



TIMCO



Certifying the interoperability of RDF database systems

Karima Raves, BorderCloud/Inria-Saclay

Julien Nauroy, Inria-Saclay

Cécile Germain, University Paris Sud and CNRS

EQUIPE PROJET

TAO

Paris-Saclay

1



Plateforme CDS de Paris-Saclay
<https://io.datascience-paris-saclay.fr>

Plateforme "Data as a Service"
du Center for Data Science

Participants

Roughly **200 researchers** in **32 laboratories**

CDS is a transversal interdisciplinary laboratory of the Paris Saclay University.
Part of the strategic plans for scientific computational environments currently developed by the French agency for scientific research (CNRS) and Ministry of Higher Education and Research.



Themes

bioinformatics

IBISC/UEvry LRI/UPSud
Hepatinov MIAj-MIG/INRA
CESP/UPSud-UVSQ-INSERM
IGM-I2BC/UPSud MIA/Agro
LMAS/Centrale

signal processing

LTCI/Telecom LSS/Supélec
CMA/Polytechnique LIMSI
CVN/Centrale DTIM/ONERA
CMLA/Cachan

statistics

LMO/UPSud LS/ENSAE
LSS/Supélec MIA/Agro
LMAS/Centrale
CMA/Polytechnique

Earth sciences

LATMOS/UVSQ

particle physics

LAL/UPSud

visualization

INRIA LIMSI

machine learning

LRI/UPSud LTCI/Telecom
LS/ENSAE
CMLA/Cachan LSS/Supélec
MIA/Agro IBISC/UEvry
CMA/Polytechnique
DTIM/ONERA LIX/Polytechnique
LMAS/Centrale LIST/CEA
CVN/Centrale

