



# Rice Data Interoperability Working Group Updates

Pierre Larmande, IRD  
Shaik Meera, IIRR  
Ramil Mauleon, IRRI

RDA11 Plenary March 2018  
In collaboration with GODAN

---

# Summary of the Problem

- 
- Problem Statement
  - Impact to (various communities) of the unresolved problem

# Rice Data Interoperability: WG Deliverables

---

1. A survey/report on the existing resources (vocabularies, ontologies, data formats, metadata standards).
2. A cookbook intended for the Rice data managers community: guidelines on metadata, vocabularies and ontologies plus a decision tree based on data and metadata description recommendations and file format recommendations.
3. Repository of controlled vocabularies and ontologies compliant with the Linked Data standards
4. A prototype of Rice-specific data registry
5. Good practices/method(s) for digitization of rice legacy data

# WG Deliverables - Status

1. A survey/report on the existing resources (vocabularies, ontologies, data formats, metadata standards)
- 

A report on the survey of existing standards among rice research and development organizations. Focus on data availability, accessibility and applicability, formats, ontology, standards and metadata used. A complete analysis of interoperability (or otherwise of) among rice databases and repositories.

2. A cookbook intended for the Rice data managers community: guidelines on metadata, vocabularies and ontologies plus a decision tree based on data and metadata description recommendations and file format recommendations.

Synthesis is being done from Wheat WG

A set of recommendations on good practices, ontologies, tools and examples to create, manage and share data related to Rice. This work will be based on the existing Wheat Data WG Guidelines , The WG will Identify and adopt those relevant to rice data, and will customize accordingly. New types of data might be added according to the results of point 1. The expected output is a Rice Data Framework specification (cookbook)

# WG Deliverables - Status

## 3. Repository of controlled vocabularies and ontologies compliant with the Linked Data standard

---

Recommendations for a Rice ontology which should align existing rice ontologies, thesauri, controlled vocabularies. This should be the basis for a prospect on multi-lingual conversion of ontologies (TH KU/JP NARO/IRRI/ IIRR / Bioversity) which will not be covered by this WG as a deliverable.

## 4. A prototype of Rice-specific data registry

Draft prepared – 6 months

## 5. Good practices/method(s) for digitization of rice legacy data

Draft ready – 6 months

# Initial Adopters

- 
- IRRI
  - IRD
  - Bioversity
  - NARO
  - IIRR
  - CIRAD
  - CIAT
  - FAO of the UN
  - INRA
  - PhilRice
  - AfricaRice.. Hopefully many rice research roganizations

# Expected Impact of the Deliverable

---

- *Understanding ongoing initiatives*
- *Create a prototype data registry for test in line with IRRI's ongoing work*
- *Collect semantics and initiate a framework for a Rice ontology*
- *Best practices for digitization of rice legacy data based on Indian and Thai experiences*
- *Encouraging rice based organizations to adopt RDI standards*

# Feedback Desired from RDA Community

---

- Need an extension of 6 months to 1 year
- Sensitizing the organizations takes time



# Detailed Updates

---

# RDI activities / deliverables updates

## Work on the cookbook

- Intent is to pattern after Wheat WG

## Evaluation of a prototype on Rice specific data registry

- Will look at result of survey to identify other data registries in use by community
- Dataverse system (<https://dataverse.org/>), has been evaluated due to it's usage in CGIAR centers (IRRI, CIMMYT, ICRISAT)
- Also the COPO platform by the Earlham Institute (<https://copo-project.org/>)
- other CGIAR examples on data management systems and approaches such as
  - an ontology-based agronomy fieldbook (Breeding4Rice)
  - a genotypic data management system (the GOBII genotyping platform (IRRI, CIMMYT and ICRISAT) and
  - RHoMIS, the Rural Household Multi-Indicator Survey database (ILRI).

# RDI activities / deliverables updates

- Open data , open access and the FAIR (Findable, accessible, interoperable, reproducible - <https://www.force11.org/group/fairgroup/fairprinciples> ) principles are strongly advocated for adherence of public data registries, including rice data registries.
- Some concepts to consider:
  - pre-open access data sharing - 100% compliance OADM policy;
  - ontologies (Bioversity);
  - DACE repository to communications workflow tool (CIFOR), and
  - using DOIs vs URIs (IITA + IFPRI).
- Galaxy as data analysis method registry using RDA outputs of Persistent Identifier Information Types (PIT) and Data Type Registries (DTR) Working Groups and software services at Advanced Institute of Science and Technology Japan (in collaboration with the Pacific Rim Applications and Grid Middleware Architecture – PRAGMA, <http://www.pragma-grid.net/> )

# RDI activities / deliverables updates

## Recommendations for ontologies appropriate for managing rice data

- RDI working closely with Ontology Working Group of CGIAR Big Data Platform
- Rice Slice of ontology dedicated in Agroportal

