OPINION ARTICLE

Developing data interoperability using standards: A wheat community use case [version 1; peer review: 1 approved]

Esther Dzale Yeumo[1], Michael Alaux[2], Elizabeth Arnaud[3], Sophie Aubin[1], Ute Baumann[4], Patrice Buche[5], Laurel Cooper[6], Hanna Ćwiek-Kupczyńska[7], Robert P. Davey[8], Richard Allan Fulss[9], Clement Jonquet[10,11], Marie-Angélique Laporte[3], Pierre Larmande[12,13], Cyril Pommier[2], Vassilis Protonotarios[14], Carmen Reverte[15], Rosemary Shrestha[9], Imma Subirats[16], Aravind Venkatesan[12], Alex Whan[17], Hadi Quesneville[2]

[1]INRA, UAR 1266 DIST Délégation Information Scientifique et Technique, Centre de recherche Ile-de-France-Versailles-Grignon, Versailles, 78000 , France
[2]INRA, UR 1164 URGI Unité de Recherche Génomique-Info, Université Paris-Saclay, Versailles, 78000, France
[3]Bioversity International, Montpellier, 34397, France
[4]School of Agriculture, Food and Wine, University of Adelaide, Glen Osmond, SA, 5064, Australia
[5]Institut National de la Recherche Scientifique, Centre National De La Recherche Scientifique, Montpellier, 34000, France
[6]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, 97331, USA
[7]Department of Biometry and Bioinformatics, Institute of Plant Genetics, Polish Academy of Sciences, Poznań, 60-479, Poland
[8]Earlham Institute , Norwich, NR4 7UZ, UK
[9]International Maize and Wheat Improvement Center, Texcoco, 56237, Mexico
[10]Center for Biomedical Informatics Research, University of Montpellier, Stanford, CA, 94305, France
[11]Laboratory of Informatics, Robotics and Microelectronics of Montpellier , Stanford University, Montpellier, 34090 , USA
[12]Institut de Biologie Computationnelle, Institut de Recherche pour le Développement , Montpellier, 34090, France
[13]Université Montpellier, Marseille, 13572, France
[14]NEUROPUBLIC S.A., Piraeus, GR18545, Greece
[15]IRTA. Ctra. de Poble Nou, Sant Carles de la Ràpita, E-43540, Spain
[16]Food and Agriculture Organization of the United Nations, Rome, 00153, Italy
[17]Commonwealth Science and Industrial Research Organisation, Agriculture and Food, Canberra, ACT, 2601, Australia

v1 **First published:** 16 Oct 2017, **6**:1843
https://doi.org/10.12688/f1000research.12234.1
Latest published: 06 Dec 2017, **6**:1843
https://doi.org/10.12688/f1000research.12234.2

Abstract

In this article, we present a joint effort of the wheat research community, along with data and ontology experts, to develop wheat data interoperability guidelines. Interoperability is the ability of two or more systems and devices to cooperate and exchange data, and interpret that shared information. Interoperability is a growing concern to the wheat scientific community, and agriculture in general, as the need to interpret the deluge of data obtained through high-throughput technologies grows. Agreeing on common data formats, metadata, and vocabulary standards is an important step to obtain the required data interoperability level in order to add value by encouraging data sharing, and subsequently facilitate the extraction

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 06 Dec 2017	✓ view	✓ view
version 1 16 Oct 2017	✓ view	

1. **Ramil Mauleon**, International Rice

of new information from existing and new datasets.
During a period of more than 18 months, the RDA Wheat Data Interoperability Working Group (WDI-WG) surveyed the wheat research community about the use of data standards, then discussed and selected a set of recommendations based on consensual criteria. The recommendations promote standards for data types identified by the wheat research community as the most important for the coming years: nucleotide sequence variants, genome annotations, phenotypes, germplasm data, gene expression experiments, and physical maps. For each of these data types, the guidelines recommend best practices in terms of use of data formats, metadata standards and ontologies. In addition to the best practices, the guidelines provide examples of tools and implementations that are likely to facilitate the adoption of the recommendations.
To maximize the adoption of the recommendations, the WDI-WG used a community-driven approach that involved the wheat research community from the start, took into account their needs and practices, and provided them with a framework to keep the recommendations up to date. We also report this approach's potential to be generalizable to other (agricultural) domains.

Keywords
wheat, data interoperability, metadata, ontology repository, bio-ontologies, standard vocabularies, data formats

This article is included in the Agriculture, Food and Nutrition gateway.