chemistry

EA4041/UPSud

astrophysics & cosmology

LAL/UPSud IAS/UPSud
CosmoStat/CEA

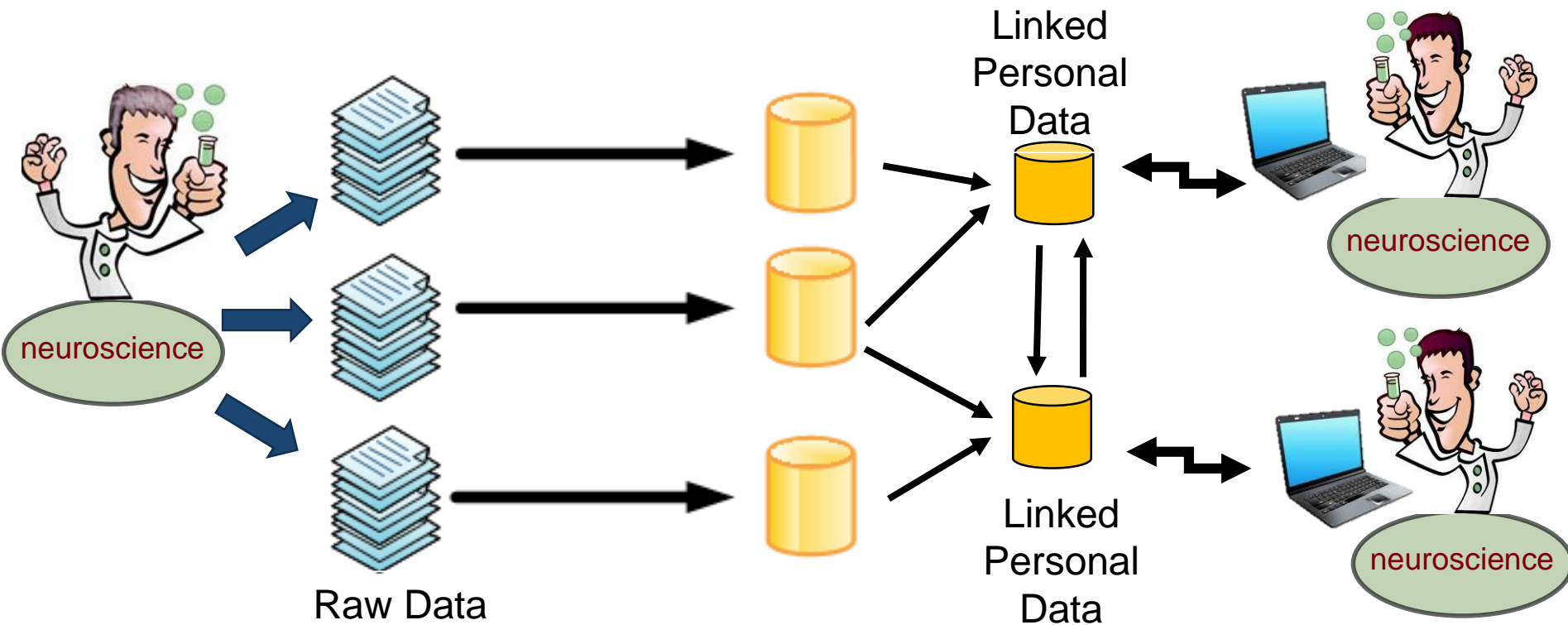
economy

LM/ENSAE RITM/UPSud
LFA/ENSAE

