



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



➤ 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.

Agriculture



Environment



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



An overview of INRA – November 2017



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



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

INRA strategic framework (oct 2016)

THREE GENERAL POLICY DIRECTIONS

[#OpenScience]

Opening science with digital tools



Document

#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-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

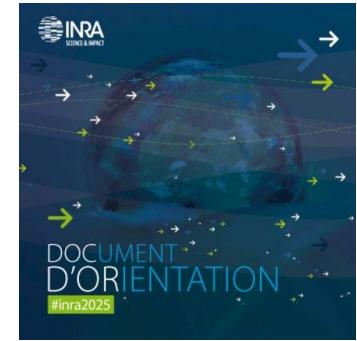
#Support-1: An efficient, agile and resilient organisation

#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



[#OpenInra]

A nation-wide actor in innovation

[#Support]

Anticipating and supporting change

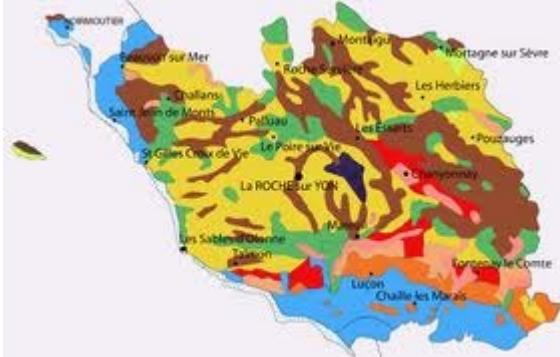
Agenda

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



Data challenge in agri-food sciences

Data in Agri-food sciences ?

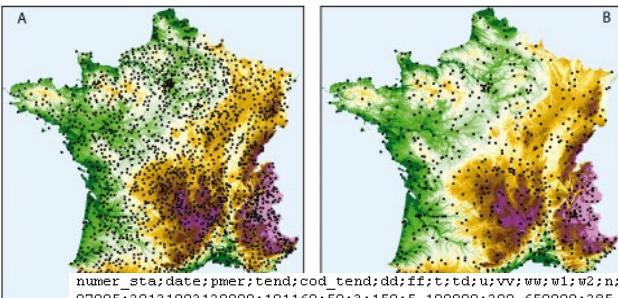


```

ACATATGACAGGGGGGGGTAGACA-
TTTTTTTTTTTTTTTTTTTTTTTTC
AAAAAAAAAAAAAAAAGA
ACCCCCCCCCTTTCATATTACCCCCCA
CTCTCAGGGTGTGCGGGGTGTTTT
TTTAGACCCCCCCCCCCCCCCCCCCCCC
CGGGGGTGTTTTAAAAGGGGGGGGGG
TTTAGACCCCCAGATTTACACAGTACAGG
ATAGATAACCCAGATATAGAGAGACCCATAGAG
CACCCCAACCCCCCTTTCATATTACCCCCCA
ATAGATAACCCAGATATAGAGAGACCCATAGAG
ACCCCCAACCCCCCTTTCATATTACCCCCCA
GATAACCCAGATATAGAGAGACCCATA