## Good practices/method(s) for digitization of rice legacy data

- Bringing legacy data – into the RDI fold- Guidelines developed

# Legacy data – Process documentation

## How to organize and manage legacy data?

- Collecting Discipline wise data sets
- Digitalize the data sets
- Define keywords and phrases in each data set
- Tagging data set with keywords, Unique title phrase, discipline, season and year
- Populating these data sets in repositories with FAIR principles

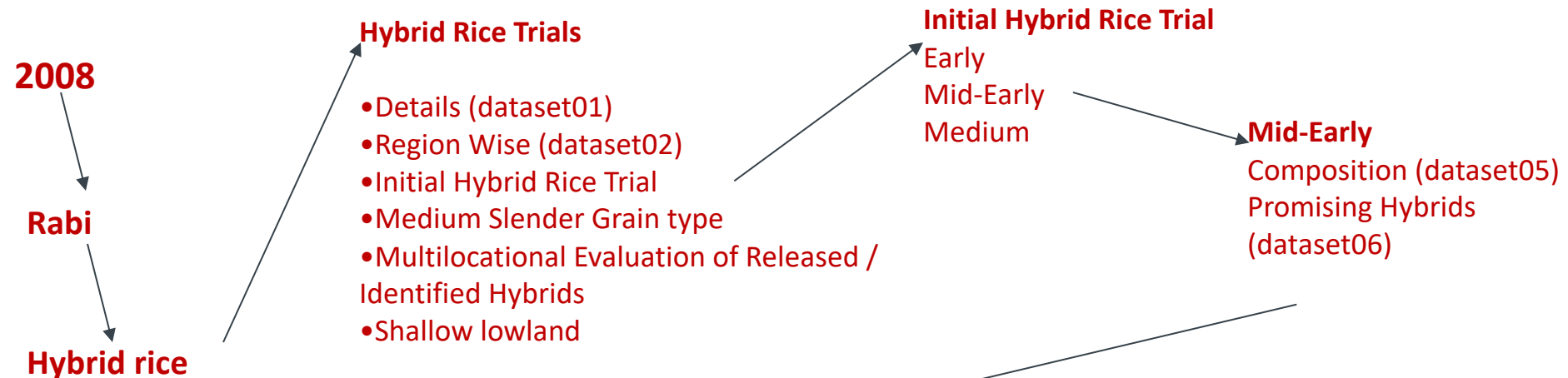
# Legacy data – Process documentation

Promising hybrids of Mid early in Initial Hybrid Rice trials, Dry Season-1989

Search for  Year  Discipline

Season  

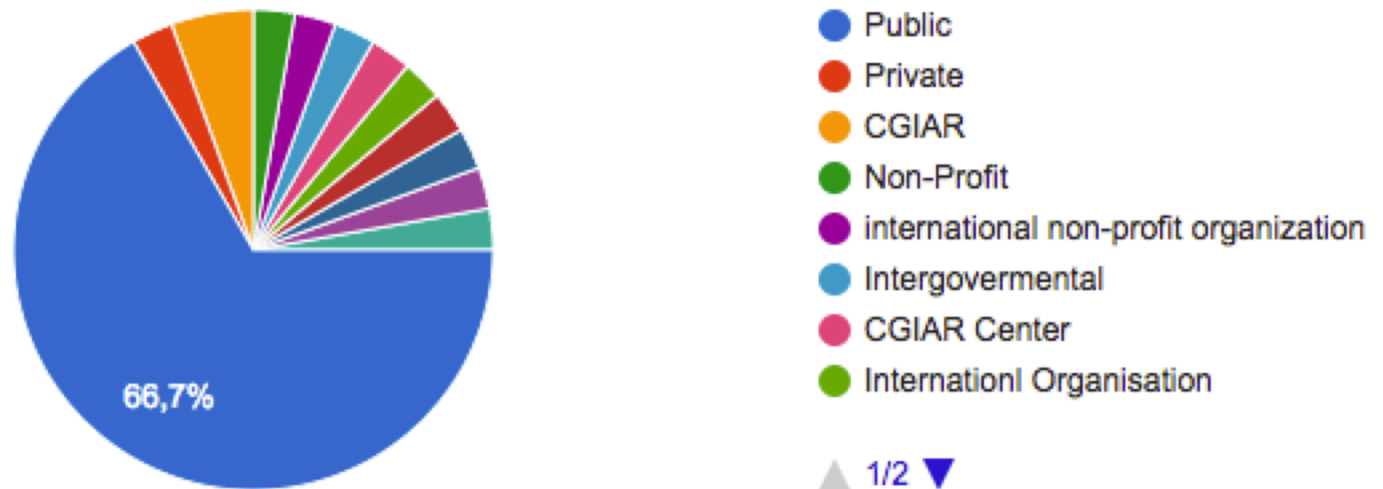
Hybrid Rice  
 SELECT  
 Agronomy  
 Entomology  
 Hybrid Rice  
 Soil Science  
 Physiology  
 Pathology  
 Plant Breeding  
 Bio-Technology  
 Production Oriented Survey  
 FLDs



