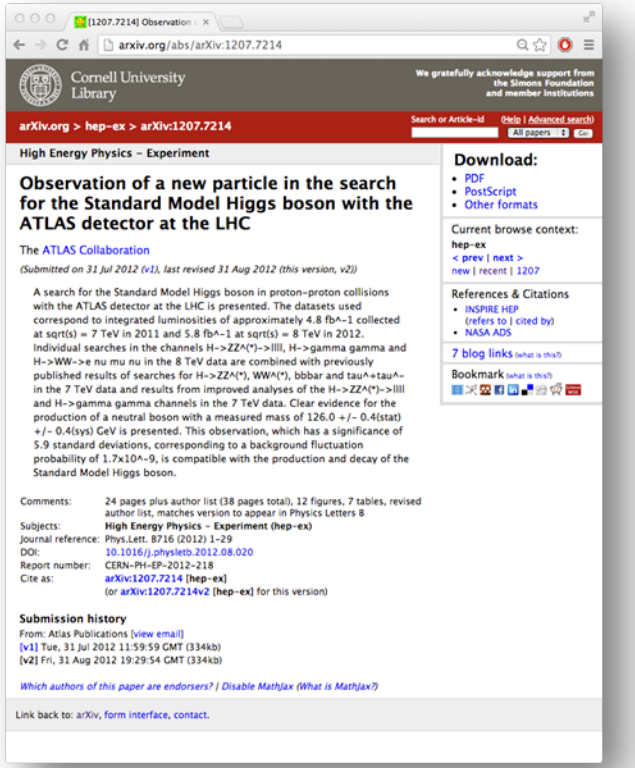


L‘“output” del Metodo Scientifico


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.


[1207.7214] Observation
arxiv.org/abs/arXiv:1207.7214
Cornell University Library
We gratefully acknowledge support from the Simons Foundation and member institutions
arXiv.org > hep-ex > arXiv:1207.7214
Search or Article-id (Help | Advanced search) All papers Go
High Energy Physics – Experiment
Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC
The ATLAS Collaboration
(Submitted on 31 Jul 2012 (v1), last revised 31 Aug 2012 (this version, v2))
A search for the Standard Model Higgs boson in proton–proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb^-1 collected at sqrt(s) = 7 TeV in 2011 and 5.8 fb^-1 at sqrt(s) = 8 TeV in 2012. Individual searches in the channels H->ZZ^(*)->llll, H->gamma gamma and H->WW->e nu mu nu in the 8 TeV data are combined with previously published results of searches for H->ZZ^(*), WW^(*), bbbar and tau^+tau^- in the 7 TeV data and results from improved analyses of the H->ZZ^(*)->llll and H->gamma gamma channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of 126.0 +/- 0.4(stat) +/- 0.4(sys) GeV is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7x10^-9, is compatible with the production and decay of the Standard Model Higgs boson.
Comments: 24 pages plus author list (38 pages total), 12 figures, 7 tables, revised author list, matches version to appear in Physics Letters B
Subjects: High Energy Physics – Experiment (hep-ex)
Journal reference: Phys.Lett. B716 (2012) 1-29
DOI: 10.1016/j.physletb.2012.08.020
Report number: CERN-PH-EP-2012-218
Cite as: arXiv:1207.7214 [hep-ex] (or arXiv:1207.7214v2 [hep-ex] for this version)
Submission history
From: Atlas Publications [view email]
[v1] Tue, 31 Jul 2012 11:59:59 GMT (334kb)
[v2] Fri, 31 Aug 2012 19:29:54 GMT (334kb)
Which authors of this paper are endorsers? | Disable MathJax (What is MathJax?)
Link back to: arXiv, form interface, contact.
Download:
PDF
PostScript
Other formats
Current browse context: hep-ex
< prev | next >
new | recent | 1207
References & Citations
INSPIRE HEP (refers to | cited by)
NASA ADS
7 blog links (what is this?)
Bookmark (what is this?)


Physics Letters B
Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC
ATLAS Collaboration
ARTICLE INFO
ABSTRACT
1. Introduction

Marked a real Scientific Revolution but…
it is the same since almost 4 centuries!