Research Institute, Metro Manila, Philippines

2. **Peter McQuilton**, University of Oxford, Oxford, UK

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Hadi Quesneville (hadi.quesneville@inra.fr)

Author roles: Dzale Yeumo E: Conceptualization, Data Curation, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Alaux M**: Data Curation, Investigation; **Arnaud E**: Data Curation, Investigation; **Aubin S**: Data Curation, Investigation; **Baumann U**: Data Curation, Investigation; **Buche P**: Data Curation, Investigation; **Cooper L**: Data Curation, Investigation; **Ćwiek-Kupczyńska H**: Data Curation, Investigation; **Davey RP**: Data Curation, Investigation; **Fulss RA**: Data Curation, Investigation; **Jonquet C**: Data Curation, Investigation; **Laporte MA**: Data Curation, Investigation; **Larmande P**: Data Curation, Investigation; **Pommier C**: Data Curation, Investigation; **Protonotarios V**: Data Curation, Investigation; **Reverte C**: Data Curation, Investigation; **Shrestha R**: Data Curation, Investigation; **Subirats I**: Data Curation, Investigation; **Venkatesan A**: Data Curation, Investigation; **Whan A**: Data Curation, Investigation; **Quesneville H**: Conceptualization, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The WDI-WG was partially funded by IWI funding received by the WheatIS Expert Working Group.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Dzale Yeumo E, Alaux M, Arnaud E *et al.* **Developing data interoperability using standards: A wheat community use case [version 1; peer review: 1 approved]** F1000Research 2017, **6**:1843 https://doi.org/10.12688/f1000research.12234.1

First published: 16 Oct 2017, **6**:1843 https://doi.org/10.12688/f1000research.12234.1

Introduction

Wheat was one of the first domesticated food crops, and for 8000 years it has been the basic staple food of major civilizations in Europe, West Asia, and North Africa. According to the International Wheat Initiative (IWI, http://www.wheatinitiative.org/), a framework to establish strategic research and organization priorities for wheat research at the international level in both developed and developing countries, wheat is the most widely grown cereal grain, cultivated in about 17% of the total arable land globally, and the staple food for 35% of the world's population, providing 20% of all calories consumed by people worldwide and more protein in the human diet than any other crop (http://www.wheatinitiative.org./about-wheat/factsheets-infographics). According to the Consultative Group on International Agricultural Research's research program on Wheat (http://wheat.org), an estimated 1.2 billion poor people depend on wheat, a crop that is particularly vulnerable to climate change.

The IWI has identified easy access and interoperability of all wheat-related data as a top priority for the wheat research community, which is in line with FAIR data principles[1]. Interoperability is the ability of two or more systems and devices to cooperate and exchange data, and interpret that shared information[2]. An important goal is to make the best possible use of the existing and upcoming wealth of genetic, genomic, and phenotypic data in fundamental and applied wheat science. Hence, data interoperability has become a hot topic in this community, given the ever-growing data deluge coming from improvements in data generating technologies and large-scale computational methods for handling DNA and RNA sequencing, high throughput genotyping and phenotyping, high throughput imaging, and satellite monitoring. However, achieving data interoperability is difficult not only because of data and tool heterogeneity, i.e., the '*technical debt*', but also because of social and scientific issues, such as lack of curation experts, lack of value chains for data generators, and lack of a first class digital citizen recognition for data managers, i.e. the '*cultural debt*'.

To help address these debts, the Wheat Data Interoperability Working Group (WDI-WG, https://www.rd-alliance.org/groups/wheat-data-interoperability-wg) was created as one of the Research Data Alliance working groups (https://www.rd-alliance.org/groups), under the umbrella of the WheatIS Expert Working Group (http://wheatis.org/), which is endorsed by the IWI to build an international information system for wheat genetic, genomic and phenotypic data. The Working Group included wheat scientists, as well as ontologists and data experts from different organizations and countries, and its mission was to provide a common framework for describing and representing data with respect to existing open data standards. From the outset, the objective of the WDI-WG was to deter communities from creating new standards, which would have made the already-complex landscape of existing data standards even more complex. The WDI-WG collected valuable information through two surveys of the wheat research community, comprising responses regarding existing data formats, practices, and the use of ontologies and controlled vocabularies. From these surveys, the WDI-WG then developed a set of specific recommendations, and worked to facilitate data interoperability through the harmonization of data formats, data models and vocabularies usage, thus aiming to address the main interoperability issues. The proposed recommendations have been endorsed by the WheatIS Expert Working Group and the Technical Advisory Board of the RDA (RDA-TAB).

This paper describes the results and the collaborative methodology used by the WDI-WG, which we believe will be of interest to formalize data interoperability in other crop research communities.

Developing the recommendations

A community driven methodology

From the preparation to the publication of the recommendations, the WDI-WG strongly based its work on the wheat research community. Similarly, the maintenance of the recommendations will be reliant on feedback of the community and the review of a steering group, which includes representatives of the adopters of the guidelines. The main steps of the methodology adopted by the WDI-WG are represented in Figure 1 and are described in more detail in the rest of this section.