```

ASPECT ENVIRONNEMENTAL			
L'environnement matériel			
01.Je suis satisfait(e) de l'état général du site où je travaille.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
02.Je cotoie régulièrement des animaux qui dérangent mes travaux.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
03.Je suis satisfait(e) de l'apurement de mes postes de travail.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
04.Je suis satisfait(e) des zones de détente mises à mon disposition (bâches, cabines...).	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
L'environnement relationnel			
05.Je trouve que les relations établies au sein du site sont cordiales.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
06.Les relations dans mon unité ou mon service sont cordiales.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
07.Je le considère d'être respecté par les personnes qui je côtoie au sein du site.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
08.Je le considère d'être respecté par les personnes dont je côtoie au sein du service les autres.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
La sécurité et la santé au travail			
09.Je suis satisfait(e) de la sécurité au travail.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.Je suis satisfait(e) de la santé au travail.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



```

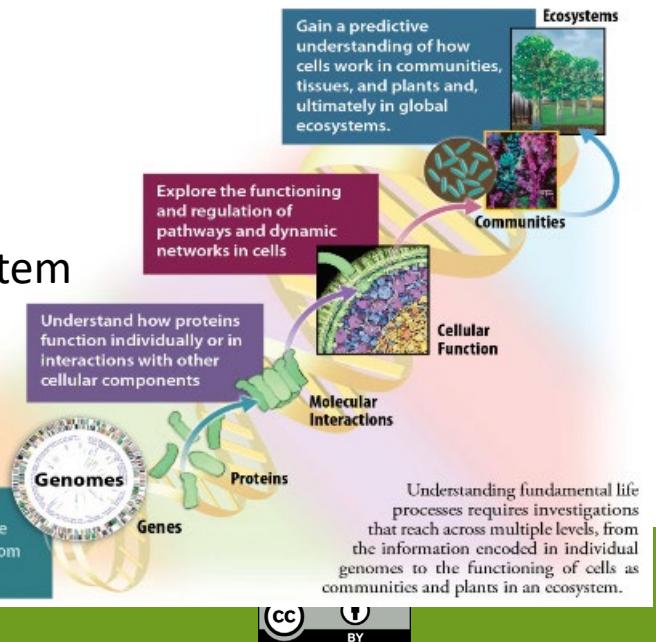
numer_sta;date;pmer;tend;cod_tend;dd;ifft;td;uvv;ww;wi;eo;n
070005;20131003120000;101160;50;3;150;5.100000;290.650000;285.
070075;20131003120000;101150;-140;6;130;6.200000;293.750000;28
07020;20131003120000;100470; -110;5;170;11.800000;293.950000;2
Agrandir
07027;20131003120000;100670; -60;7;160;5.700000;294.250000;291.
Source : 07037;20131003120000;100980; -50;8;100;5.100000;289.550000;286
07072;20131003120000;101310; -50;8;140;5.100000;293.250000;286
07110;20131003120000;100300;90;1;190;8.700000;293.550000;289.
07117;20131003120000;100320;120;1;200;6.700000;295.450000;289
07130;20131003120000;100620;50;1;180;8.700000;296.350000;291.
07139;20131003120000;100860; -50;5;170;6.200000;294.250000;290
07149;20131003120000;101120;30;0;130;2.100000;294.650000;290.
07168;20131003120000;101230;10;1;170;6.200000;296.050000;289.
07181;20131003120000;101490;mrg;mrg;100;4.600000;291.650000;284
07190;20131003120000;101880; -70;8;40;2.600000;287.950000;278.
07207;20131003120000;100550;40;0;160;7.200000;294.350000;291.

```

Different families :

- Omics
- Observation
- Social sciences, cohort
- From genome to ecosystem

The genome determines dynamic biological structure and function at all scales, from genes to ecosystems

