Coordinating global research for wheat

WHEAT DATA INTEROPERABILITY (WDI) WORKING GROUP MEETING MINUTES

Meeting of the Wheat Data Interoperability RDA working group

29 - 30 June 2015
INRA Headquarters, 147 rue de l'Université, Paris
Chaired by Esther Dzalé Yeumo (INRA, France)

Participants

Alaux Michael (INRA, France)	Laporte Marie-Angélique (Bioversity, France)
Aubin Sophie (INRA, France)	Letellier Thomas (INRA, France)
Baumann Ute (Adelaide Uni, Australia)	Lucas Hélène (INRA, France)
Buche Patrice (INRA, France)	Pommier Cyril (INRA, France)
Cooper Laurel (Planteome, USA)	Shrestha Rosemary (INRA, France)
Fulss Richard (CIMMYT, Mexico)	Subirats Imma (FAO, Italy)
Hologne Odile (INRA, France)	Thanopoulos Charalampos (Agro-Know, Greece)
	Venkatesan Aravind (IRD, France)

Apologies and Absentees

Larmand Pierre (IRD, France)
Whan Alex (CSIRO, Australia)

Attachments

Presentations: Updates from the WheatIS, Presentation of the Wheat Initiative, Biosharing related-work and standards

MINUTES

- **Welcome and introduction** by Esther Dzalé Yeumo
    - Welcome
    - Objectives of the workshop
    - Presentation of the agenda
  - **Tour de table**

- **Repositioning the RDA working group in its global context**
    - Updates from the WheatIS by Michael Alaux
    - Overview of CIMMY data management practices by Richard Fulss and Rosemary Shrestha
    - Presentation of the Wheat Initiative by Hélène Lucas
    - Biosharing related-work and standards by Susanna Sansone
        - o Note: Biosharing portal provides information on standards and databases and makes them discoverable. Biosharing has links with Elixir and works

with journals. The portal will include a bioportal widget for the visualisation of the standards. Biosharing is willing to collaborate with the WDI working group. The working group could for example have its own collection of standards and databases in the portal.

▶ Overview of INRA phenotyping-related data management practices by Cyril Pommier
▶ Agro-know's potential contribution to the wheat research community by Charalampos Thanopoulos
▶ Presentation of the ontologies survey by Laurel Cooper and Imma Subirats

▶ **Parallel work sessions**
  ▶ Parallel session 1 - Germplasm metadata: define an annotated tabular data model for multi-crop passport data (MCPD) plus a CSV template to facilitate the reuse of tabulated passport data.
    o The Json tabular framework is available at: https://github.com/edzale/wheat/blob/master/mcpd.json
    o Follow-up actions:
      • Finalize the version 1 of the Json framework
        a. How to specify a list of accepted values for a column?
        b. Can we define a primary key?
        c. Updates from the MCPD authors
      • Provide guidelines on how to publish CSV files along with the Json tabular framework
      • Provide a CSV template
      • Write a rationale for this work and provide examples with real data samples
  ▶ Parallel session 2 - Phenotyping Metadata: propose a metadata framework starting from IWIS and BMS + standardization of traits.
    o IWIS- International Wheat Information System is being phased out in favor of using the BMS. The proposed new BMS
      o Provides the researcher with the ontology terms, depending on the specie of interest- e.g. wheat, maize
      o Provides the Crop Research Ontology to fill in the details of the metadata
      o BMS converts the metadata and observations into the cvs template
    o We don't need a new metadata framework and will build on the following ones : Crop Research Ontology, GnpIS. Ephesis, Breeding API, PHIS/Phenome, PODD, USDA MOWG Metadata and Ontologies working group, Genomic Standards Consortium http://gensc.org/- MixS schema, Biosharing https://www.biosharing.org/
    o We must analyze how to implement Open data sharing general principles, like FAIR (https://www.force11.org/group/fairgroup/fairprinciples), USDA Big Data initiative, RDA Metadata working Group
    o Follow-up actions and upcoming outputs:
      • Phenomic Metadata specifications to handle whole trial or single phenotype observations (as used in genetic/GWAS) studies
      • Simple example of the MCPD success: easy to understand and implement
        o Phenotype = Genotype * Environment
        o Phenotype= measurement of simple trait (yield, flowering date, LAI, NDVI, biomass), or elaborated trait (drought tolerance)
        o Genotype= germplasm (ID + MCPD), sequence level identification (markers, alleles, GO/SO, other recommendations), etc.
        o Environment= Site/location, Year, Trial management practice, Crop Ontology, XEO, Experimental Factor Ontology (EFO), Phenome Env Ontology, EO, ENVO
    o Implementation in information systems (BMS, GnpIS) and Exchange Format (ISA TAB, etc.)