neuroscience

NeuroSpin/CEA
UNICOG/INSERM

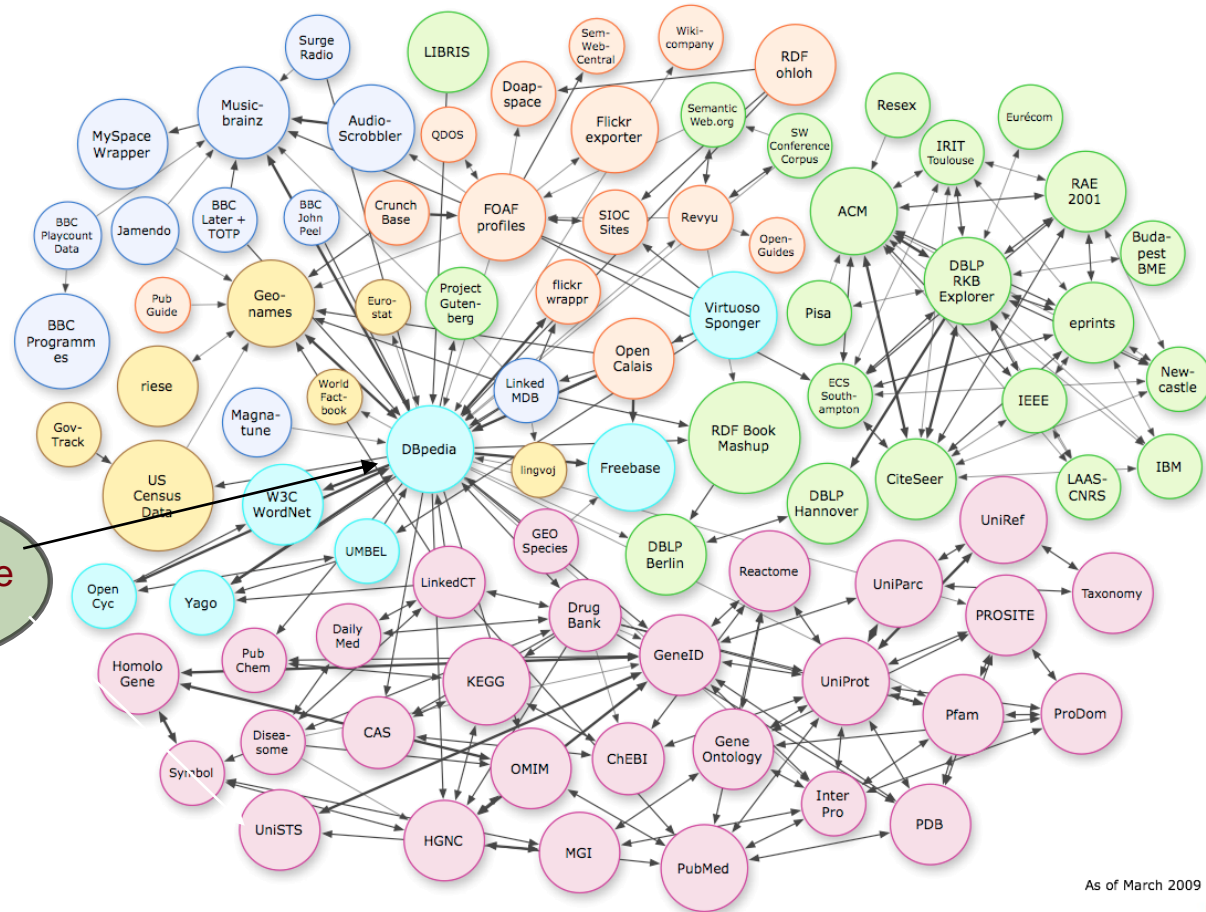
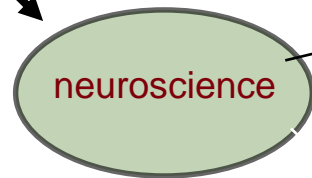
Convert 'raw' to Linked Data in the laboratories



