



The value of RDA outputs for a research institute specialized in Agri-food sciences

Odile Hologne, Head of the department of scientific information @Holo_08

In collaboration with : Imma Subirats (FAO), Sophie Aubin, Esther Dzalé, Michael Chelle (INRA)



Challenges & missions of INRA

OBJECTIVE: meet the challenges of agricultural research

Food & nutrition



Agriculture



Environment



➤ Develop agricultural systems that are:

- Competitive and respectful of the environment, regional differences and natural resources;
- Adapted to human dietary needs and to new uses of agricultural products.

➤ Through international and multidisciplinary scientific research of excellence and by disseminating results to all stakeholders

 An overview of INRA – November 2017



2 The world's second greatest producer of publications in the agricultural sciences



250 laboratories including **45** experimental units across

- **13** scientific divisions and
- **17** research centres



An annual budget of **850.89** million euros

- **77%** of funds coming from the Ministry of Research
- and **20%** from other publically-funded sources



A community of **13 000** people including

- **7 903** permanent staff,
- **1 849** full-time researchers,
- **2 353** interns,
- and **556** PhD students with INRA grants

INRA strategic framework (oct 2016)

THREE GENERAL POLICY DIRECTIONS

[#OpenScience]

Opening science with digital tools



- #OpenScience-1: Interconnected research infrastructures
- #OpenScience-2: Data organisation for sharing and reuse
- #OpenScience-3: Predictive approaches in biology
- #OpenScience-4: New approaches for knowledge dissemination
- #OpenScience-5: Evolving research professions and environments in a digital world

→ **Predictive approaches in biology and ecology**

[#OpenInra]

A nation-wide actor in innovation

- #OpenInra-1: Greater openness towards higher education and regional partnerships
- #OpenInra-2: The mobilisation of all INRA's expertise in support of public policies
- #OpenInra-3: The road to innovation benefits from reinforced management
- #OpenInra-4: Science open to non-market actors in society

→ **A national strategy for increasingly specialised regions**

[#Support]

Anticipating and supporting change

- #Support-1: An efficient, agile and resilient organisation
- #Support-2: A reliable and solidarity-based financing
- #Support-3: An institute that attracts and motivates its staff
- #Support-4: INRA's actions and values visible and shared through active external and internal communication
- #Support-5: Efficient and shared institutional management



Agenda

- ❖ Data challenge in agri-food sciences
- ❖ RDA outputs
 - ✓ what we do
 - ✓ what we use
- ❖ Conclusion