- **Review of the guidelines on how to produce wheat data**
  - General structure
    - o Change "Genetic & Physical maps" to "Physical maps" as we will not provide recommendations for genetic maps
    - o Remove the "Material" section from the menu
    - o Add a "Use cases" section in the menu
    - o Move the ontologies survey in "Ontologies/vocabularies" section
    - o Move the standards survey in the "About" section
    - o Add the logo of the WDI on the Home page
    - o Move the "Adopters" from the "Partners" page and create a subsection dedicated to the "Adopters" in the "Getting involved" section. Adopters could be individuals, groups or projects.
    - o Add a subsection "WDI working group members" in the "About" section
  - For each subsection in the "Guidelines" sections, reorganise the content following the plan in the "sequence variations" subsection, that is:
    - o Recommendations
      - Summary (of the recommendations)
      - A rationale for each recommendation
    - o Most popular tools
    - o Examples

    Make sure that each recommendation is clearly stated. If for any reason it is impossible to make a recommendation, just say it.
  - Finalize the "Ontologies and vocabularies" section
  - Propose a new home page, more visual, with less text, focusing on the objectives of the guidelines.
  - Replace "cookbook" by "recommendations" or "guidelines"
  - Add a disclaimer and license information
  - Activate the statistics

- Perspectives and follow up
  - Reviewing the guidelines: we agreed the need for a review of the guidelines by biologists producing the data. The review process will take place as soon as version 1 of the guidelines is ready and will include the following steps:
    - Step 1: ask two biologists from two different fields (for example a molecular biologist and a geneticist) to review the guidelines;
    - Step 2: submit the guidelines to the Wheat Initiative Expert Working Groups for review and validation;
    - Step 3: advertise the guidelines through the Wheat Initiative, via its Expert Working Groups, and to the global wheat research community.
  - Implementing the prototype: Aravind ensured that the AgroLD tool is available; he will implement a use case in AgroLD which is "Identification of wheat genes that control root growth". Ute will provide the data that are necessary for the implementation of the use case.
  - Promoting the guidelines: a paper will be prepared for publication in an open access journal. In addition, the work should be promoted through the partner's websites and newsletters.
  - Request an extension of 6 months of the working group to RDA: the working group has been endorsed by RDA to work within its umbrella during 18 months. The 18 months will be reached in September. We agreed on the need to request an extension of the duration of the group in the context of RDA.
  - Assess the relevance of extending the guidelines to more wheat data types in accordance with the Wheat Initiative Expert Working Groups.

Action	Who (the leads are underlined)	Deadline
Reorganize the "Genome annotations" subsection	<u>Aravind</u>, Pierre	End of July
Reorganize the "Phenotypes" subsection	<u>Cyril</u>, Rosemary, Marie-Angélique, Patrice	End of July
Reorganize the "Germplasm" subsection	Cyril, <u>Rosemary</u>, Marie-Angélique, Patrice	End of July
Reorganize the "Gene expression" subsection	Laurel, <u>Ute</u>	End of July
Reorganize the "Genetic and physical maps" subsection	<u>Michael</u>, Thomas, Hélène	End of July
Finalize the "Ontologies and vocabularies" section	<u>Imma</u>, <u>Laurel</u>, Marie-Angélique, Sophie, Rosemary, Esther	End of July
Propose a new home page, more visual, with less text, focusing on the objectives of the guidelines.	Richard, <u>Imma</u>	End of July
Replace cookbook by recommendations or guidelines	All	End of July
Add the logo of the WDI to the home page	Esther	End of July
Find 2 experts for the review of the guidelines	Ute	End of August
Submit the guidelines to the wheat expert working groups	Michael	End of August
Request a statement from the Wheat Initiative	Hélène	Depends on the reviewing
Activate the statistics on the website	Esther	End of July
Implement a use case in AgroLD which is "Identification of wheat genes that control root growth".	Aravind	End of August

Provide the data that are necessary for the implementation of the use case.	Ute	End of July
Add a disclaimer and license information	Imma, Richard	End of July
Promote the guidelines	Richard and All	
Phenomic Metadata specification to handle whole trial or single phenotype observations (as used in genetic/GWAS) studies + Implementation in information systems (BMS, GnpIS) and Exchange Format (ISA TAB, …)	Cyril, Rosemary, Marie-Angélique, Laurel,	
Assess the relevance of extending the guidelines to more wheat data types in accordance with the wheat expert working groups.	Michael, Esther, Richard	End of December
Finalize the version 1 of the Json schema for the passport data exchange	Esther, Aravind, Imma, Ute, Sophie, Michael, Thomas	End of September
Provide guidelines on how to publish CSV files along with the schema Provide a CSV template	Esther, Aravind, Imma, Ute, Sophie, Michael, Thomas	End of September
Write a rationale for this work and provide examples with real data samples	Esther, Aravind, Imma, Ute, Sophie, Michael, Thomas	End of October