# Rice Data Interoperability Survey output

n= 37

## Type of Organizations



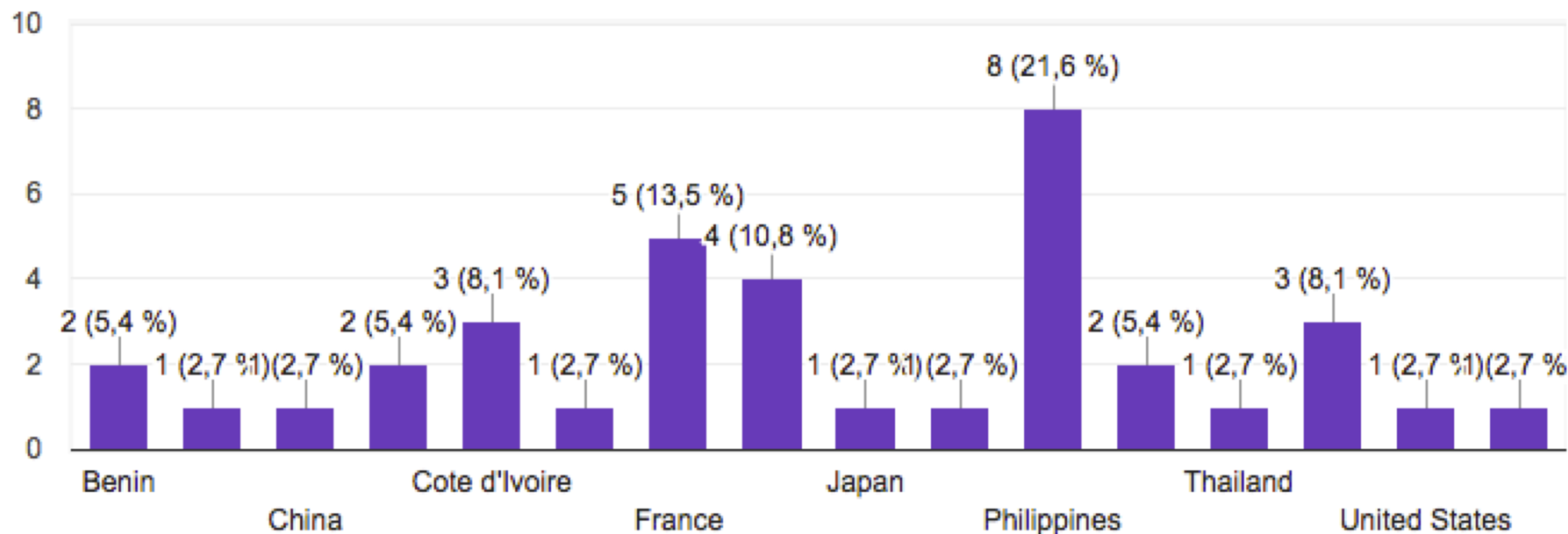
# Rice Data Interoperability Survey output

## n= 37

### Country

#### Country

37 réponses





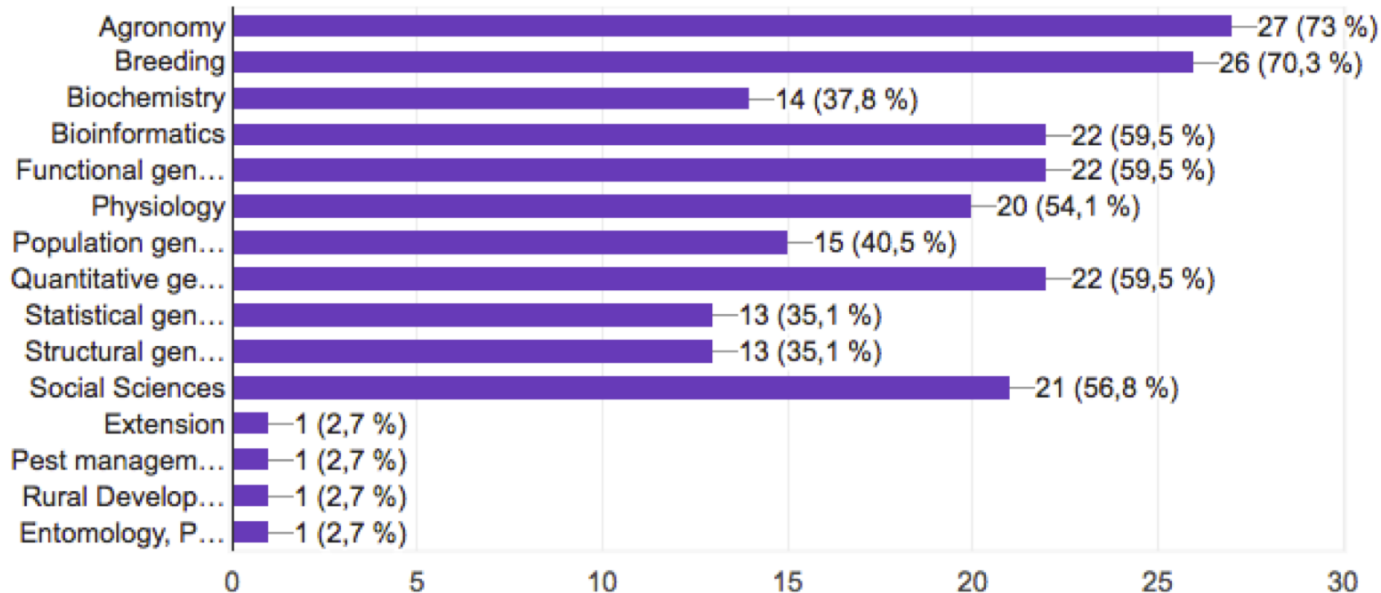
# Rice Data Interoperability Survey output