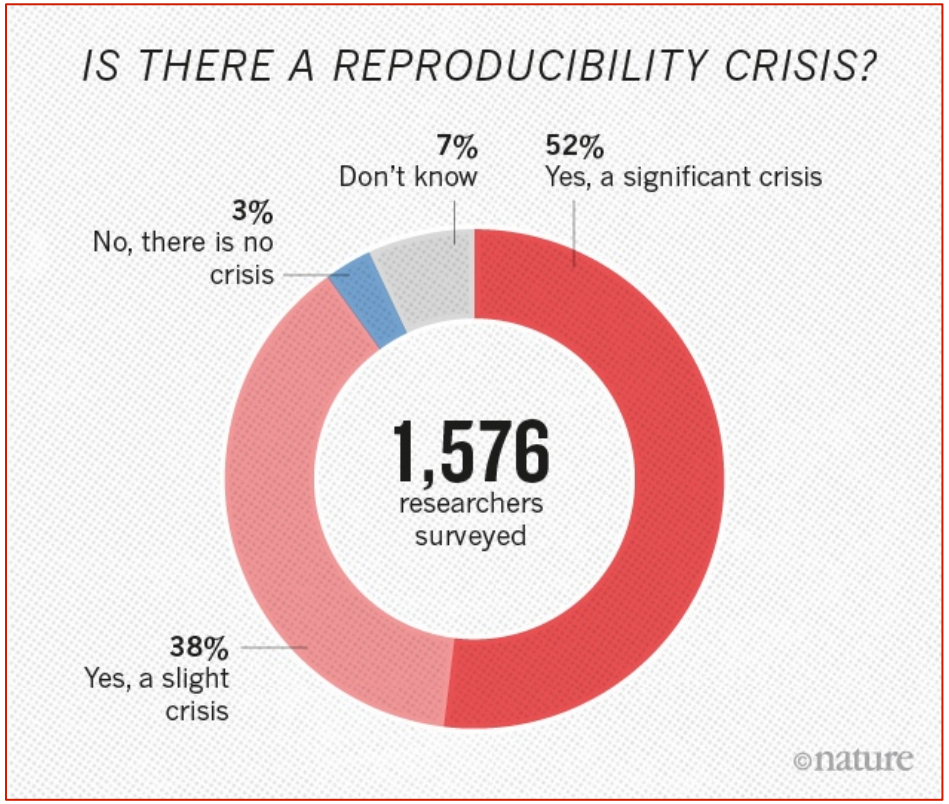
I «pilastri» del Metodo Scientifico

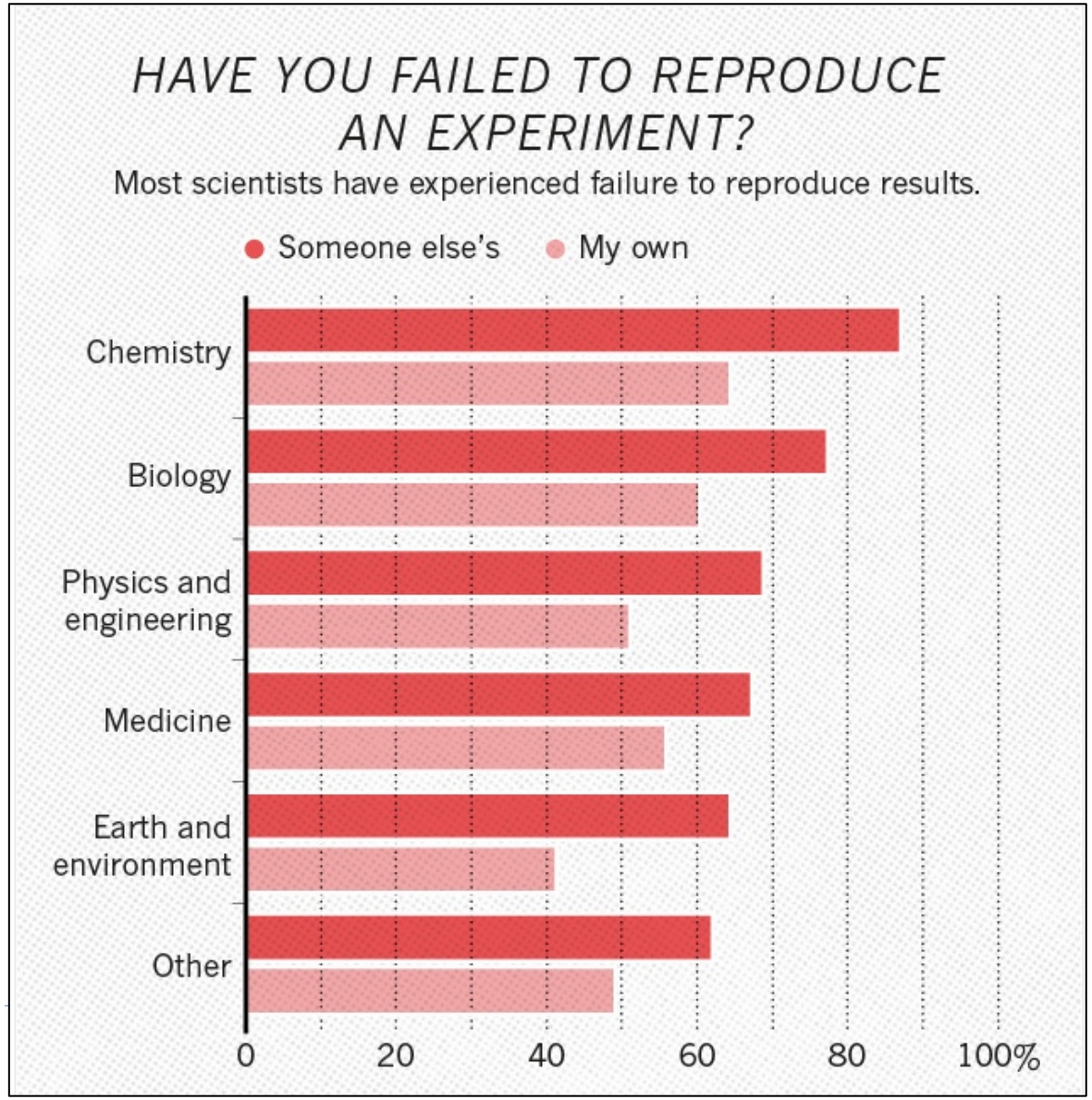
- **Repeatability**
  - The closeness of agreement between independent results obtained with the same method on identical test material, under the same conditions (same operator, same apparatus, same laboratory and after short intervals of time)
  - Affected by *random errors*
- **Reproducibility**
  - The closeness of agreement between independent results obtained with the same method on identical test material but under different conditions (different operators, different apparatus, different laboratories and/or after different intervals of time)
  - Affected by *systematic errors*

La scienza è veramente riproducibile?

Challenges in irreproducible research
(https://www.nature.com/collections/wjsrmrdnsm)

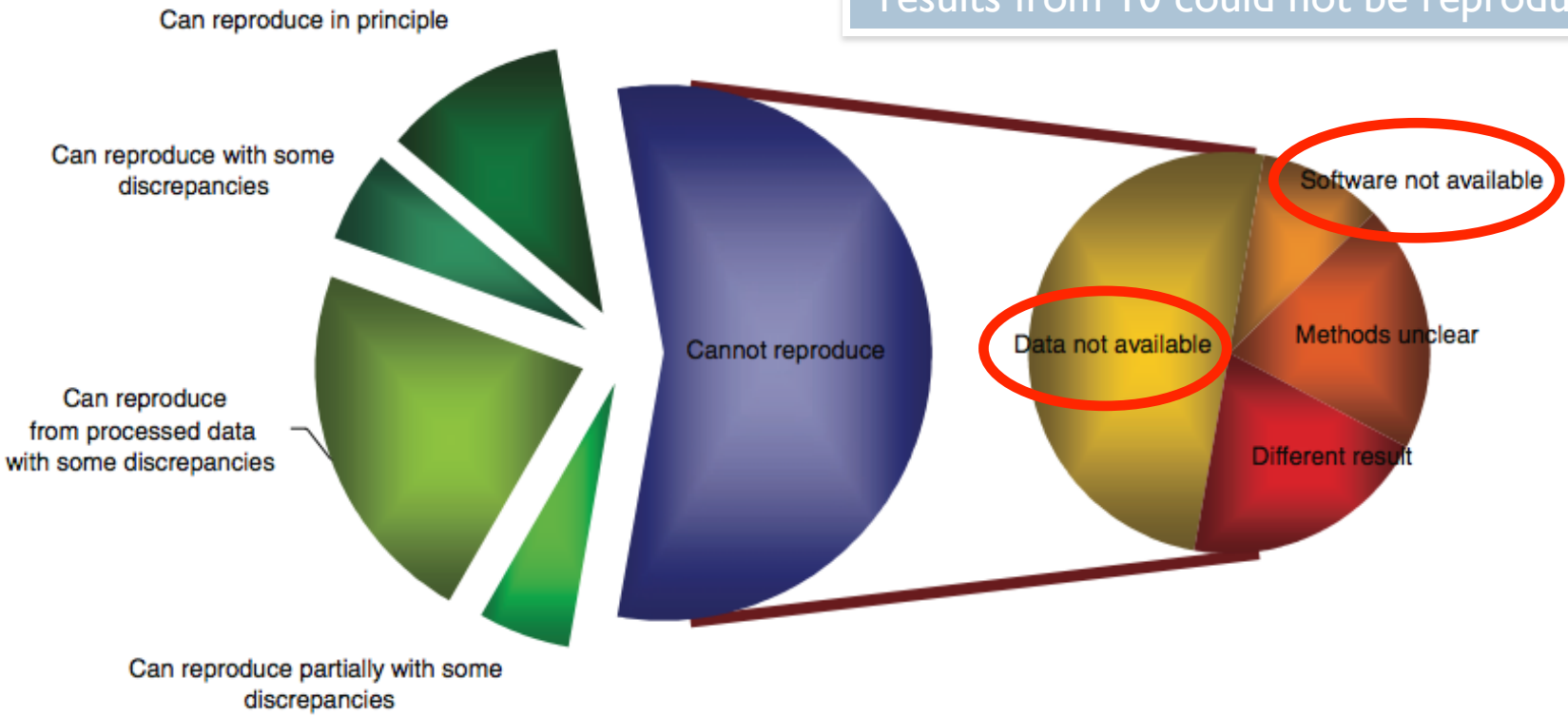
La «crisi di riproducibilità»