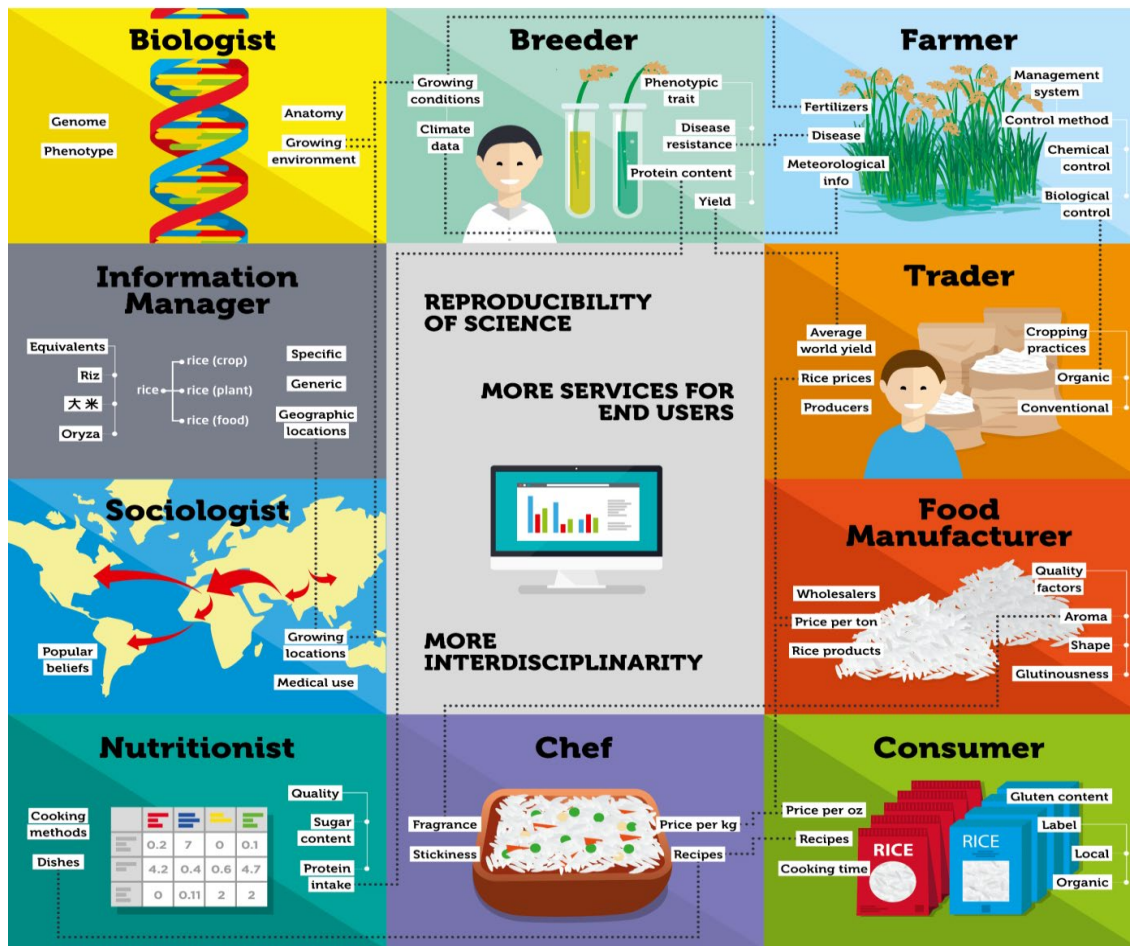
Data challenge in agri-food sciences

Data challenges in agri-food sciences

- ❖ Massive data production in labs (sensors, robots, models) but also in farms and by the citizens (of huge interest for science)
 - ✓ Big data : more variety than volume
- ❖ Data silos, poorly documented, not easy to find, nor to access (same for semantic resources)
- ❖ Different level of maturity of the practices about data (management, sharing, analysis)
- ❖ Not only about data : code, workflow ...
- ❖ Disruption in the knowledge ecosystem (see next slide)

SEMANTICS - THE WAY TO RECONCILE POINTS OF VIEW AND DATA

THE EXAMPLE OF "RICE"



ISSUES

Diversity of focus
Conflicting view points
Scale / granularity
Language
Synonymy & ambiguity
Silos

SOLUTIONS

Ontologies & skos resources
Network of ontologies
Documentation
Standards (RDF... W3C)
Persistent identifiers
Shared infrastructures

Aubin S, RDA Agrisemantics Working Group and RDA Rice Data Interoperability Working Group. Semantics – The way to reconcile points of view and data [version 1; not peer reviewed]. F1000Research 2017, 6:1871 (poster) (doi: 10.7490/f1000research.1114998.1)

Agri-food science: The need to combine, integrate data!



Blind men exploring an elephant (Hokusai, 1818)

Digital

- Big data technology (store, query, compute)
- FAIR data
- Semantic level
- Complex system
- Simulation
- Deep learning
- Interoperability (data / model)
- Sharing
- Collaborative