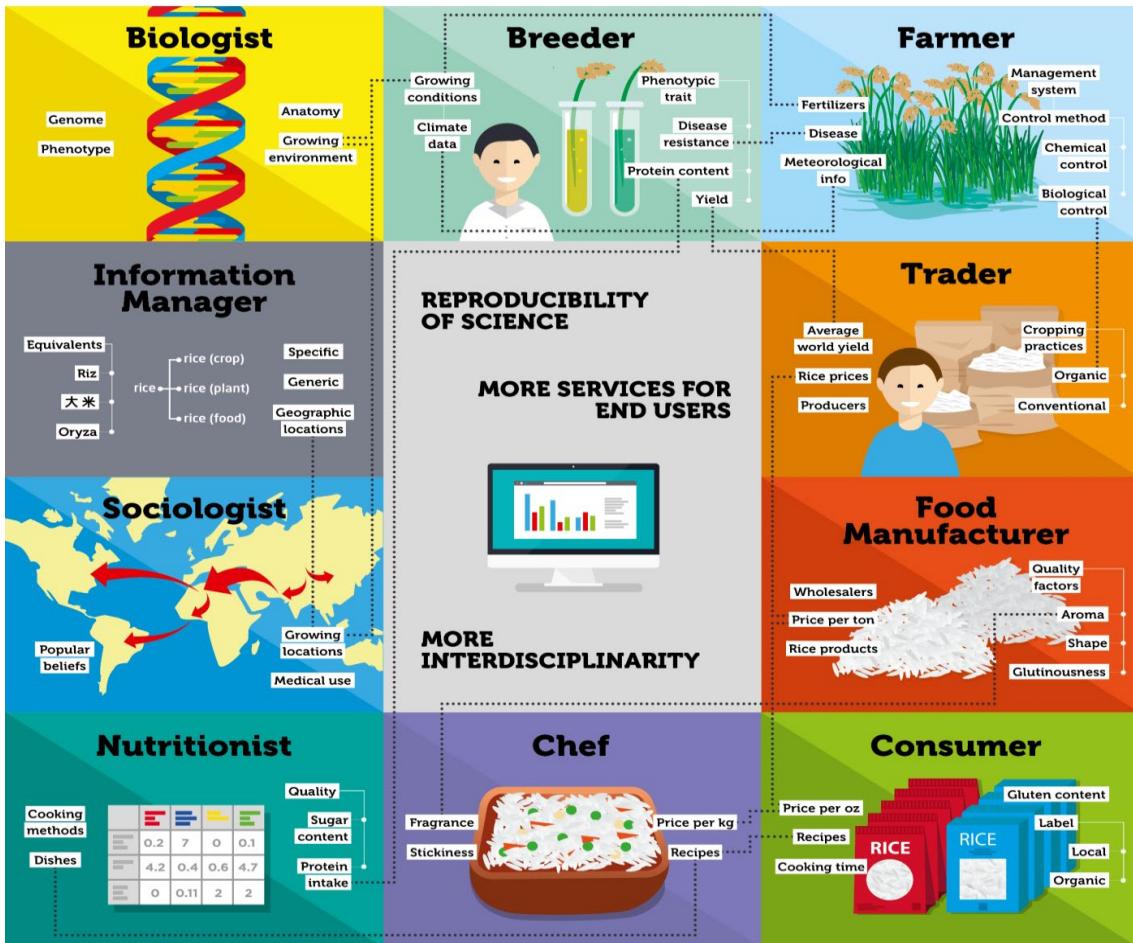


Data challenges in agri-food sciences

- ❖ 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)

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



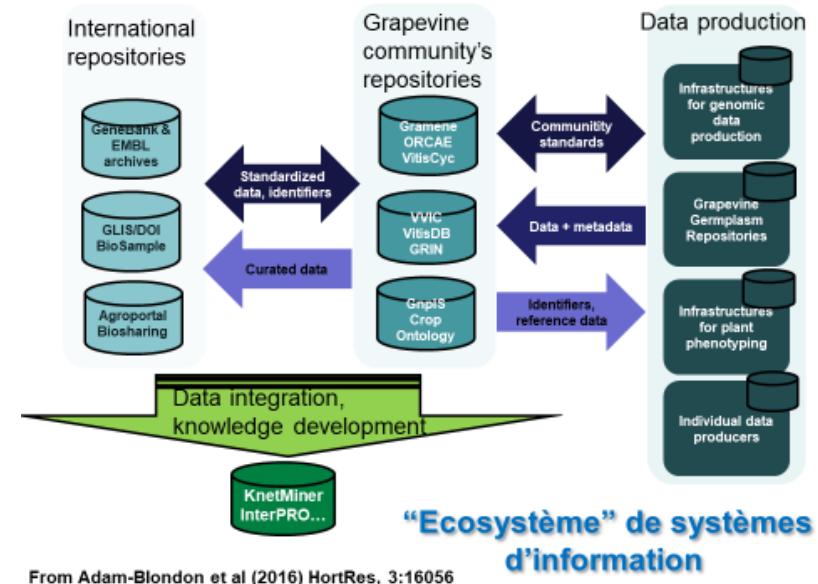
=> e-infrastructure

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)



Food and Agriculture Organization
of the United Nations



created in 2014, WDI is today a maintenance Group



→ recommandations continually updated



Improve wheat data interoperability

- 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.