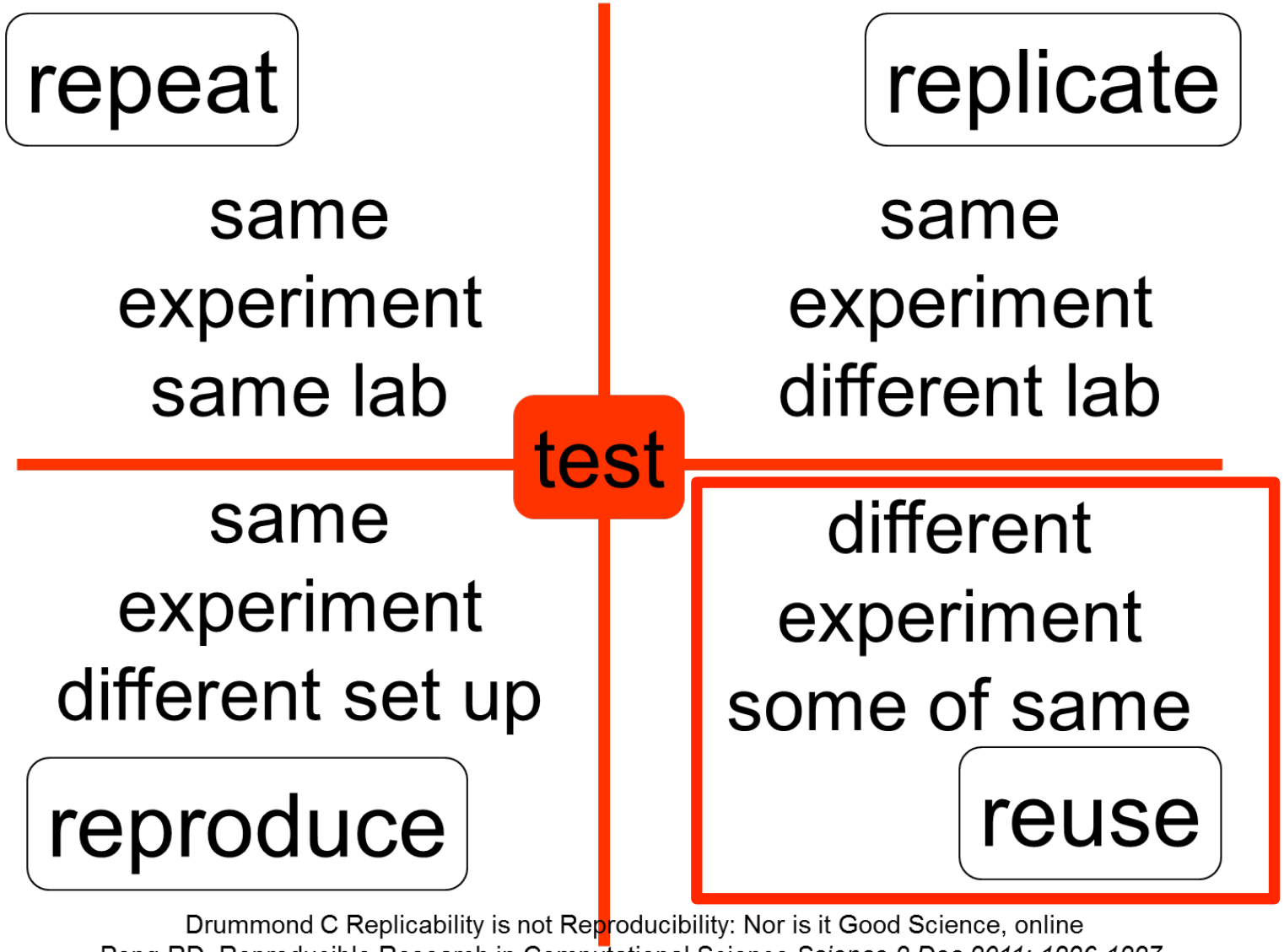
Le ragioni della «crisi di riproducibilità»

Out of 18 seminal microarray papers, results from 10 could not be reproduced



1. Ioannidis et al., 2009. Repeatability of published microarray gene expression analyses. *Nature Genetics* 41: 14
2. Science publishing: The trouble with retractions http://www.nature.com/news/2011/111005/full/478026a.html
3. Bjorn Brembs: Open Access and the looming crisis in science https://theconversation.com/open-access-and-the-looming-crisis-in-science-14950

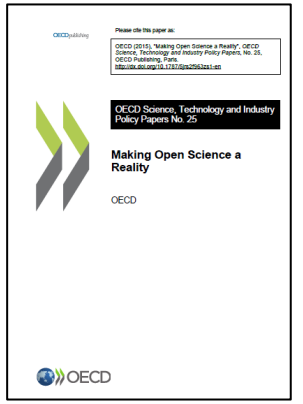
Evoluzione: la ripetibilità e la riproducibilità non sono tutto…



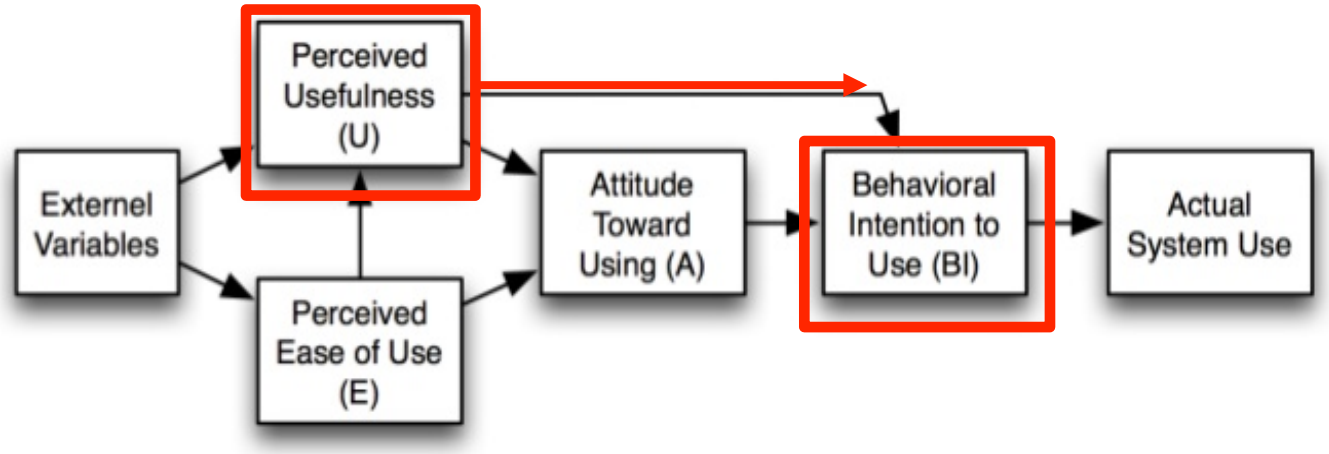
Drummond C Replicability is not Reproducibility: Nor is it Good Science, online
Peng RD, Reproducible Research in Computational Science *Science 2 Dec 2011: 1226-1227.*