=> e-infrastructure

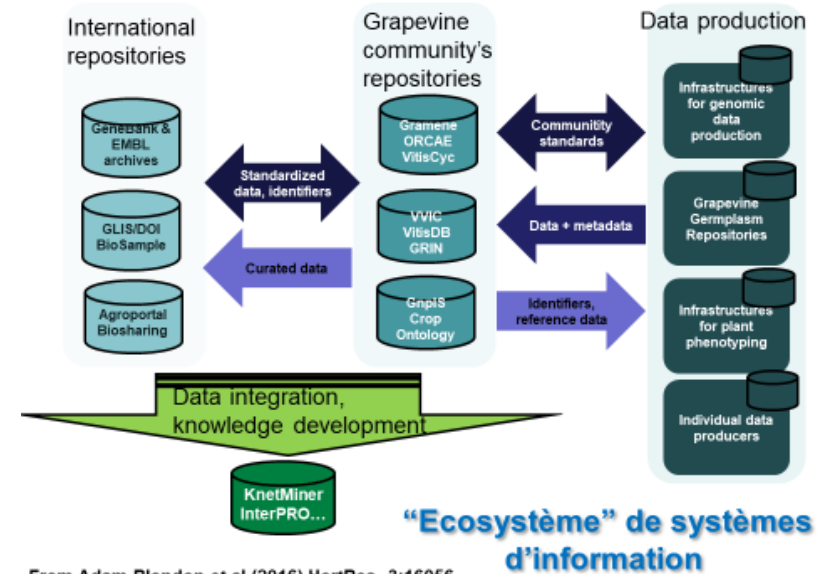
Source : Michael Chelle, Inra

How to build a federative and distributed e-infrastructure to support agri-food sciences ?

❖ Inra internal challenges:

- ✓ evolution of our digital infrastructure to provide « big data » and « FAIR data » services

- New services
- New skills



❖ Use external energies :

- ✓ Scientific communities, (i.e : plant sciences)
- ✓ Funders : H2020 : EOSC projects, Belmont forum
- ✓ Initiatives : GO FAIR, RDA



The role of RDA

What we do ...

What we use ...

Interest group “Agricultural Data” (IGAD)



Created in 2013, IGAD gather more than 200 people

Co-chairs : Patricia Bertin (Embrapa, Brésil),
Imma Subirats (FAO-ONU, int.)

web : <https://bit.ly/2SW4TL6>

A workshop before each RDA plenary

Published on
<https://f1000research.com/gateways/godan/okad?selectedDomain=slides>

Objectives:

- **To promote good practices** in the research domain: data sharing policies, data management plan, data interoperability
- **To provide a platform for networking** and cross-fertilization of research ideas in data management and interoperability
- **To solicit and promote interactions** and projects among the major international institutions and groups worldwide which work on agricultural research and innovation
- **To achieve data interoperability**

4 Working groups :

Wheat Data Interoperability, Rice Data Interoperability, AgriSemantics, On-Farm Data Sharing, Capacity Development for Agricultural Research Data



Wheat Data Interoperability WG

Co-chairs: Esther Dzalé (INRA, France), Richard Allan Fulss (CIMMYT, Mexico), Rosemary Shrestha (CIMMYT, Mexico)



Improve wheat data interoperability



created in 2014, WDI is today a maintenance Group

- Promote compliance of the data format with available software applications (preferably open).
- Facilitate recombination between sets of data from different sources.
- Promote common standards for metadata and vocabularies to facilitate the interpretation and linking of data across disciplines.
- Encourage mappings when unusual or project-specific formats or vocabularies are unavoidable.