## n= 37

What kind of rice research data is generated and shared by your organization?



37 réponses



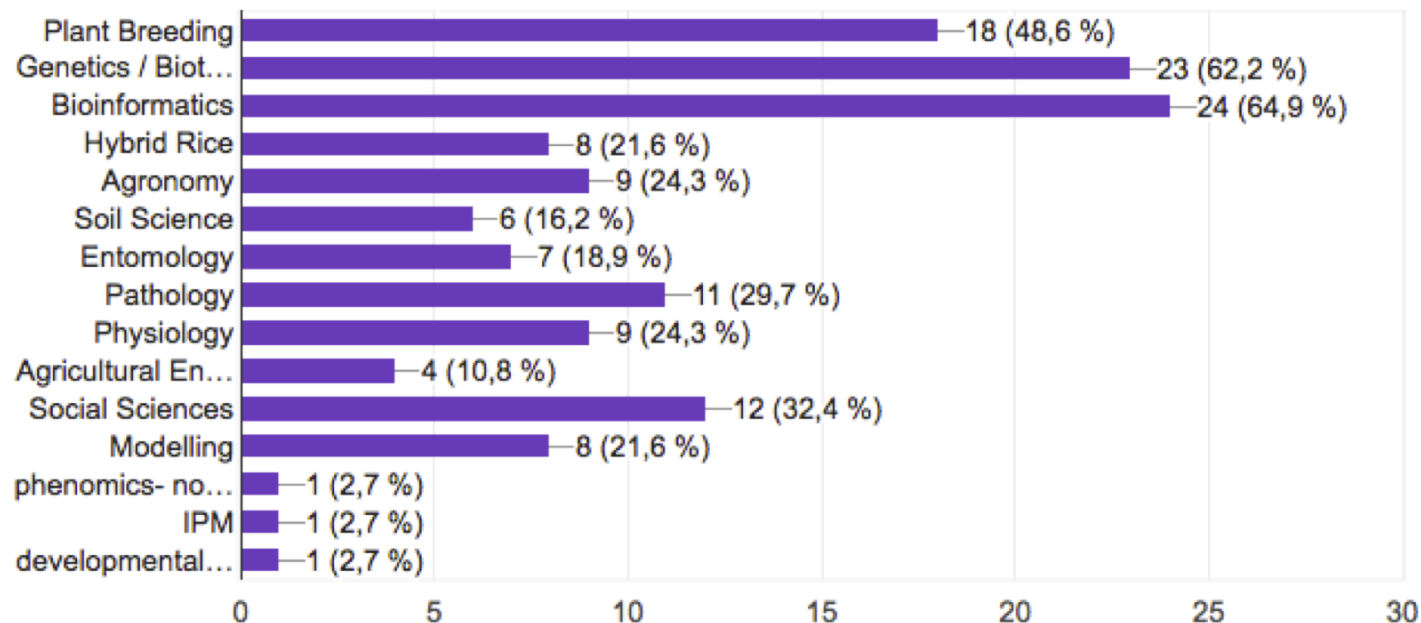
# Rice Data Interoperability Survey output

## n= 37

Within Rice research you are interested in what kind of data (discipline-wise)?