La «rivoluzione» della Scienza Aperta

(http://book.openingscience.org 10.1787/5jrs2f963zs1-en) http://dx.doi.org/

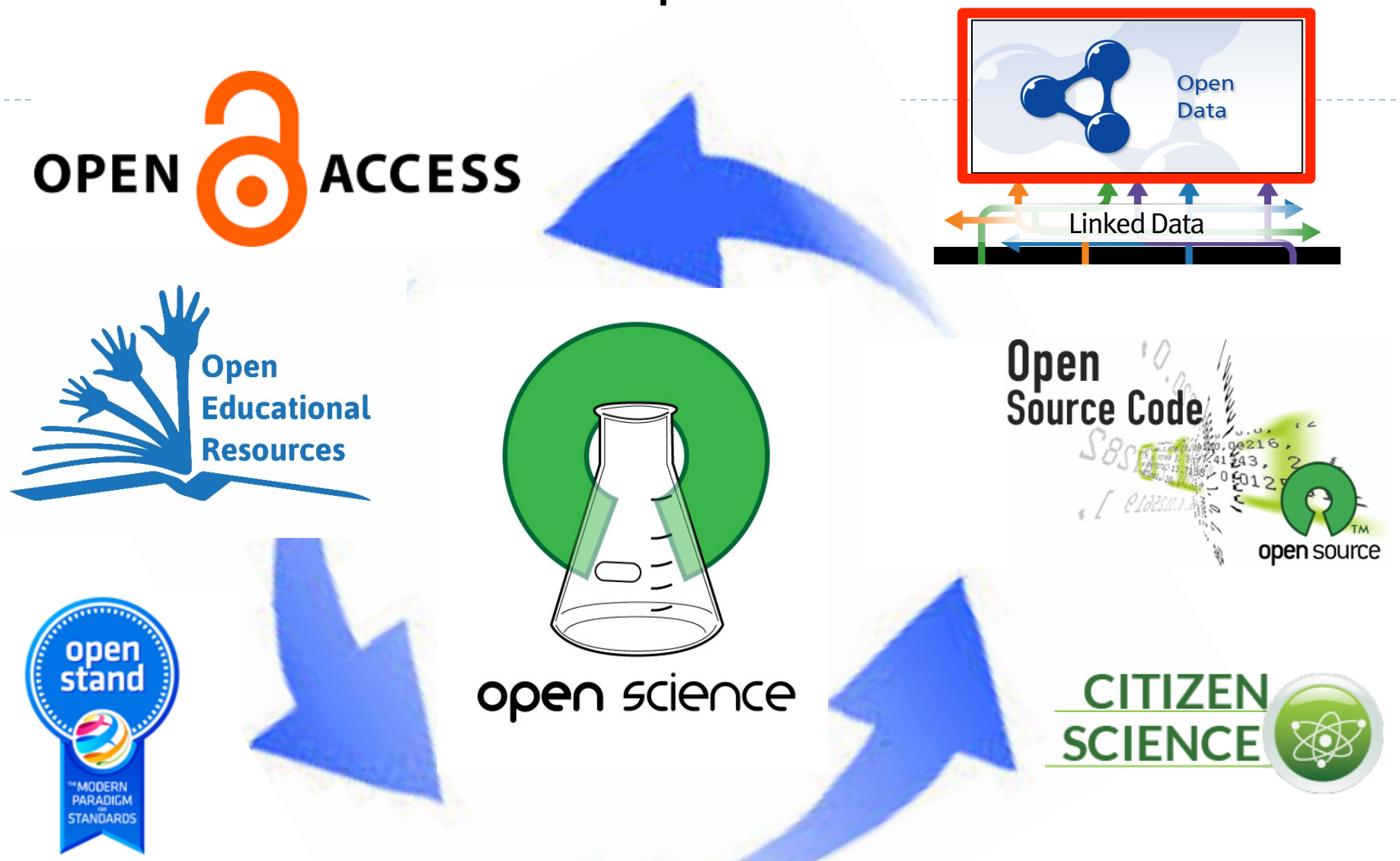


- *"**Open Science** refers to a scientific culture that is characterized by its openness. Scientists **share results** almost immediately and with a very wide audience"*
- *"**Open science** is a means and not an end in itself and it is much more than just open access to publications or data; **it includes many aspects and stages of research processes** thus enabling full reproducibility and re-usability of scientific results."*



Source: Davis, F. D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly* **13(3)**: 319–340

Gli «enzimi» della Scienza Aperta

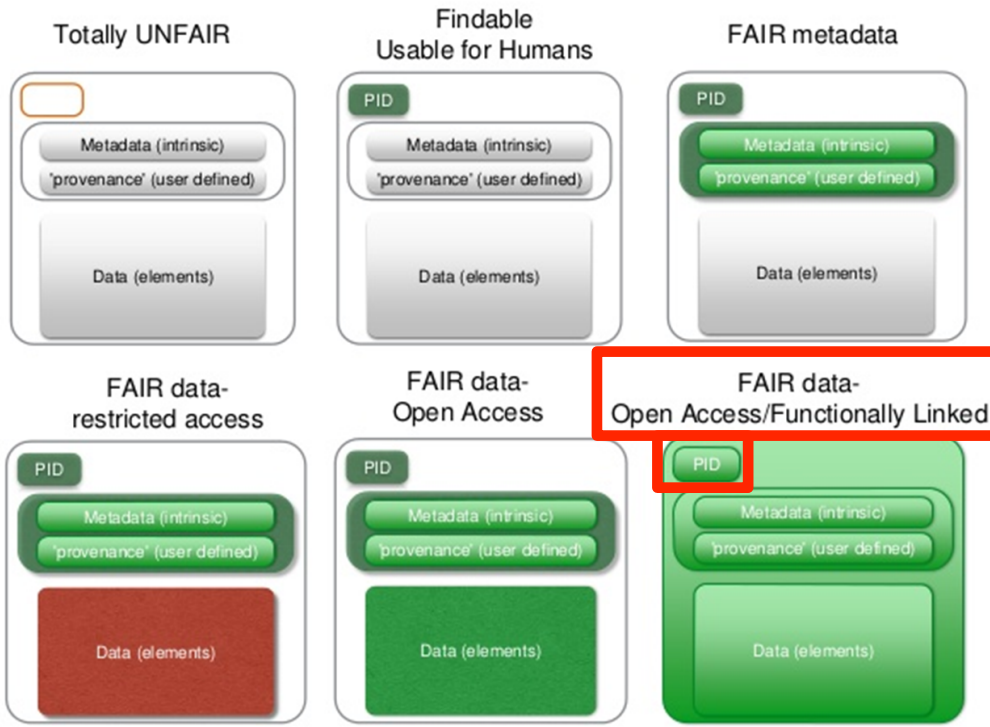


La «mutazione» dei dati: i principi **FAIR**

(https://www.force11.org/group/fairgroup/fairprinciples)

- *"One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows."* **Research data have to be**