→ recommendations continually updated

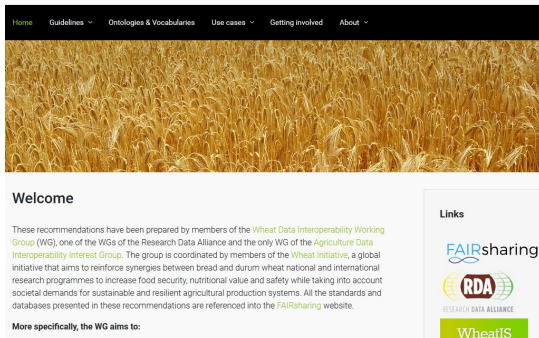


Wheat Data Interoperability WG Outputs

Guidelines : web portal

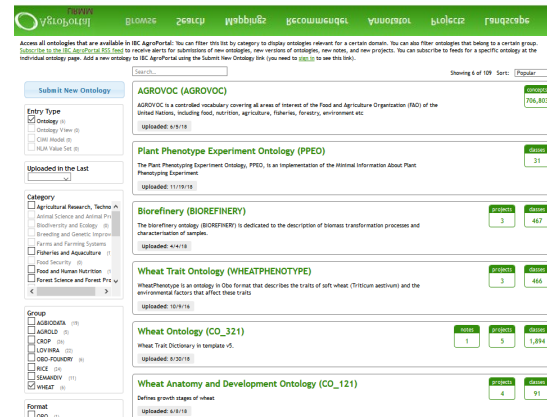
<http://wheatis.org/DataStandards.php>

Wheat Data Interoperability Guidelines



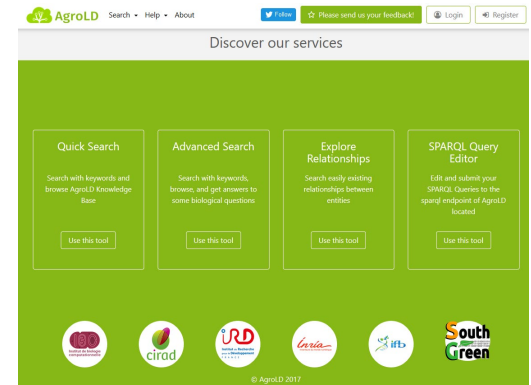
Ontologies collection in Agroportal

<http://wheat.agroportal.lirmm.fr/ontologies>



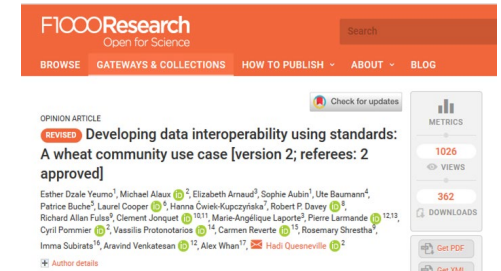
A prototype on AgroLD

<http://www.agrold.org>



An article

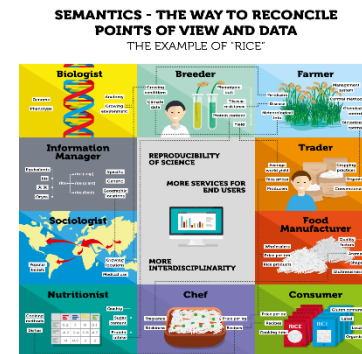
<https://doi.org/10.12688/f1000research.12234.2>



- ★ Recommended formats
- ★ Best practices to describe data
- ★ Best practices for data sharing
- ★ Tools
- ★ Catalog of relevant ontologies and vocabularies for wheat.
- ★ Give access to ontologies and vocabularies via REST API and SPARQL endpoint.
- ★ Tools to search, align, ...

The Agrisemantics WG

Jan. 2017 - Mar. 2019, ~100 members



Objective of the group : Envision the seamless use and creation of semantic resources supporting agricultural and food data findability and interoperability

- ❖ A report on **Semantics Landscape for Agricultural Data** (applications, research trends, resources, toolkits)
<http://bit.ly/AgSemLandscape>
- ❖ A set of **20 use cases and a list of community requirements** (access, reusability, tools and services for creation and management, use in applications, standards and best practices) <http://bit.ly/AgSemReqUC>
- ❖ A document on **Recommendations** to facilitate the uptake of semantics for agricultural data (version submitted to the RDA TAB for endorsement)
<http://bit.ly/AgSemRecom>

<https://www.rd-alliance.org/groups/agrisemantics-wg.html>

What we use : some examples

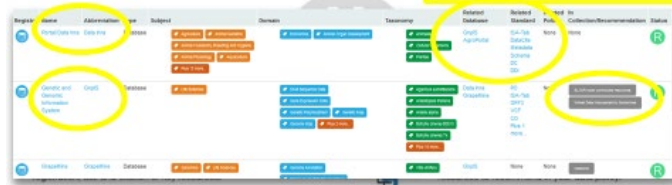


The FAIRsharing Registry and Recommendations:
Interlinking Standards, Databases and Data Policies

Researchers in academia, industry, government
Use FAIRsharing to identify and cite the standards, databases or repositories that exist for your discipline when creating a data management plan, releasing data or submitting a manuscript to a journal.