37 réponses

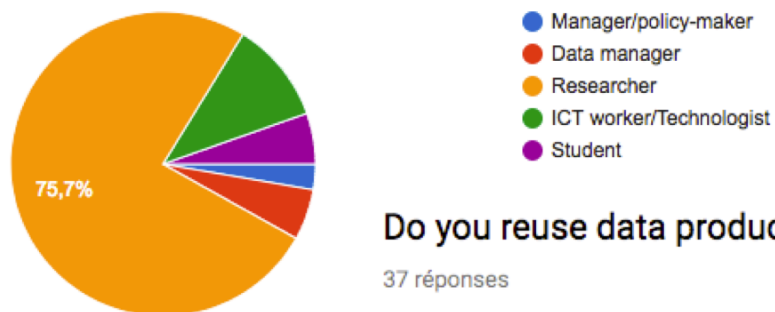


# Rice Data Interoperability Survey output

## n= 37

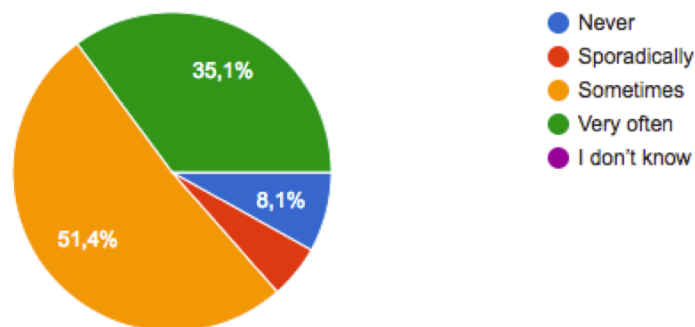
What is your current position/role?

37 réponses



Do you reuse data produced by other organizations?

37 réponses

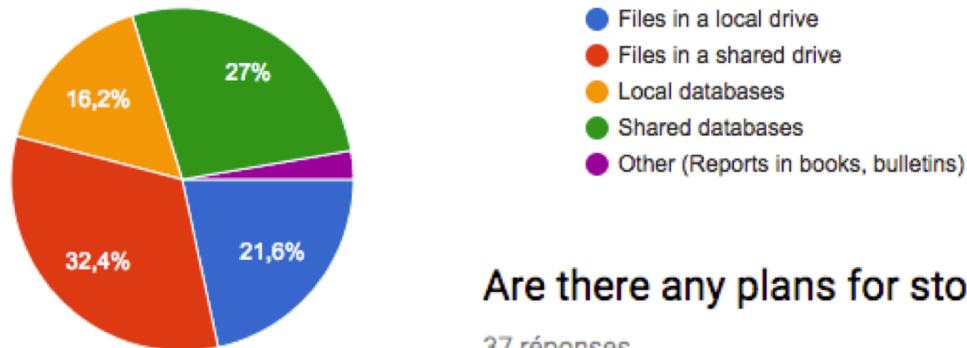


# Rice Data Interoperability Survey output

## n= 37

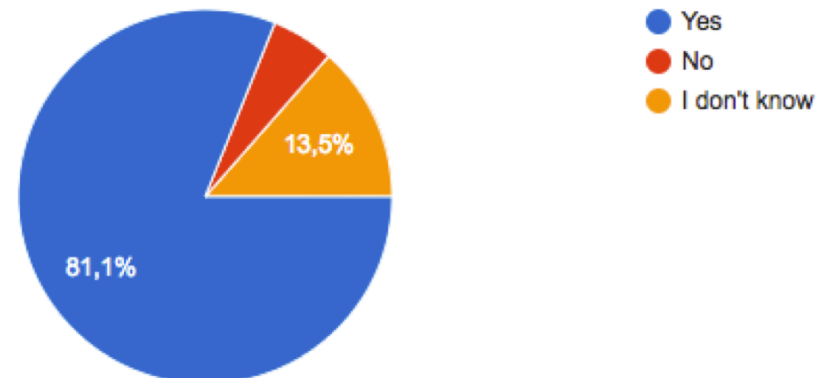
Where do you currently store your data?

37 réponses



Are there any plans for storing data in your organization ?