Building in parallel a Linked Data platform for hosting their data

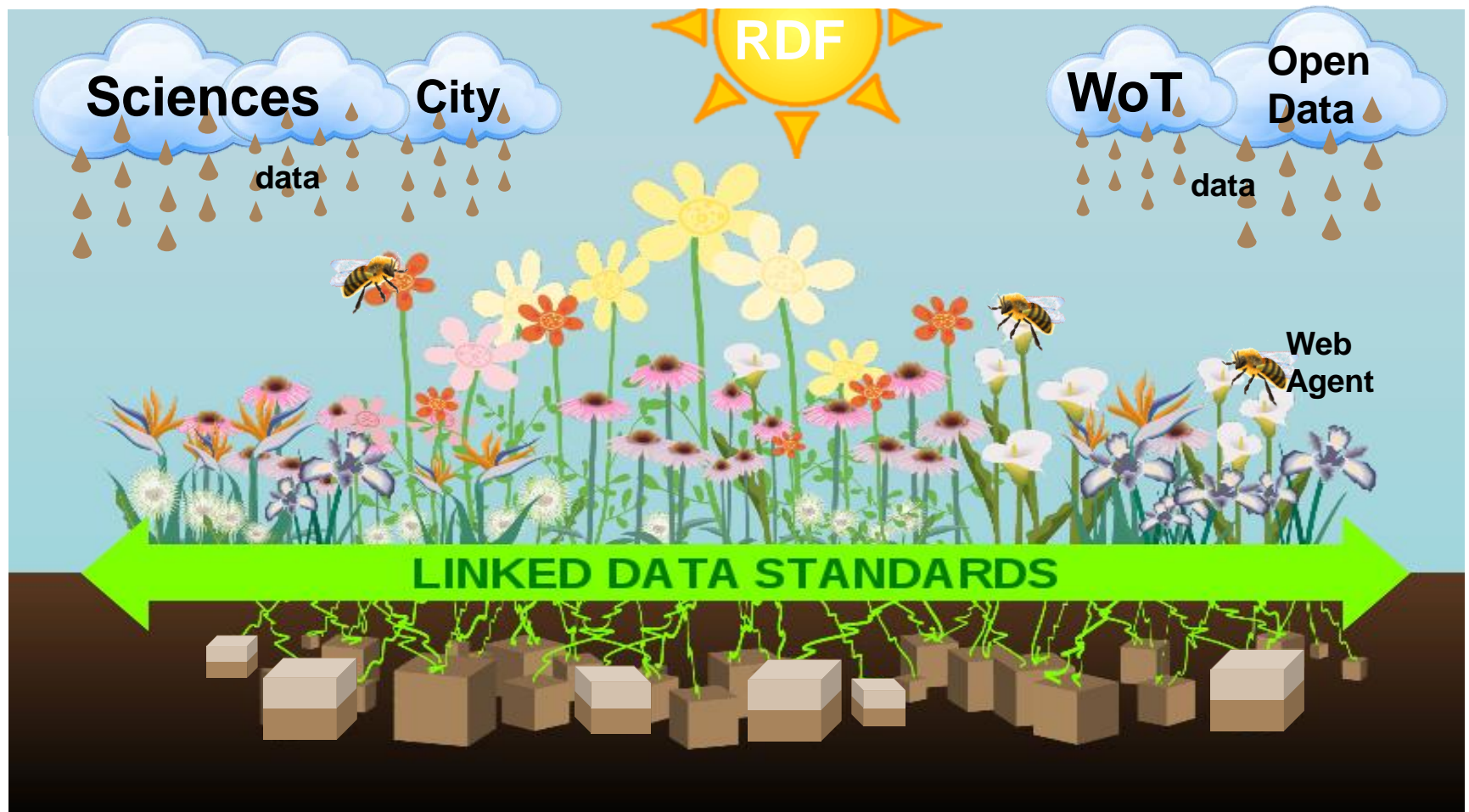


Publish



As of March 2009

**Why Linked Data ? Because in theory,
any service or tools of scientists can reuse the data**



Demo, how discover the datasets of scientists ?

<https://io.datascience-paris-saclay.fr/appDisplayArtefactResearch.php>



Article [Talk](#) [Research](#) Read [Edit](#) [View history](#) More

Herschel Space Observatory(Q209630)

List of datasets in relation with this article

- [HESIOD](#) : The Herschel IdOc Database is delivering photometric maps and spectral cubes from the PACS and SPIRE instruments (IR domain), reprocessed at IAS with the latest ESA pipelines and with high level customized pipelines. Virtual Observatory compatible. ([source](#))

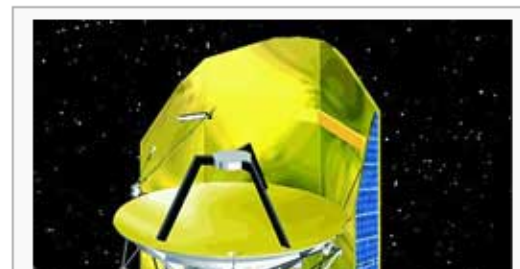


e free encyclopedia

about the space telescope. For the ground-based telescope, see [William Herschel Telescope](#).

Space Observatory was a [space observatory](#) built and [European Space Agency](#) (ESA). It was active from 2009 to a large infrared telescope ever launched,^[2] carrying a single) mirror^{[2][3][4][5]} and instruments sensitive to the [far infrared](#) wavebands (55–672 μm). *Herschel* was the fourth on in the ESA science programme, along with [Rosetta](#), . NASA is a partner in the *Herschel* mission, with US ibuting to the mission: providing mission-enabling instrument