Developers and curators of resources
Make your standard, database or repository discoverable, by adding or claiming it in FAIRsharing. Increase exposure and credit outside of your immediate community and promote adoption.



How to analyze the technical ecosystem ?



Recommendations for Implementing a Virtual Layer for Management of the Complete Life Cycle of Scientific Data

January 2017

Edited by: Tobias Weigel, Peter Wittenburg.

Supported by: Bridget Almas, Reinhard Budich, Sandra Collins, Michael Diepenbrock, Ingrid Dillo, Françoise Genova, Frank Oliver Glöckner, Rebecca Grant, Wilco Hazeliger, Margareta Hellström, Keith Jefferey, Francisca de Jong, Tibor Kalman, Rebecca Koskela, Dimitris Koureas, Wolfgang Kuchinke, Leif Laaksonen, Larry Lannom, Michael Lautenschlager, Damien Lecarpentier, Jianhui Li, Jay Pearlman, Luca Pezzati, Ralph Müller-Pfefferkorn, Beth Piale, Stefano Nativi, Raphael Ritz, Ulrich Schwarndmann, Rainer Stotzka, Achim Streit, Dieter van Uytvanck, Anwar Vahed, Doris Wedlich, Colin Wright, Ramin Yahyapour, Thomas Zastrow, Carlo Maria Zucchi

[/www.rd-alliance.org/sites/default/files/recommendation-jan-2017-v8.pdf](http://www.rd-alliance.org/sites/default/files/recommendation-jan-2017-v8.pdf)



23 Things: Libraries for Research Data

An overview of practical, free, online resources and tools that you can begin using today to incorporate research data management into your practice of librarianship.

RDA 11th, Berlin, IGAD pre-meeting, 03/20/18

Horizon 2020 research and innovation programme - grant agreement No 730088

Research Data Sharing Without Barriers

Learning Resources

Librarians are learning how to apply the principles of library science to solve problems and to provide new services related to research data.

1. A "top ten" list of recommendations for libraries to get started with research data management from LIBER, <http://bit.ly/3qUvK63>
2. Relevant concepts are presented and mapped in the e-Science Thesaurus, <http://bit.ly/3LE04h8>
3. Understanding the life of research data with the DCC Curation Lifecycle Model, <http://bit.ly/3MoGGGv>

Learning Resources
Data Reference and Outreach
Data Management Plans
Data Literacy
Citing Data
Data Licensing and Privacy
Digital Preservation
Data Repositories
and a Community of Practice
...to help librarians engage in research data management!

Data Reference & Outreach

Librarians are answering questions about data from patrons and conducting outreach to assess the data needs of their researchers and students.

10. Questions about data answered by experts on the DataQ forum, <http://bit.ly/3MoH4Vg>

Data Management Plans

Librarians are becoming familiar with funder requirements and consulting with researchers to help them write and implement effective data management plans.

11. One example is the DMPTool that lists funder requirements in the United States and builds a plan by asking the researcher to answer a series of questions. Other countries such as the U.K. and Canada have similar tools, <http://bit.ly/3LuNZMH>



The FAIRsharing Registry and Recommendations: Interlinking Standards, Databases and Data Policies

Researchers in academia, industry, government
Use FAIRsharing to identify and cite the standards, databases or repositories that exist for your discipline when creating a data management plan, releasing data or submitting a manuscript to a journal.



Developers and curators of resources
Make your standard, database or repository discoverable, by adding or claiming it in FAIRsharing; increase exposure and credit outside of your immediate community and promote adoption.

Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy	Related In Collection/Recommendation	Status
	Portail Data Inra	Data Inra	Database	Agriculture Animal Genetics Animal Husbandry, Breeding And Hygiene Animal Physiology Aquaculture Plus 12 more...	Economics Animal Organ Development	Animals Cellular organisms Plantae	GnplS AgroPortal	ISA-Tab DataCite Metadata Schema DC DDI	None	None	
	Genetic and Genomic Information System	GnplS	Database	Life Sciences	DNA Sequence Data Gene Expression Data Genetic Polymorphism Genetic Map Genome Map Plus 2 more...	Agaricus subrefescens Arabidopsis thaliana Arabis alpina Botrytis cinerea B0510 Botrytis cinerea T4 Plus 15 more...	Data Inra GrapeMine	PO ISA-Tab GFF3 VCF CO Plus 1 more...	None	 	
	GrapeMine	GrapeMine	Database	Genomics Life Sciences	Genome Annotation	Vitis vinifera	GnplS	None	None		

How to analyze a technical ecosystem ?



Recommendations for Implementing a Virtual Layer for Management of the Complete Life Cycle of Scientific Data

January 2017

Edited by: Tobias Weigel, Peter Wittenburg,

Supported by: Bridget Almas, Reinhard Budich, Sandra Collins, Michael Diepenbrook, Ingrid Dillo, Françoise Genova, Frank Oliver Glöckner, Rebecca Grant, Wilco Hazeleger, Margareta Hellström, Keith Jefferey, Franciska de Jong, Tibor Kalman, Rebecca Koskela, Dimitris Koureas, Wolfgang Kuchinke, Leif Laaksonen, Larry Lannom, Michael Lautenschlager, Damien Lecarpentier, Jianhui Li, Jay Pearlman, Luca Pezzati, Ralph Müller-Pfefferkorn, Beth Plale, Stefano Nativi, Raphael Ritz, Ulrich Schwarzmann, Rainer Stotzka, Achim Streit, Dieter van Uytvanck, Anwar Vahed, Doris Wedlich, Colin Wright, Ramin Yahyapour, Thomas Zastrow, Carlo Maria Zwölf