Wheat Data Interoperability WG Outputs

Guidelines : web portal

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

Wheat Data Interoperability Guidelines



Welcome

These recommendations have been prepared by members of the [Wheat Data Interoperability Working Group \(WG\)](#), one of the WGs of the Research Data Alliance and the only WG of the Agriculture Data Interoperability Interest Group. The group is coordinated by members of the Wheat Initiative, a global initiative that aims to support the development of data standards and international research programmes to increase food security, nutritional value and safety while taking into account societal demands for sustainable and resilient agricultural production systems. All the standards and databases presented in these recommendations are referenced into the [FAIRsharing](#) website.

More specifically, the WG aims to:

- ★ Recommended formats
- ★ Best practices to describe data
- ★ Best practices for data sharing
- ★ Tools

Ontologies collection in Agroportal

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

The screenshot shows a search interface for ontologies. The search bar is empty. Below it, a table lists several ontologies:

Ontology	Concepts
AGROVOC (AGROVOC)	706,803
Plant Phenotype Experiment Ontology (PPEO)	31
Biorefinery (BIOREFINERY)	3 467
Wheat Trait Ontology (WHEATPHENOTYPE)	3 466
Wheat Ontology (CO_321)	1 5 1,894
Wheat Anatomy and Development Ontology (CO_121)	4 91

Each row includes a 'Search' button and an 'Upload' link.

- ★ Catalog of relevant ontologies and vocabularies for wheat.
- ★ Give access to ontologies and vocabularies via REST API and SPARQL endpoint.
- ★ Tools to search, align, ...

A prototype on AgroLD

<http://www.agrold.org>

The screenshot shows the AgroLD homepage with a green header. It features four main search tools:

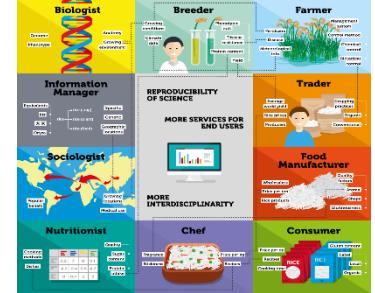
- Quick Search**: Search with keywords and browse AgroLD Knowledge base. Includes a 'Use this tool' button.
- Advanced Search**: Search with keywords, browse, and get answers to some biological questions. Includes a 'Use this tool' button.
- Explore Relationships**: Search easily existing relationships between entities. Includes a 'Use this tool' button.
- SPARQL Query Editor**: Edit and submit your SPARQL Queries to the sparql endpoint of AgroLD located. Includes a 'Use this tool' button.

Logos for various partners like INRA, Cirad, URD, Iriis, and South Green are at the bottom.

An article

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

The screenshot shows the F1000Research article page. The title is "Developing data interoperability using standards: A wheat community use case [version 2; referees: 2 approved]". The authors listed are Elizabeth Arnould¹, Sophie Alauzen¹, Jérôme Baumard⁴, Patricia Baudet², Laurent Cogné³, Bertrand Deville-Kespriano³, Robert Deneuve⁵, Richard Allen Fulte², Clément Janaget³, Marie-Angélique Laporte², Pierre Lemande^{12,13}, Cyril Pommier¹², Vassilis Protoporfiris¹², Carmen Reverte¹², Rosemary Shrestha¹, Imma Subira¹⁶, Aravind Venkatesan¹², Alex Whan¹⁷, Hadi Quesneville¹². The page includes metrics: 1026 views, 362 downloads, and links to PDF and XML versions.



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.



RESEARCH DATA ALLIANCE
UNITED STATES

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.

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/1qUvKG3>
2. Relevant concepts are presented and mapped in the e-Science Thesaurus, <http://bit.ly/1LEo4h8>
3. Understanding the life of research data with the DCC Curation Lifecycle Model, <http://bit.ly/1MoGGGv>

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!