Herschel Space Observatory



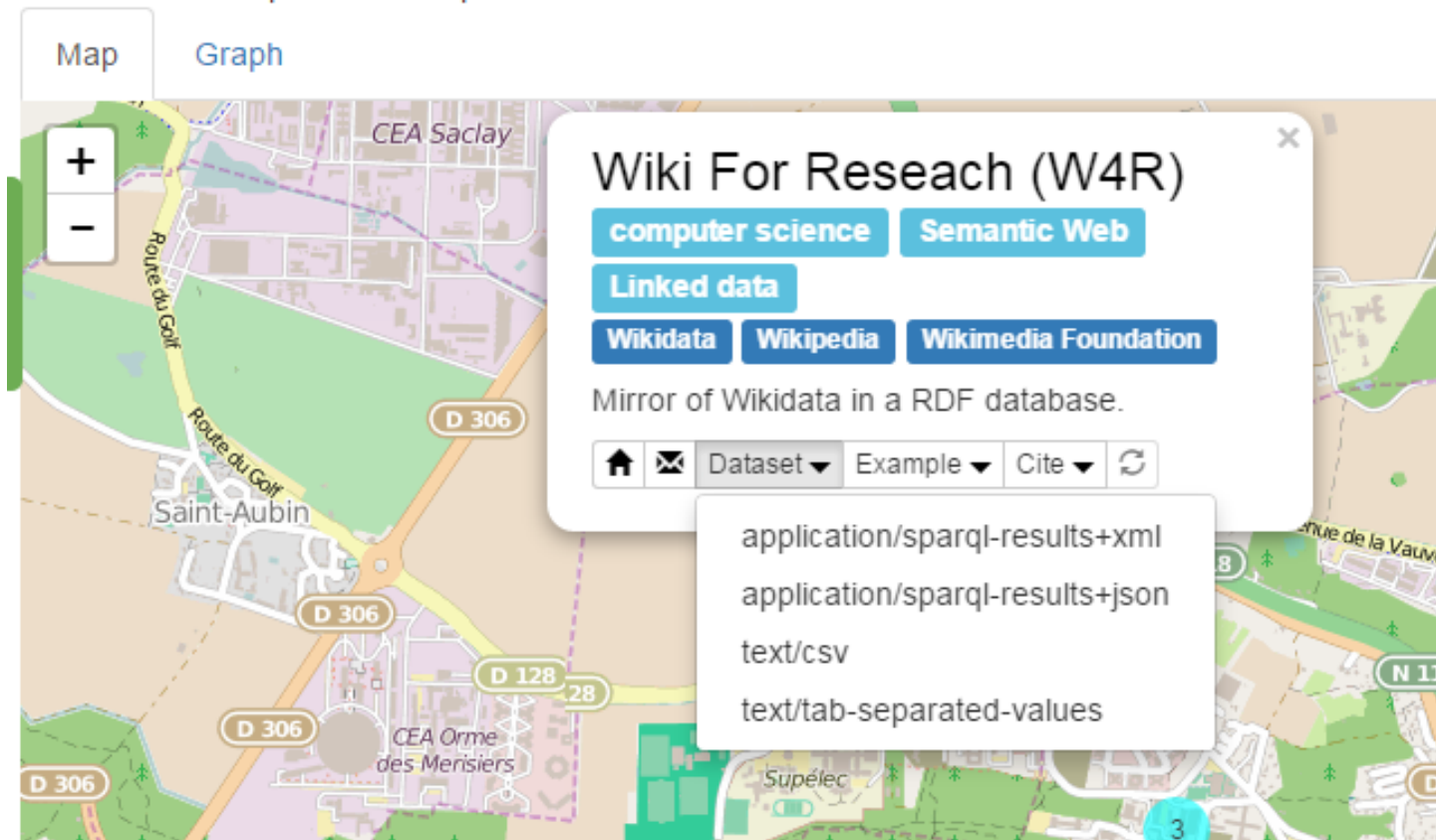
First result :

Front and back office to publish their datasets

<https://io.datascience-paris-saclay.fr>

Search an Open Dataset at Paris-Saclay

Locate on the map the actual open datasets.



The screenshot displays a web interface for searching open datasets. At the top, there are two tabs: 'Map' (selected) and 'Graph'. Below the tabs is a map of the Paris-Saclay area, showing roads like 'Route du Golf' and 'D 306', and locations like 'CEA Saclay', 'Saint-Aubin', and 'Supélec'. A zoom control with '+' and '-' buttons is on the left. A dataset card for 'Wiki For Research (W4R)' is overlaid on the map. The card includes the title 'Wiki For Research (W4R)', tags for 'computer science' and 'Semantic Web', a 'Linked data' section with links to 'Wikidata', 'Wikipedia', and 'Wikimedia Foundation', and a description: 'Mirror of Wikidata in a RDF database.' Below the description are icons for home, email, and a dropdown menu for 'Dataset'. A secondary dropdown menu is open, showing options: 'application/sparql-results+xml', 'application/sparql-results+json', 'text/csv', and 'text/tab-separated-values'.