37 réponses

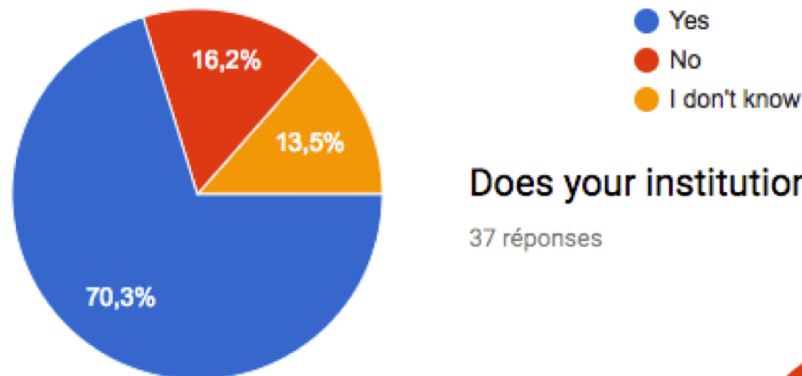


# Rice Data Interoperability Survey output

## n= 37

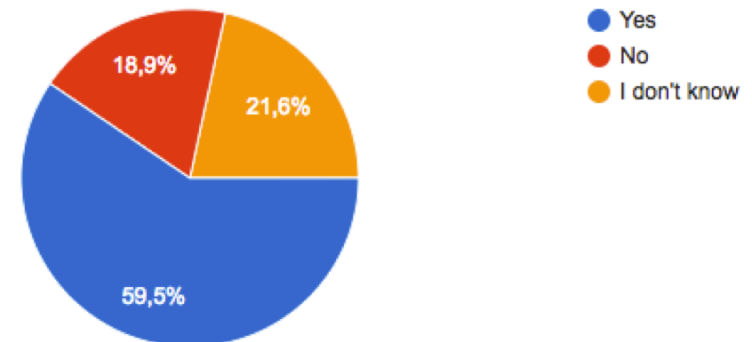
Does your organization have a data policy or guidelines for data management?

37 réponses



Does your institution have its own repository for storing data?

37 réponses

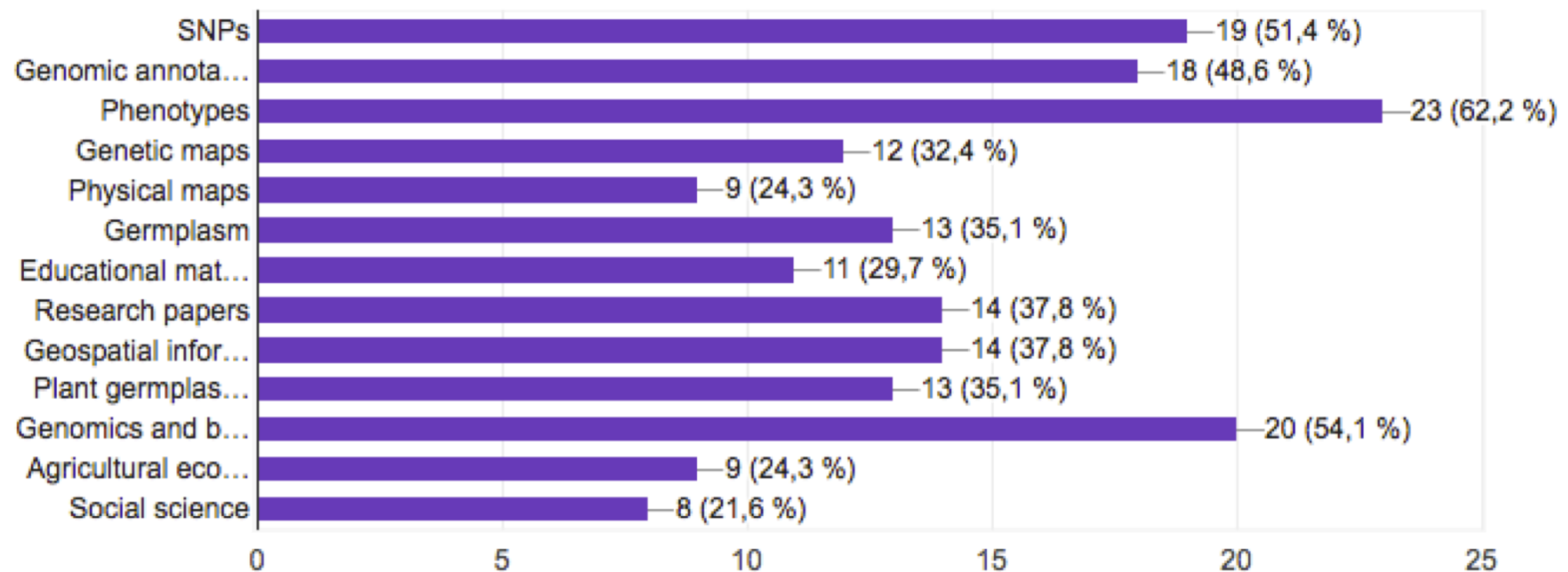


# Rice Data Interoperability Survey output

## n= 37

In your opinion, which datasets will be the most important for rice research in next five years?

