WheatIS:
Progress report

Michael Alaux
WheatIS data submission
DSpace

• Beta-version to test: http://urgi.versailles.inra.fr/xmlui/

• At the moment, available submission formats:
  • Genotyping
  • Phenotyping
  • SNP discovery

• An account has been send to each EWG members:
  • You can test -> the data will be deleted
  • Please give feedback: michael.alaux@versailles.inra.fr
DSpace

• To do:
  • Developments based on feedbacks
  • Add new formats
DSpace Repository

DSpace is a digital service that collects, preserves, and distributes digital material. Repositories are important tools for preserving an organization’s legacy; they facilitate digital preservation and scholarly communication.

Communities in DSpace

Select a community to browse its collections.

- Wheat Community

Recently Added

Test SNP Disco Vitis 2
Unknown author

View more
DSpace

Item submission

General Information → Trial → Variable → Plant Material → Treatment → Upload → Review

License → Complete

Describe Item

Title*: Enter the main title of the item.

Principal Contact:
Enter principal contact informations of this item below.

Last name* First name*

Email*
• Phenotyping data submission demo by Thomas: http://urgi.versailles.inra.fr/xmlui/
WheatIS distributed search tool
• **Beta-version** to test: [https://urgi.versailles.inra.fr/wheatis](https://urgi.versailles.inra.fr/wheatis)

• Rely on the transPLANT model

• Data types queried at the moment:
  • Sequences
  • Markers
  • Accession
  • Experiment
  • QTL
  • Genetic maps
SoIR search

• Nodes queried at the moment:
  • URGI
    □ GnpIS
  • EBI
    □ Ensembl Plants
  • IPK
    □ CR-EST
    □ GEBIS
    □ MetaCrop
SoIR search

• To do:
  • Add more nodes and data types
  • Please give feedback:
    michael.alaux@versailles.inra.fr
SoIR search

Gene: Traes_4DS_D3D974286

Location: Scaffold IWGSC_CSS 4DS scaffold 2322117: 3,099-7,813 reverse strand.

Transcripts: This gene has 1 transcript (splice variant) [Show transcript table]

Summary

Gene type: Protein coding
Prediction Method: Triticum aestivum genes annotated by MIPS

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

View genomic alignments of all homoeologues

Gene Legend:
- Import / Other
- Protein coding
SoIR search

Filter by data type

Filter by wheat species

Filter by database « nodes »
SolR search

• Demo: https://urgi.versailles.inra.fr/wheatis
Report of the RDA wheat group
RDA group main objectives

Focus:
- Polymorphisms (SNPs)
- Genome annotations
- Phenotypes
- Genetic & Physical Maps
- Germplasm
- expression data

Differences in representation formats of each data type
- Many data sources with their own data structure
- Differences in the interpretation of the meaning of the data

Standards

Translations

Interoperability framework

From Esther Dzalé Kaboré  Michael Alaux
Workflow

Survey/Interviews

- Identify metadata, data formats and vocabularies used within/by the Wheat research community

Workshops

- Identify/agree on the use of common metadata, data formats and vocabularies
- Assess then improve the level of accessibility and interoperability of data formats and vocabularies
- Collect interoperability use cases

Implementation

- Interactive cookbook: recommendations + guidelines
- Hub of linked vocabularies
- Prototype - Assess the gain of interoperability based on collected use cases
Where we are

Survey/Interviews

- A survey launched in April 2014
- Answers from more than 200 respondents

Workshops

- 1-2 October 2014
- List of recommended data formats and vocabularies for each data type
- List of follow up actions for each data type (standardization e.g. for traits, minimal set of metadata e.g. for SNP files provenance or for markers, QTL and map handling, check for existing mapping tools, etc.)
- List of interoperability use cases
## Standards summary

<table>
<thead>
<tr>
<th>Recommendations</th>
<th>Follow up actions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SNPs</strong></td>
<td></td>
</tr>
<tr>
<td>Use of VCF data format</td>
<td>Look at a metadata set to contextualize the provenance of SNP files</td>
</tr>
<tr>
<td><strong>genome annotations</strong></td>
<td></td>
</tr>
<tr>
<td>Use of GFF3 data format</td>
<td>Provide description guidelines for filling in content for column 9</td>
</tr>
<tr>
<td>Use of ontologies to fill « Attributes » column (a list of feature</td>
<td></td>
</tr>
<tr>
<td>attributes in the format tag=value), column 9</td>
<td></td>
</tr>
<tr>
<td><strong>Germplasms</strong></td>
<td></td>
</tr>
<tr>
<td>Use of MPCD and Darwin Core Germplasm formats</td>
<td>Check how to integrate with tool specific formats (Grin Global, Genesys)</td>
</tr>
<tr>
<td></td>
<td>Provide a table like human readable format for DWC Germplasm</td>
</tr>
<tr>
<td><strong>Gene expression</strong></td>
<td></td>
</tr>
<tr>
<td>Follow existing format standards laid out by repositories (NCBI GEO,</td>
<td>Check for mapping and conversion tools</td>
</tr>
<tr>
<td>EBI Array Express)</td>
<td></td>
</tr>
<tr>
<td><strong>Physical maps</strong></td>
<td></td>
</tr>
<tr>
<td>Same as for genome annotations</td>
<td>Same as for genome annotations</td>
</tr>
<tr>
<td><strong>Genetic maps</strong></td>
<td></td>
</tr>
<tr>
<td>Data formats depend on tools that are used, rather concentrate on</td>
<td>Look at a minimal metadata set to handle markers, QTL, maps</td>
</tr>
<tr>
<td>metadata harmonization</td>
<td>Obtain details for linking requirements</td>
</tr>
<tr>
<td><strong>Phenotypes</strong></td>
<td></td>
</tr>
<tr>
<td>Use of isa-tab data format</td>
<td>Standardize the traits metadata</td>
</tr>
<tr>
<td></td>
<td>Improve the reference to ontologies used for traits</td>
</tr>
</tbody>
</table>
Next steps

• In progress:
  ❖ A first version of the « cookbook » will be available on a website: http://ist.blogs.inra.fr/wdi/

• Work with experts to meet the identified needs
  ❖ Metadata harmonization, minimal metadata sets
  ❖ Mapping among metadata, formats and ontologies
  ❖ Survey on wheat ontologies
    https://docs.google.com/forms/d/1YozHjRWVlnQSVx0e2983C9shnd8E0skHxekJ_zK5T00/viewform?pli=1&edit_requested=true

• Two workshops in 2015
  ❖ Refine the cookbook
  ❖ Collect more interoperability use cases
Acknowledgments
Questions

DSpace test:  
http://urgi.versailles.inra.fr/xmlui

SolR test:  
https://urgi.versailles.inra.fr/wheatis

Contact me at  michael.alaux@versailles.inra.fr