2

Linked Data, in practice

Interoperability is not optional

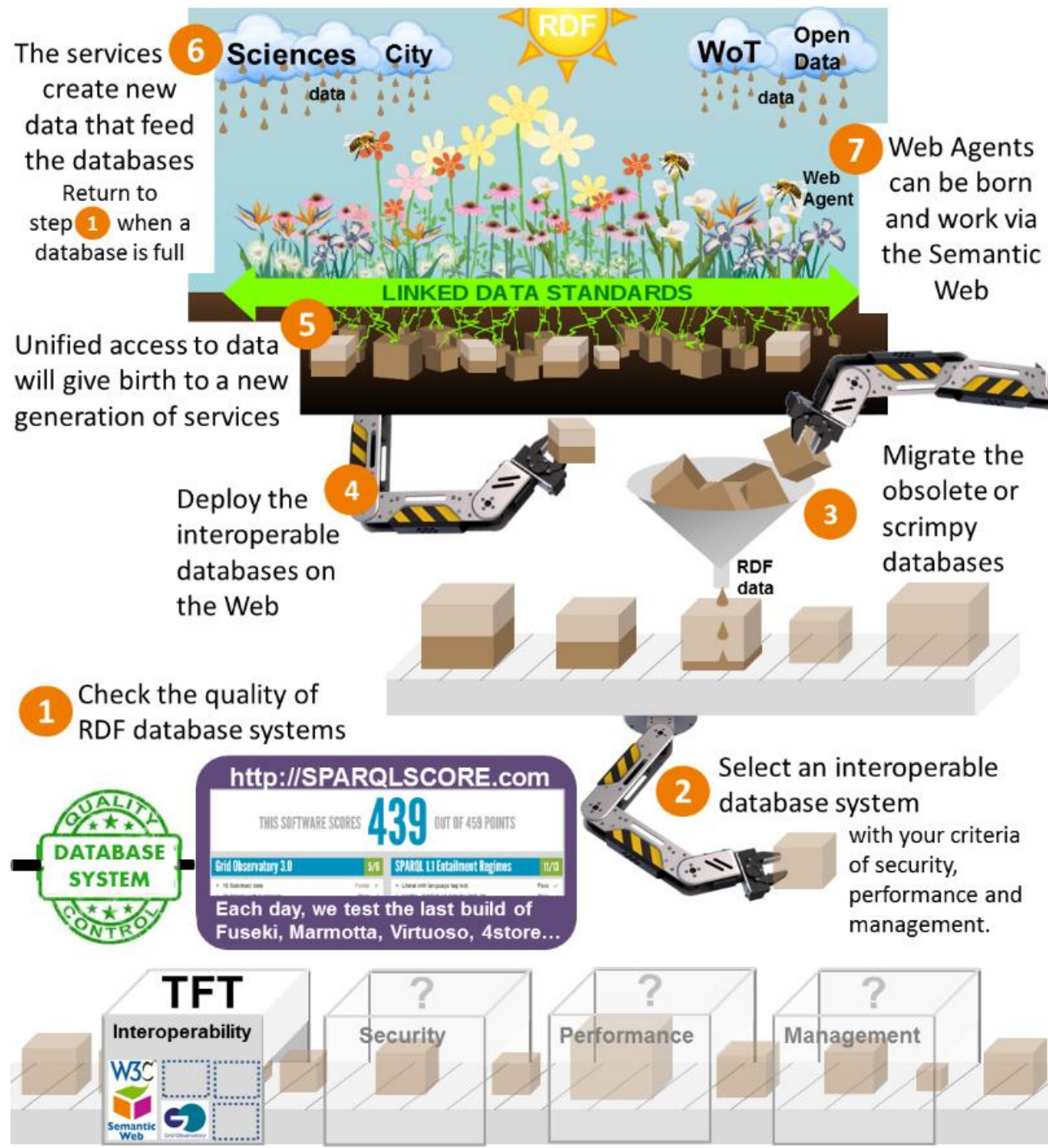
Lack of interoperability causes two complications