37 réponses

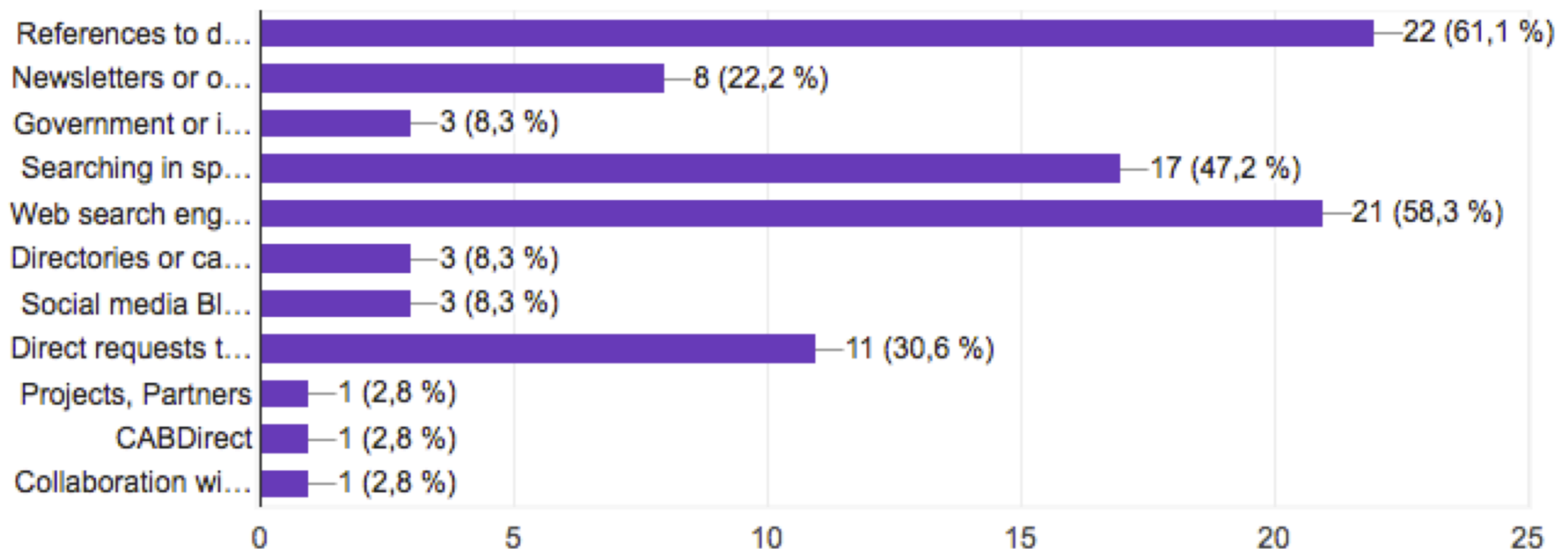


# Rice Data Interoperability Survey output

## n= 37

### How do you normally discover rice research data?

36 réponses

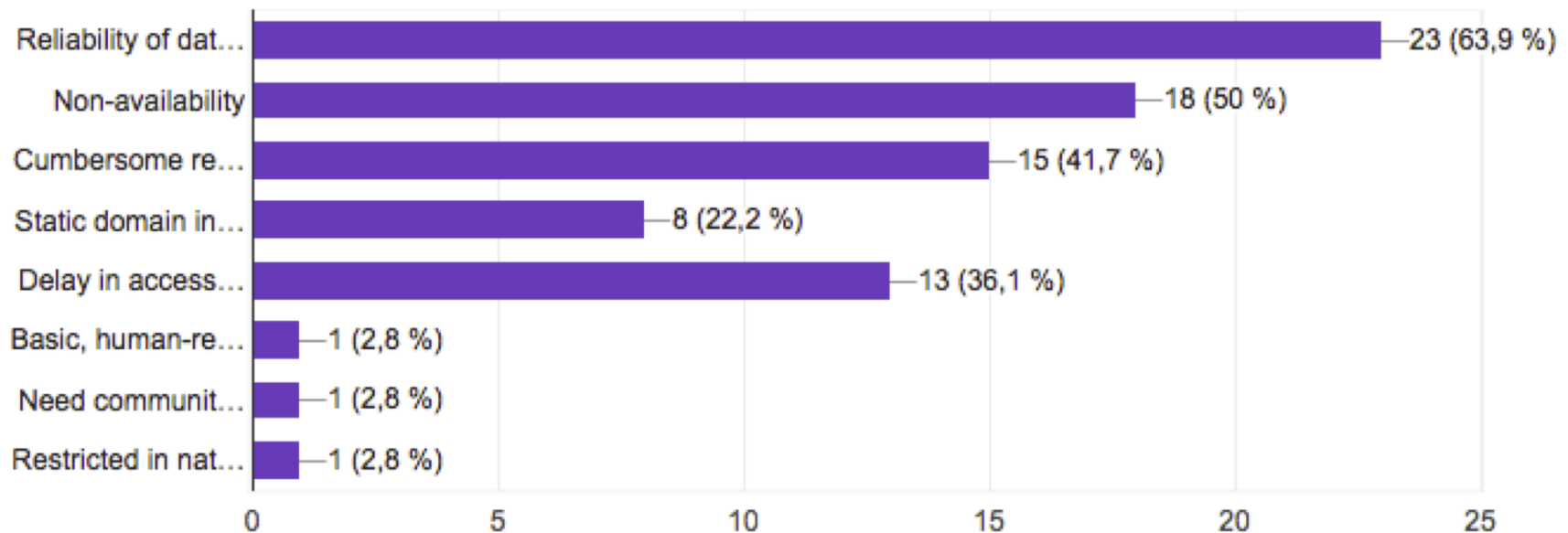


# Rice Data Interoperability Survey output

## n= 37

Where do you see the GAP in the data/ information available on the public domain?