10. Questions about data answered by experts on the DataQ forum, <http://bit.ly/1MoH4Vg>

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 DMTool 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/1LuNZMH>

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 Wittenberg,
Supported by: Bridget Almas, Reinhard Busch, Sandra Collins, Michael Dispensbrook, Ingrid Dilts,
Françoise Genova, Frank Oliver Glöckner, Rebecca Grant, Wilco Hazelaeger, Margareta Hellström,
Keith Jefferey, Franciska de Jong, Tibor Kalman, Rebecca Koskela, Dimitris Kourreas, Wolfgang
Kuchinke, Leif Leakkonen, Larry Lannom, Michael Lautenschlager, Damien Leparcopier, Jianhui Li, Jay
Pearlman, Luca Pezzati, Ralph Müller-Pfefferkorn, Bett Plale, Stefano Nativi, Raphael Ritz, Ulrich
Schwardmann, Rainer Stotzka, Achim Streit, Dieter von Uytvank, Anwar Vahed, Doris Wedlich, Colin
Wright, Ramin Yahyapour, Thomas Zastrow, Carlo Maria Zwilf

<http://www.rd-alliance.org/sites/default/files/recommendation-jan-2017-v8.pdf>

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

Horizon 2020 research and innovation programme - grant agreement No 73088

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	In Policy	In Collection/Recommendation	Status
	Portal Data Inra	Data Inra	Database	<input checked="" type="checkbox"/> Agriculture <input checked="" type="checkbox"/> Animal Genetics <input checked="" type="checkbox"/> Animal Husbandry, Breeding And Hygiene <input checked="" type="checkbox"/> Economics <input checked="" type="checkbox"/> Animal Physiology <input checked="" type="checkbox"/> Animal Organ Development <input checked="" type="checkbox"/> Aquaculture <input checked="" type="checkbox"/> Plus 12 more...	<input checked="" type="checkbox"/> Agriculture <input checked="" type="checkbox"/> Animal Genetics <input checked="" type="checkbox"/> Animal Husbandry, Breeding And Hygiene <input checked="" type="checkbox"/> Economics <input checked="" type="checkbox"/> Animal Physiology <input checked="" type="checkbox"/> Animal Organ Development <input checked="" type="checkbox"/> Aquaculture <input checked="" type="checkbox"/> Plus 12 more...	<input checked="" type="checkbox"/> Animals <input checked="" type="checkbox"/> Cellular Components <input checked="" type="checkbox"/> Plantae	GnplS AgroPortal	ISA-Tab DataCite Metadata Schema DC DDI	None	None	
	Genetic and Genomic Information System	GnplS	Database	<input checked="" type="checkbox"/> Life Sciences	<input checked="" type="checkbox"/> DNA Sequence Data <input checked="" type="checkbox"/> Gene Expression Data <input checked="" type="checkbox"/> Genetic Polymorphism <input checked="" type="checkbox"/> Genetic Map <input checked="" type="checkbox"/> Genome Map <input checked="" type="checkbox"/> Plus 2 more...	<input checked="" type="checkbox"/> Agaricus subrufescens <input checked="" type="checkbox"/> Arabidopsis thaliana <input checked="" type="checkbox"/> Araols alpina <input checked="" type="checkbox"/> Botrytis cinerea B0510 <input checked="" type="checkbox"/> Botrytis cinerea T4 <input checked="" type="checkbox"/> Plus 15 more...	Data Inra GrapeMine	PO ISA-Tab GFF3 VCF CO Plus 1 more...	None	ELIXIR node contributed resources Wheat Data Interoperability Guidelines	
	GrapeMine	GrapeMine	Database	<input checked="" type="checkbox"/> Genomics <input checked="" type="checkbox"/> Life Sciences	<input checked="" type="checkbox"/> Genome Annotation	<input checked="" type="checkbox"/> Vitis vinifera	GnplS	None	None	Incentive	

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 Diepenbroek, Ingrid Dillo, Francoise 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 Schwardmann, Rainer Stotzka, Achim Streit, Dieter van Uytvanck, Anwar Vahed, Doris Wedlich, Colin Wright, Ramin Yahyapour, Thomas Zastrow, Carlo Maria Zwölff

<https://eoscipilot.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

The top screenshot shows the INRA Datapartage website (<http://datapartage.inra.fr>). The bottom screenshot shows the INRA Portail Data Inra (<http://data.inra.fr>) based on Dataverse.