- **Findable:**
  - F1. (meta)data are assigned a globally unique and eternally persistent identifier
  - F2. data are described with rich metadata
  - F3. (meta)data are registered or indexed in a searchable resource.
  - F4. metadata specify the data identifier
- **Accessible:**
  - A1 (meta)data are retrievable by their identifier using a standardized communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
  - A2 metadata are accessible, even when the data are no longer available
- **Interoperable:**
  - I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
  - I2. (meta)data use vocabularies that follow FAIR principles
  - I3. (meta)data include qualified references to other (meta)data
- **Reusable:**
  - R1. meta(data) have a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with their provenance
  - R1.3. (meta)data meet domain-relevant community standards

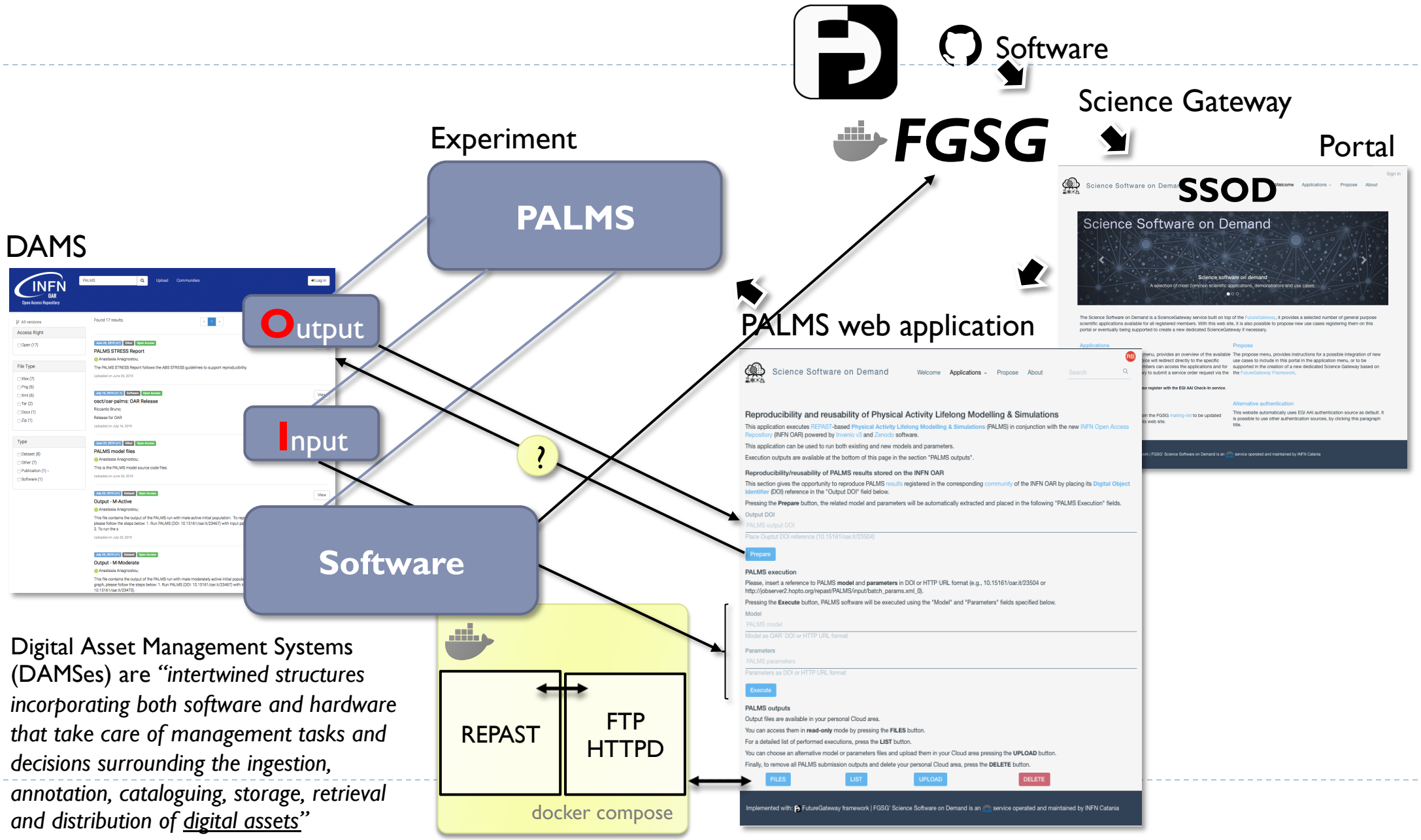


Evoluzione (→ complessità): approccio multidisciplinare a…

OK, tutto molto interessante…

… ma in pratica?

The «Reproducibility & Reusability Platform» - overview

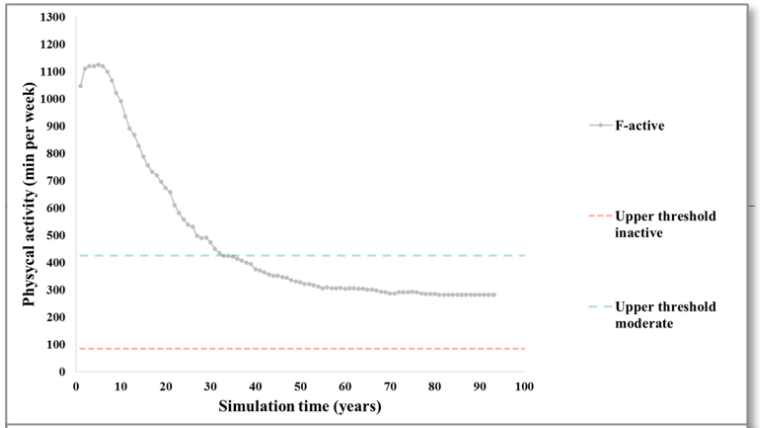


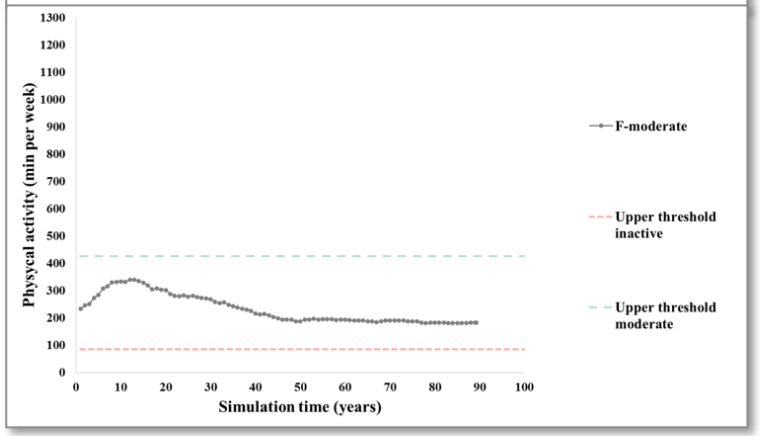
Digital Asset Management Systems (DAMSes) are *"intertwined structures incorporating both software and hardware that take care of management tasks and decisions surrounding the ingestion, annotation, cataloguing, storage, retrieval and distribution of digital assets"*