36 réponses



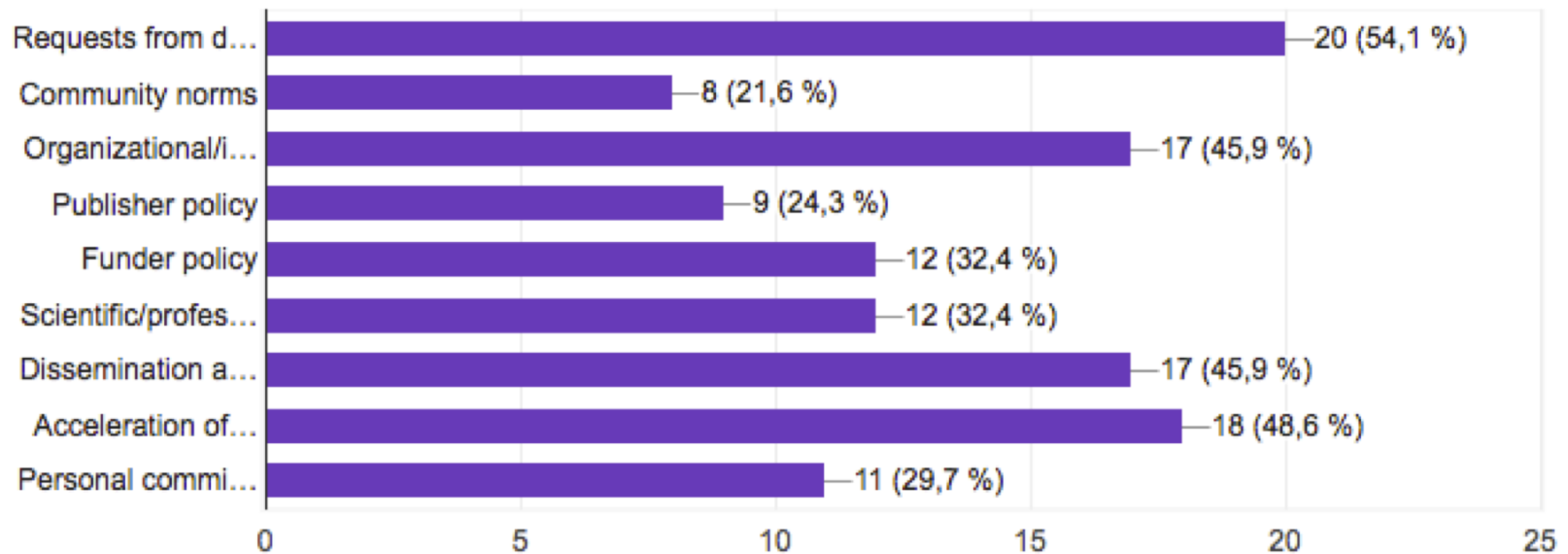


# Rice Data Interoperability Survey output

## n= 37

Which of the following are important motivators to share your work across the regions and countries?

37 réponses

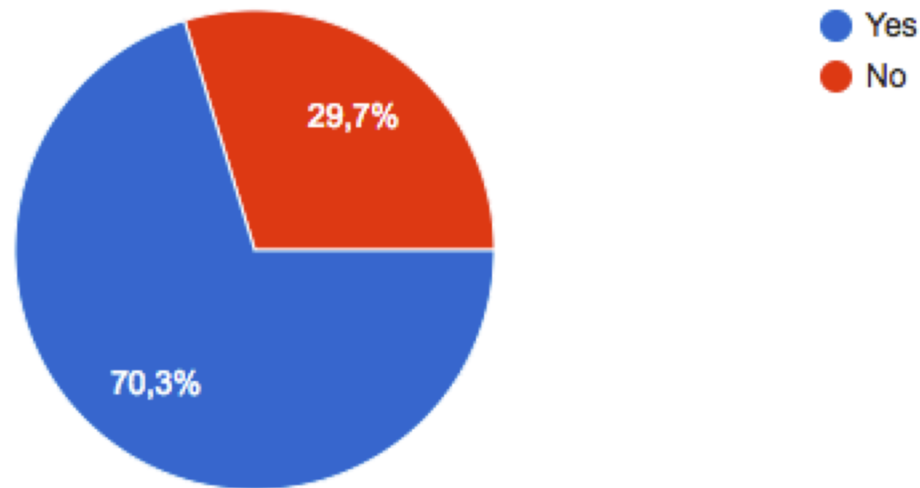


# Rice Data Interoperability Survey output

n= 37

Do you work with phenotypes / trait observations?

37 réponses



# Rice Data Interoperability Survey output

n= 37

What type of data do you usually look for?