<https://eoscpiot.eu/content/erosa-position-paper>

Johannes Keizer, e-ROSA project

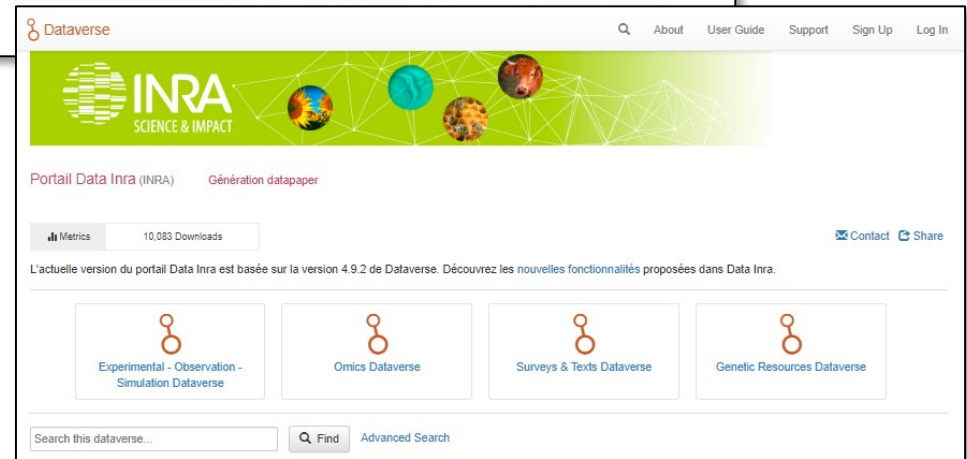
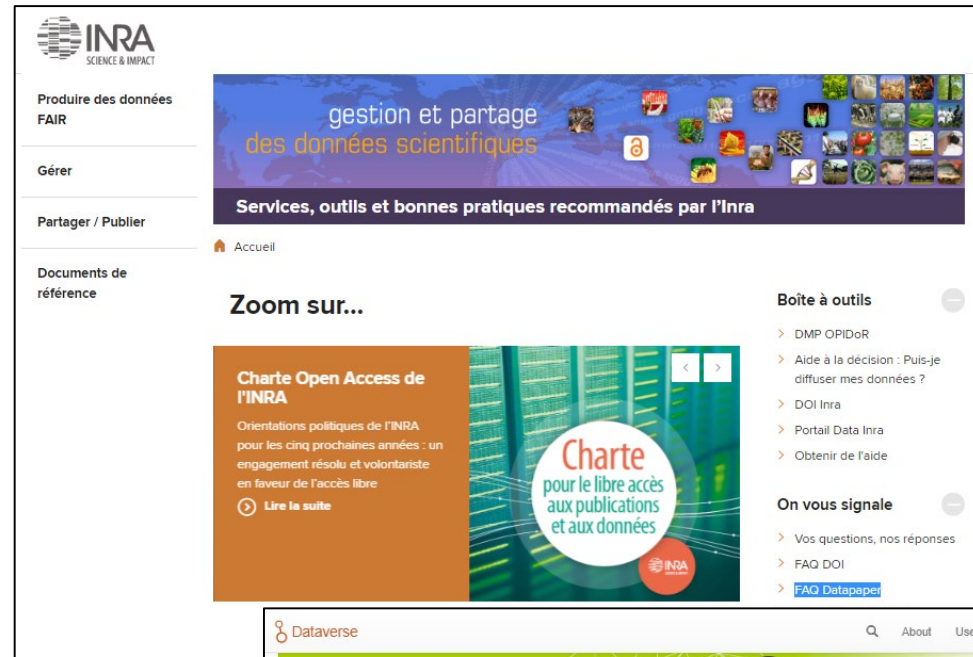
Item	Status and needed action
Network of trustworthy repositories	There are scattered repositories all over the landscape which need to be organized in the network
A registry of trustworthy repositories	CIARD RING as prototype is powerful tool to organize the network of repositories
Repository API/PID Registration	No domain specific activities should be undertaken, collaboration should be seared with other EOSC partners
Metadata Schemes registration	Work is very advanced with the former VEST registry. Coordination and discussion with other similar enterprises in other domains is needed
Concept server and registration system	The domain is comparatively advanced having available operative prototypes of Ontology Servers (Agroportal) and Concept Servers (GACS)
System of Authorisation records	No domain specific activities should be undertaken, collaboration should be seared with other EOSC partners
System of licence registries	No domain specific activities should be undertaken, collaboration should be seared with other EOSC partners
Ecosystem of tools and operating procedures	Some scattered work has been done but without systematization; it needs to be integrated with the effort to create competence centers

New Library services inspired by 23 things ...

- ❖ Information, training
- ❖ Services for FAIR data : DMP, DOI, ontologies
- ❖ Data portal based on dataverse