- Migration between databases and their updates
 - The last version is always the better in the science
I want... inference, velocity, etc.
→ To accelerate the science
 - Needs in the new platform IaaS and PaaS
To linking data to results of science
→ Reproducibility as the ultimate goal
- Development...
The same code doesn't work with another endpoint SPARQL !
→ But, SPARQL is also a protocol ? No ?

A first solution :

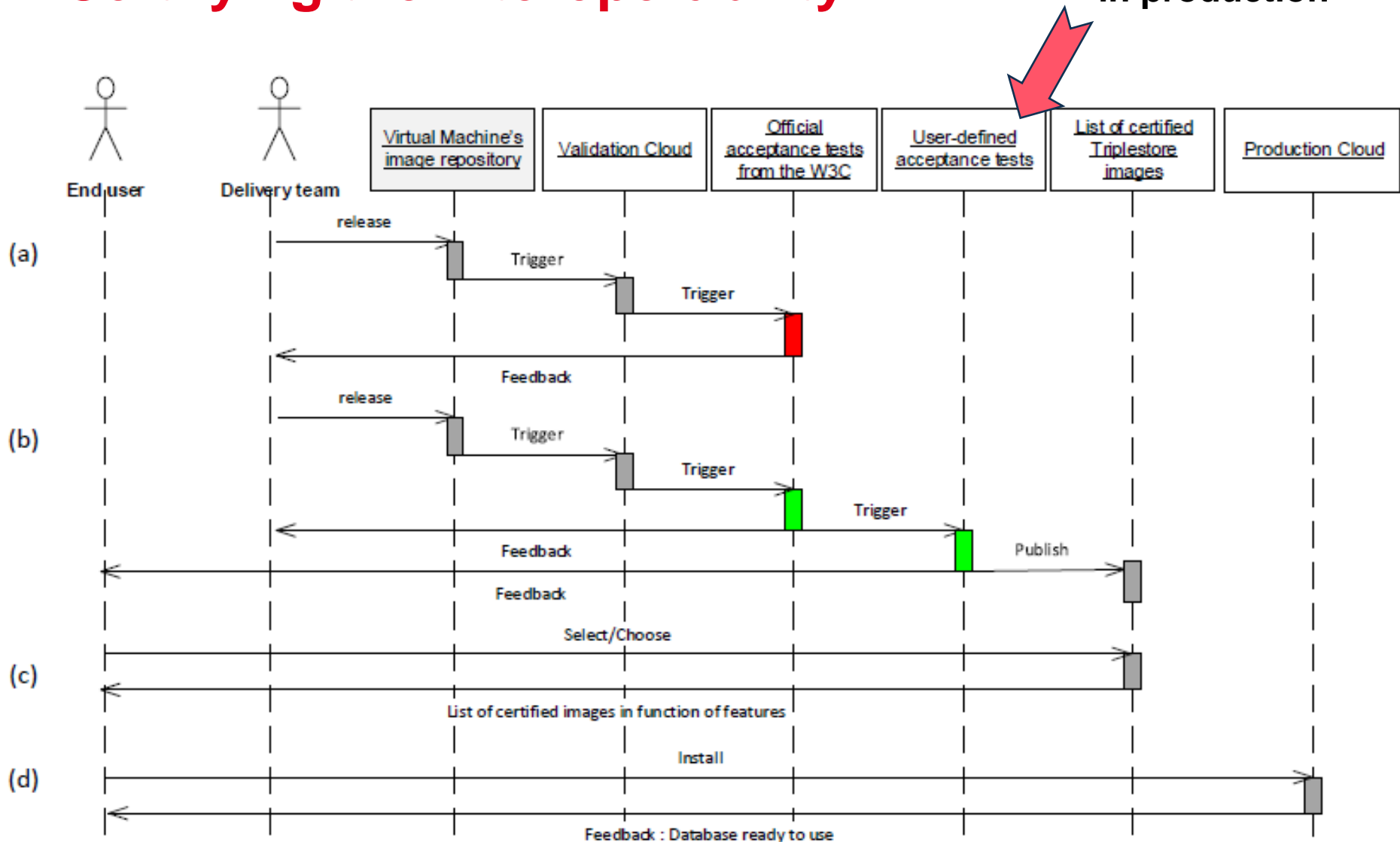
We check
the database
systems

We give
our results to
scientists or
developers
before to deploy
a new RDF database
in the cloud of the
university.



A better solution : Certifying the interoperability ?



A user can add tests, ie his query in production



Test framework named TFT (Tests for Triple stores)

<https://github.com/BorderCloud/TFT>

We push our results on the web site : Sparqlscore.com
Everybody can reproduce the same tests with TFT.

SPARQLSCORE (BETA) - the soon-to-be reference in triplestore benchmarking  Tweet 34  8+1 1

LIST OF TESTED TRIPLESTORES

Server name	version	test tool	score	test date
Fuseki	v1.1.1-SNAPSHOT	TFT v0.2	380/459	2014-09-05 04:47:55
Marmotta-KiWiStore	3.2.1	TFT v0.2	341/459	2014-09-05 05:18:12
Software1	v0.0	TFT v0.2	335/459	2014-09-05 05:41:51
Software3	v0.0	TFT v0.2	317/459	2014-09-05 06:44:24
Software2	v0.0	TFT v0.2	271/459	2014-09-05 06:19:26
4Store	v1.1.5	TFT v0.2	209/459	2014-09-05 04:04:53