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

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

Ne

RDA

RDA
Ingl
Soc
for

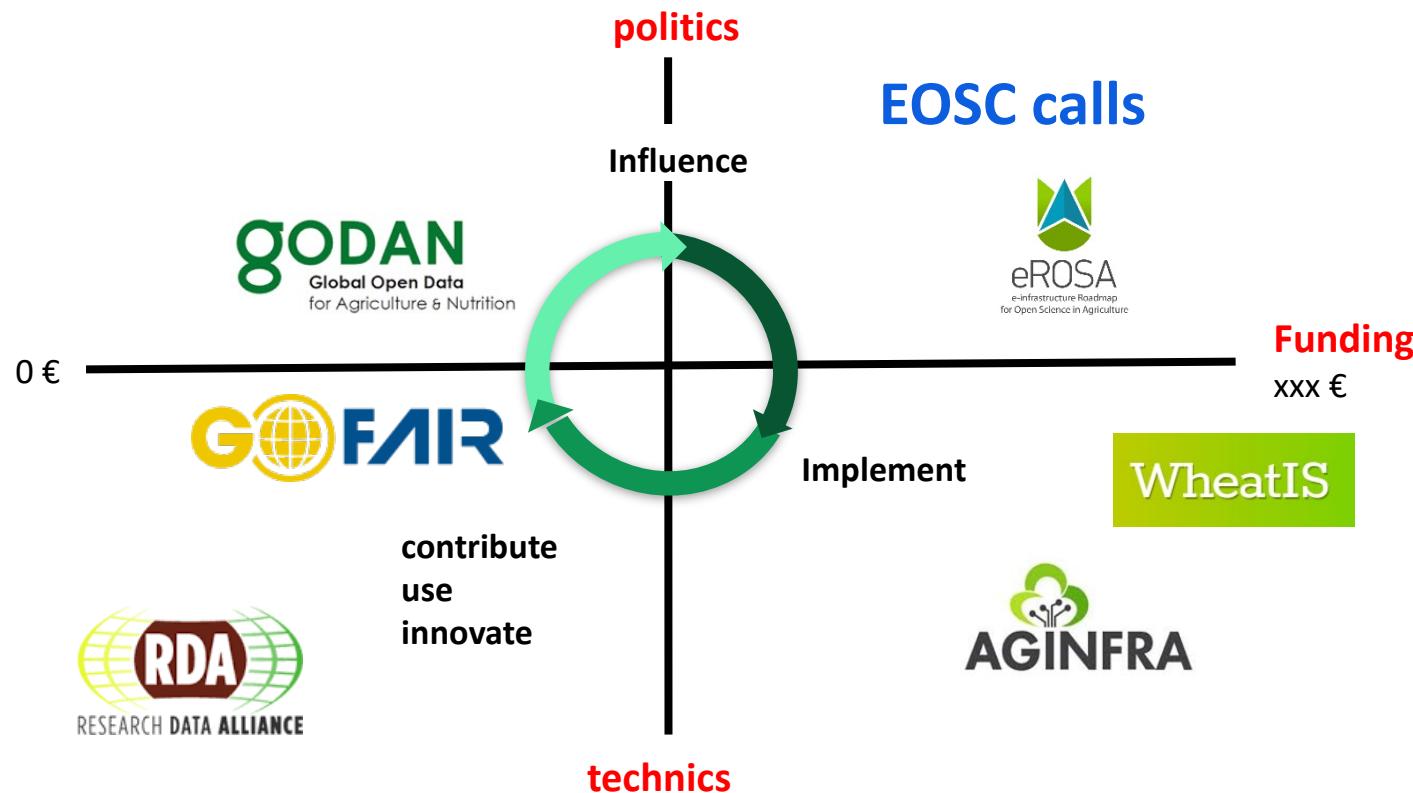
Date

Vent
Build



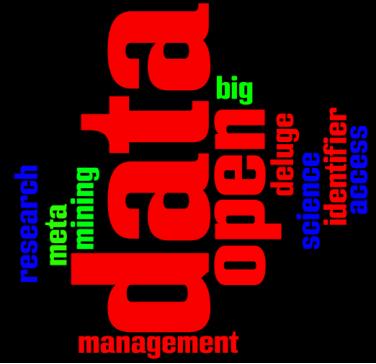
Conclusion

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.holgne@inra.fr

@Holo_08



twitter