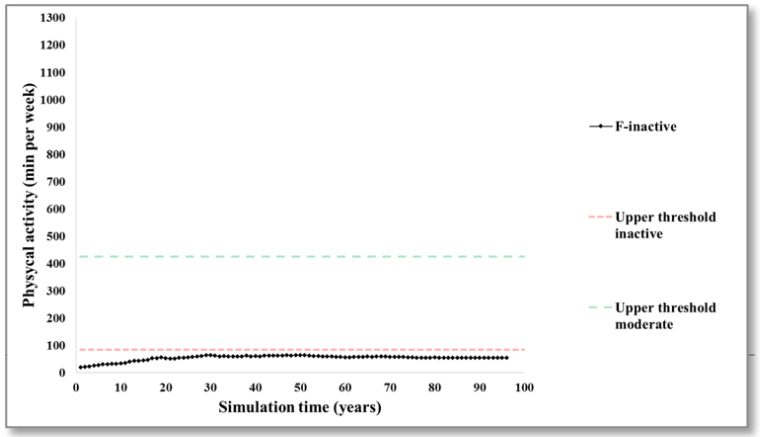
RRP ingredients: PALM simulations

- **P**hysical **A**ctivity **L**ifelong **M**odelling & **S**imulations
  - 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
  - Produces individual and aggregated quantitative outputs for quality of life and health conditions related costs

- The software
  - Uses REPAST [1], a «de facto» standard open source agent-based modeling and simulation platform
  - A specific dockerhub image exists for PALMS executions (osabuoun/repast) [2]
  - Two inputs necessary: model file (REPAST) and a parameters' file







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

RRP ingredients: The FutureGateway Framework

futuregatewayframework.github.io

INFN software project aiming to build secure and reliable Science Gateways [1]

Three core components:

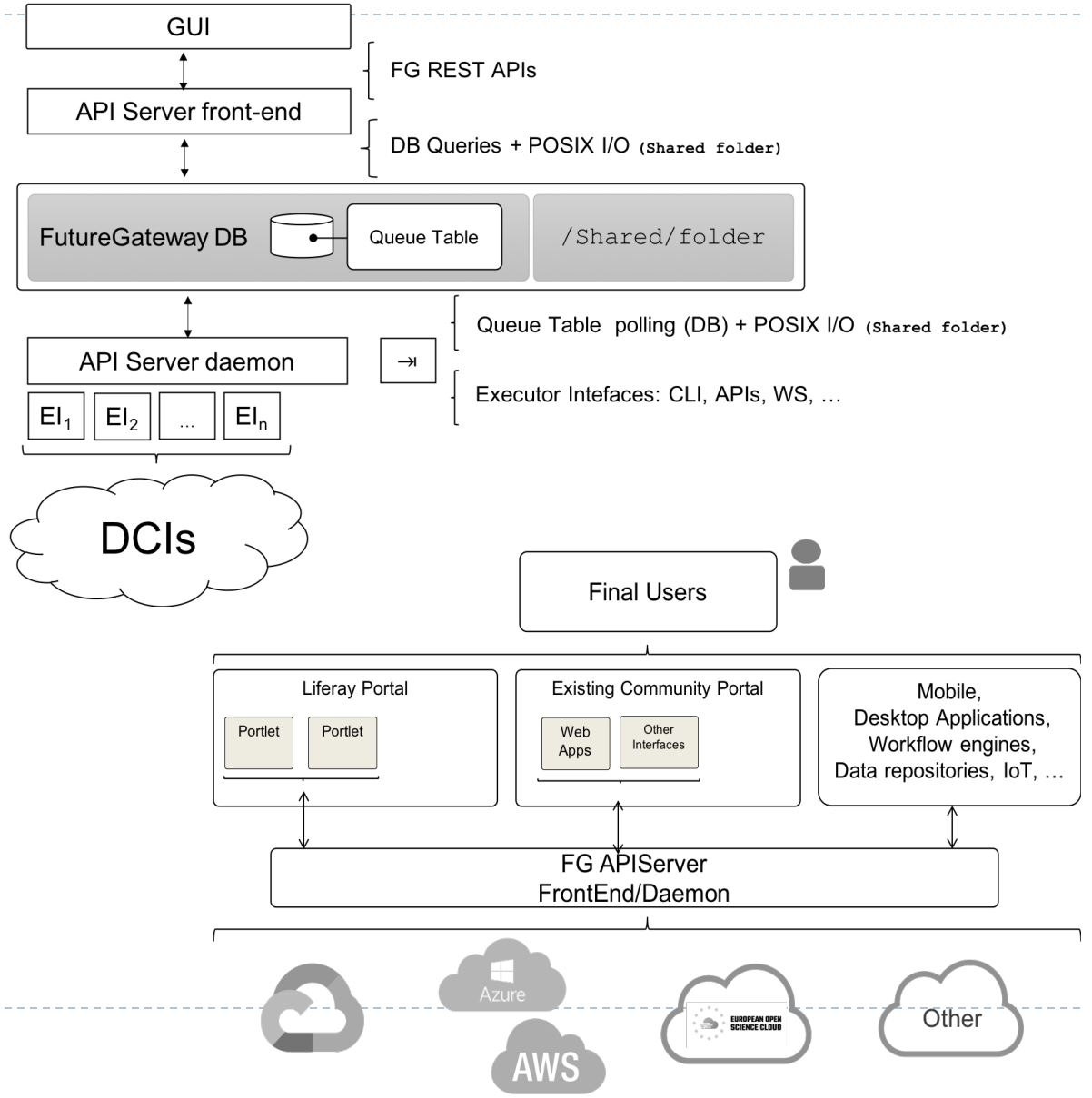
- Database, APIServer front-end, APIServer daemon + Executor Interfaces

The framework:

- Core components are enriched with a suite of tools, APIs and installation + maintenance scripts
- Open Source code available on GitHub

Targets:

- Desktop and Mobile applications, Workflow Engines, IoT and **Open Science**

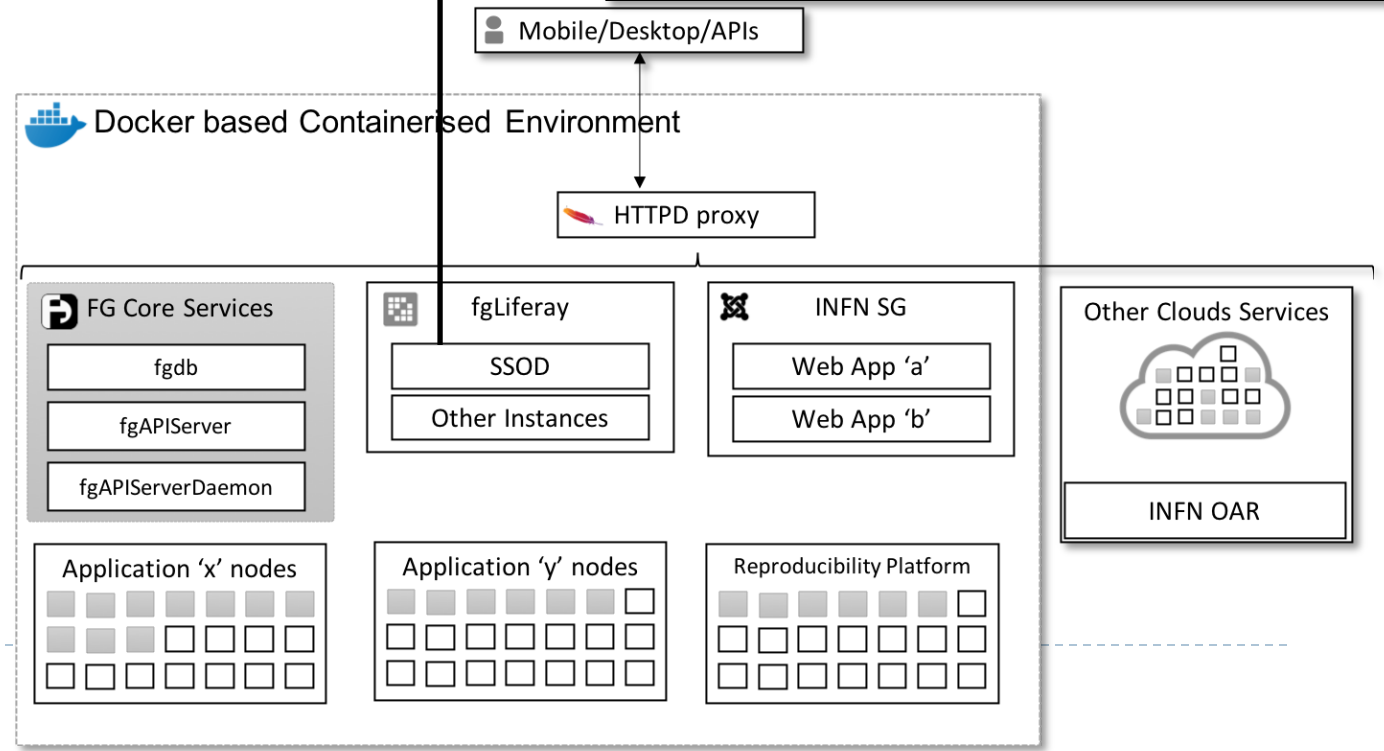


[1] https://www.xsede.org/ecosystem/science-gateways

RRP ingredients: FGSG – FutureGateway based Science Gateway

Fully docker containerised environment built in the context of the EOSC-hub project [1], to provide a General purposes Science Gateway: the EGI [2] **Science Software on Demand** (SSOD) [3]

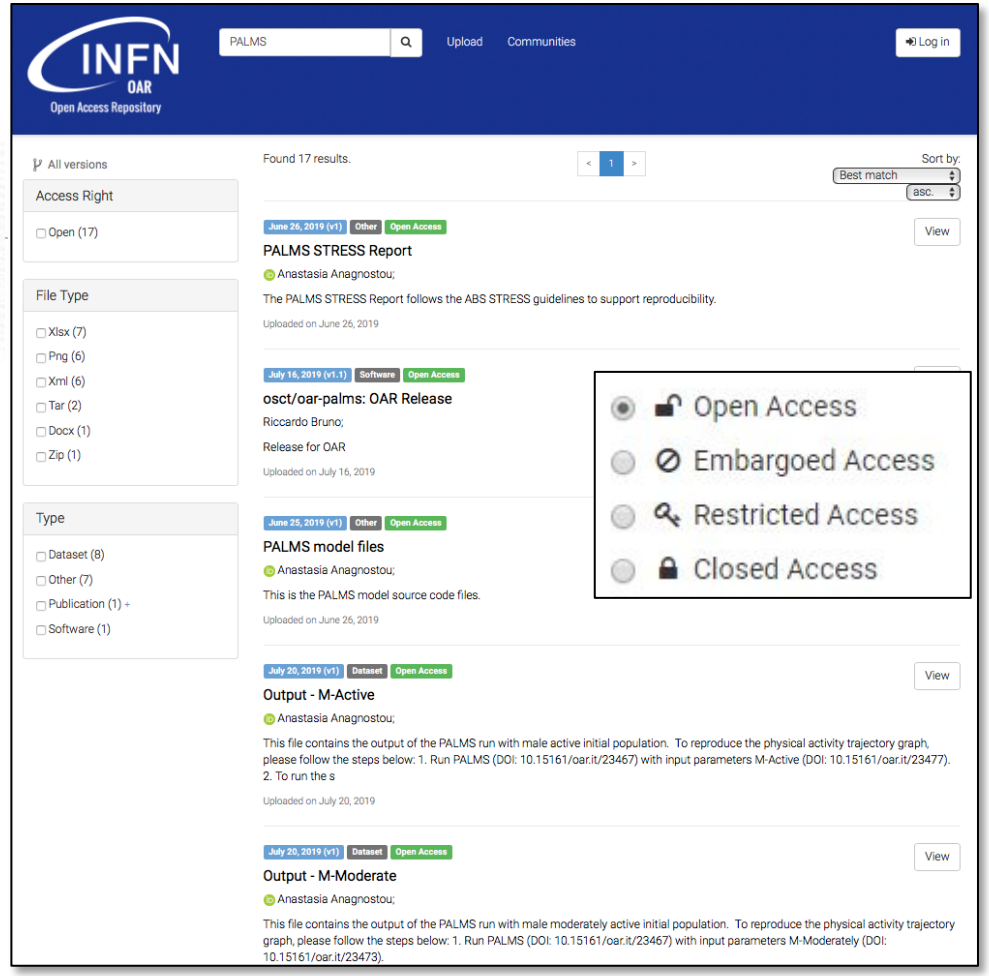
- The system allows to dynamically instantiate and destroy docker containers (it supports docker compose as well as docker swarm)
- FG core services + SSOD service
- SSOD service powered by an enterprise portal framework (Liferay)
  - One section dedicated to the **Reproducibility & Reusability Platform**
  - The platform exploits the FutureGateway and the INFN Open Access Repository

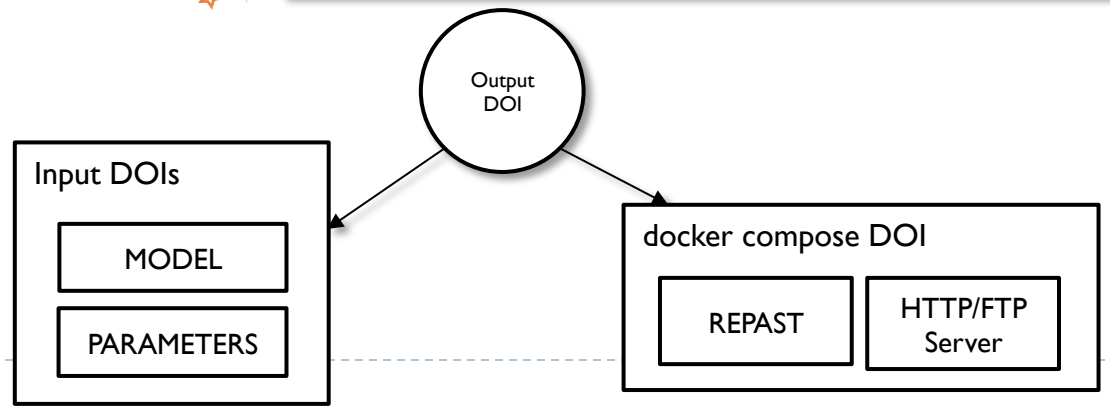


[1] https://www.eosc-hub.eu/
[2] https://www.egi.eu/
[3] https://fgsg.egi.eu/egissod/web/ssod/