THIS SOFTWARE SCORES **380** OUT OF 459 POINTS

Triplestore tested: Fuseki v1.1.1-SNAPSHOT

Testing software used: TFT v0.2

[Back to the list of triplestores](#)

grid observatory tests

Grid Observatory 3.0

4/6

▼ 10 Subtract date	Partial	○
10 Subtract date : Test the protocol.	Pass	✓
10 Subtract date : Test the response.	Fail	✗
▶ View the original test suite	Pass	✓
▶ Results:	Pass	✓
▶ View the errors	Pass	✓
▶ 50 Filter date & order by date	Pass	✓
▶ 60 Query to calculate ERT-ART	Fail	✗

sparql 1.1 tests

SPARQL 1.1 Entailment Regimes

10/13

▶ Literal with language tag test	Pass	✓
▶ bind01 - BIND fixed data for OWL DL	Pass	✓
▶ bind02 - BIND fixed data for OWL DL	Pass	✓
▶ bind03 - BIND fixed data for OWL DL	Pass	✓
▶ bind04 - BIND fixed data for OWL DL	Pass	✓
▶ bind05 - BIND fixed data for OWL DL	Pass	✓
▶ bind06 - BIND fixed data for OWL DL	Pass	✓
▶ bind07 - BIND fixed data for OWL DL	Pass	✓

Conclusion :

Who will check in the Semantic Web the Linked Data Quality about the protocol ?

- Benchmarking the velocity but without the protocol is insufficient
- An open benchmark is possible and can help to converge
- The SPARQL 1.1 is a recommendation but not the tests.
 - There are again works... How to pay that ?
- How help to create a really interoperable ecosystem ?
 - Like HTML5 with "Test the Web Forward" or ?

Thanks Questions ?

Karima.rafes@gmail.com