<http://datapartage.inra.fr>

<http://data.inra.fr>



The whole story

Data standards for Agriculture: INRA adoption of RDA Outputs

15
JAN
2019

RDA
ADOPTION
STORY

On how RDA has
contributed to
many aspects of
INRA's Open
Science policy



RDA

Image: no one cares at Unsplash

On how RDA has contributed to many aspects of INRA's Open Science policy

Next


M
Lec
31 July 20
National
2019

RDA
Ingr
Soc
for

Date

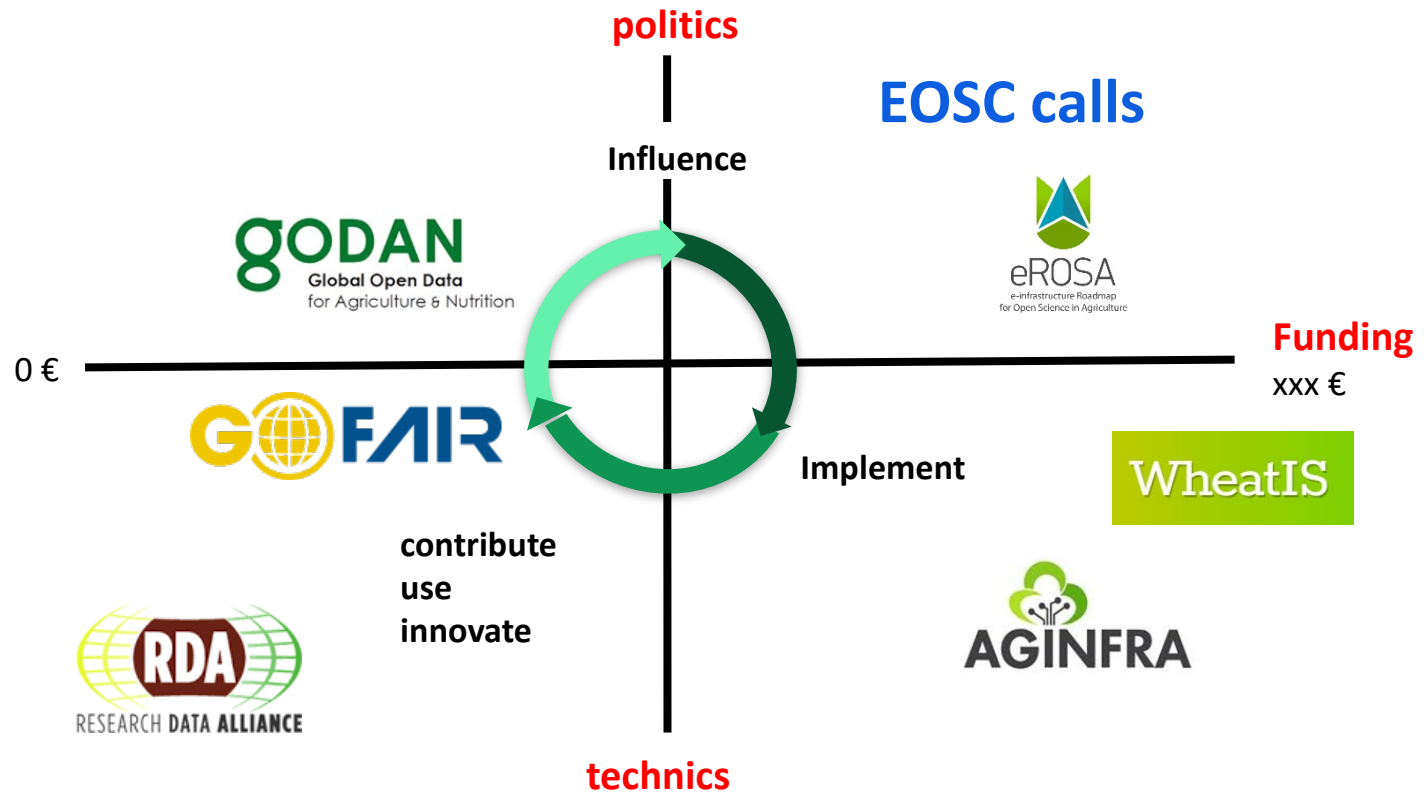
Vent
Build

<https://www.rd-alliance.org/rda-adoption-story-inra-france>



Conclusion

RDA and the others ...



Conclusion

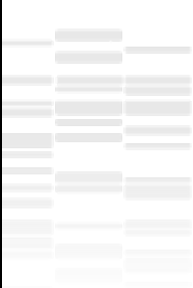
❖ RDA is a win-win environment, at different scales :

✓ INRA perspective :

- Institution : Useful results supporting our e-infra strategy, specially in wheat research
- Support team (library) : Increased the legitimacy to work with researchers
- Individuals : results recognition

✓ RDA perspective :

- Success stories easy to explain : feed 9 billion people ...
- An active community interacting with others – but also a challenge to keep these interactions



Merci pour votre attention...

odile.hologne@inra.fr

@Holo_08