RRP ingredients: the INFN Open Access Repository (OAR)

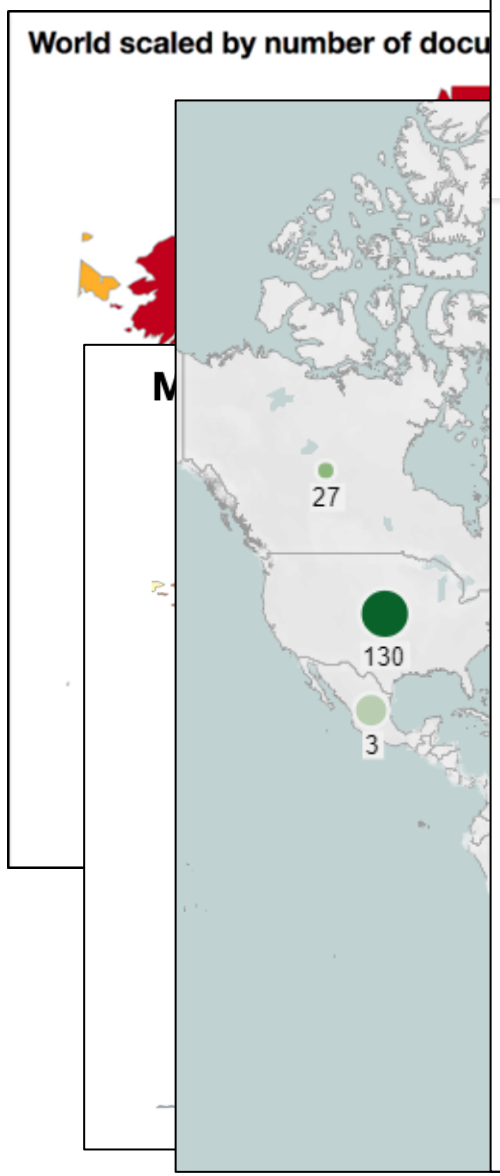
- INFN joined the Plan S [1] initiative to promote open access
- INFN OAR is a DAMS hosted at INFN Catania [2] running on a dedicated kubernetes cluster
- It uses the zenodo [3] open source software for DAMSes
- PALMS input, output, software and papers files are registered with DataCite DOIs (linked to ORCID profiles of authors)
- INFN OAR allows to define references among registered DOIs and supports DOI versioning
- Software used by *FGSG* to run PALMS simulations is published as well on the INFN OAR (linked from GitHub)

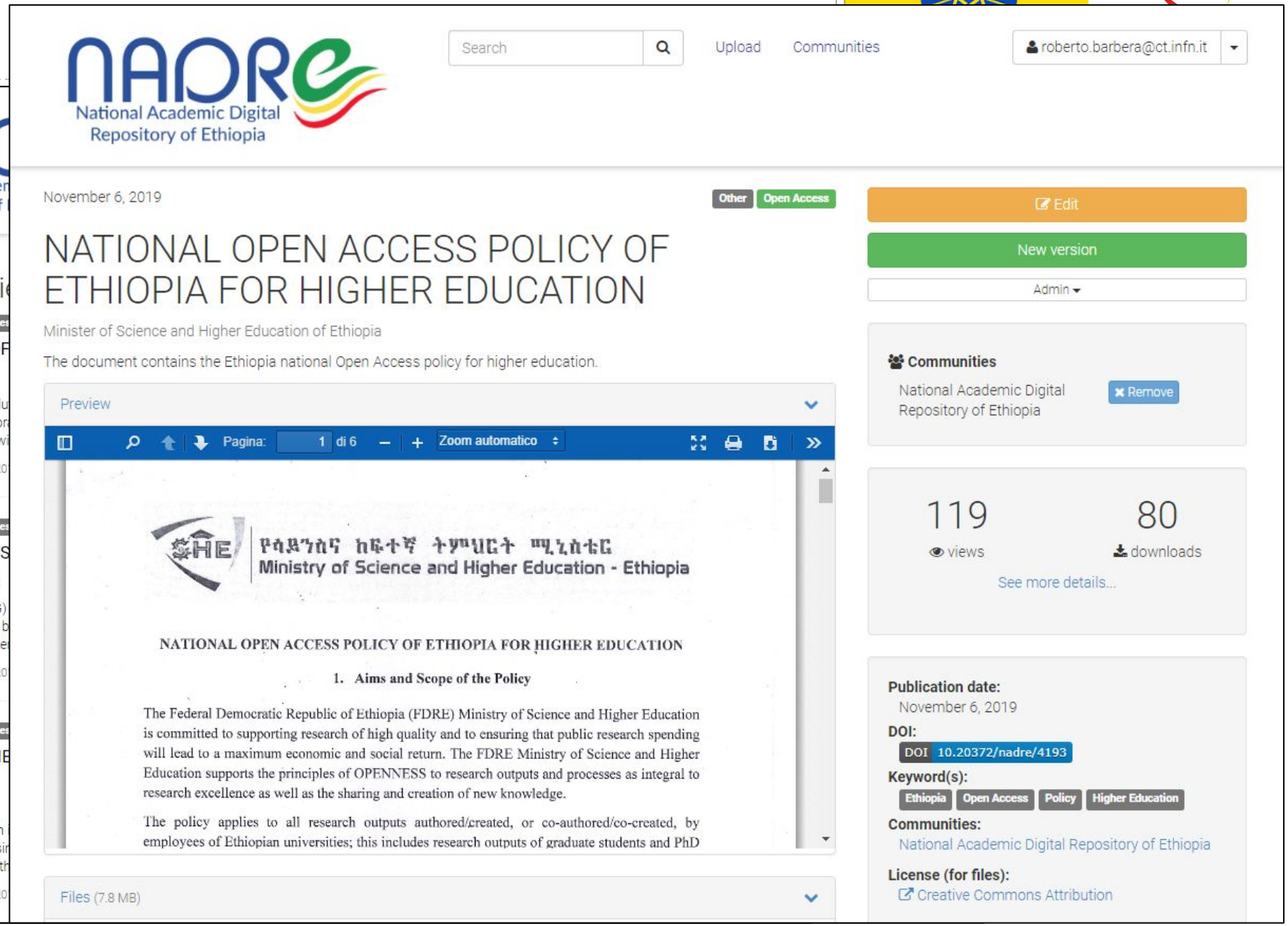




[1] https://www.coalition-s.org
[2] https://www.openaccessrepository.it
[3] https://github.com/zenodo/zenodo

Evoluzione «per contagio»: NADRE





Sommario e conclusioni

- The PALMS use case demonstrates Open Science in action
- The R&R Platform developed for PALMS can be easily reused to implement other use cases
- The INFN OAR and the R&R Platform are open to any interested communities
- FutureGateway is a mature product successfully used by:
  - Web, Desktop and Mobile Applications, Workflow engines, IoT and **Open Science demonstrators**
- Investigations are in progress to extend this work to a more general and widely adoptable solution

Grazie !