Building on existing standards and practices

The WDI-WG standpoint was to build on prior practices in use in the community, reusing existing standards as much as possible. Gaps, if they existed, could then be filled through the development of new standards. This principle led the working group to start with two surveys, interrogating the wheat research community through the IWI communication channels. The first, "Data standards in the wheat research community wheat data interoperability WG", studied the usage of data standards in the wheat research community through a series of questions sent out to researchers and stakeholders in wheat science. The questions and answers are summarized in a report[11]. The results allowed the group to identify the most commonly used data formats and controlled vocabularies in the wheat research community. The second survey, "Towards a Comprehensive Overview of Ontologies and Vocabularies for Research on Wheat", focusing on ontologies and vocabularies, allowed the WDI-WG to collect information about the visibility, interoperability, domain, and content of relevant ontologies and vocabularies. The questions and answers of this survey are also summarized in a report[12].

Converging towards the recommendations

Two meetings of the WDI-WG were organized in 2014 and 2015, as well as regular face-to-face and online meetings, to analyze the survey results in order to draw recommendations. Calls for participation were regularly posted on the websites of RDA and IWI and channeled by the stakeholders. During these working sessions, wheat research scientists, data and information managers, and semantic web experts discussed and collectively agreed on a set of recommendations to cover the widest set of requirements of the communities they supported. The criteria used to guide the recommendation process were the following: (i) reuse existing standards and reinforce existing good practices with regards to interoperability to preserve synergies that work well in the community and (ii) promote emerging standards and practices where gaps exist.

Figure 1. A community driven methodology for data interoperability guidelines design.

For the following data types of interest to the WDI-WG, the surveys confirmed the existence of adequate consensus regarding data exchange formats: DNA sequence and any associated variants, genome/transcriptome annotations, gene expression data, and physical maps. As such, the WDI-WG recommended the formats that were the most used and/or compliant with the most popular tools and/or already interoperable with other data formats. For example, the GFF3 file format (http://gmod.org/wiki/GFF3) is found to be widely used by the community to represent genome annotations. Moreover, a Genbank-to-GFF3 script converter is available (http://www.hpa-bioinformatics.org.uk/biosnippets/snippets/115), in addition to a GFF3 validator tool (http://genometools.org/cgi-bin/gff3validator.cgi). Thus, the WDI-WG recommended GFF3 for the representation of genome annotations.

However, unlike the aforementioned data types, the wheat data standards survey did not show good consensus for phenotypes and germplasm in terms of data exchange formats and data description practices. For these data types, the WDI-WG collectively agreed to recommend emerging standards, such as (i) Minimum Information About Plant Phenotyping Experiment (MIAPPE)[4] and its ISA-TAB implementation[4]; (ii) the Crop Ontology[5], especially the Wheat Trait Ontology for phenotypes (http://agroportal.lirmm.fr/ontologies/CO_321); and (iii) the FAO-IPGRI Multi-Crop Passport Ontology (http://agroportal.lirmm.fr/ontologies/CO_020) for germplasm.

Validation of the recommendations

Prior to their endorsement by RDA, the resulting recommendations have been reviewed by the WheatIS expert working group. As a deliverable of a RDA working group, the recommendations received feedback from the RDA community and validation from the RDA-TAB.

The WDI-WG also used many of the available channels in order to obtain feedback from the wheat research community. In particular, feedback was requested, and was obtained, from communications through the Wheat Initiative's website, the Food and Agriculture Organization Agricultural Information Management Standards (AIMS) newsletter, and various national and institutional mailing lists.

Publishing the recommendations

The recommendations are published on the b2share repository[13], and a website (http://datastandards.wheatis.org), which provides the option to submit comments and suggestions. Thus, recommendations can be updated as required by the wheat community, which is of significance since technologies and practices are constantly evolving. Hence, this kind of media allows keeping the guidelines relevant and useful for the wheat research community.

Disseminating the recommendations

The Wheat Data Interoperability Guidelines website

The first and main output of the WDI-WG is a set of recommendations for describing, representing and linking wheat data. These recommendations are available at http://datastandards.wheatis.org and cover the following data types: sequence variations, genome annotations, phenotypes, physical maps, germplasm, and gene expression. The navigation menu of the website includes four main items (Figure 2): "Guidelines", "Ontologies and vocabularies", "Use cases", and "Getting involved". The guidelines menu contains a section for each of the data types addressed by the WDI-WG. Each data type-specific page (Figure 3) contains the following sections: (i) a summary of the recommendations for the indicated data type; (ii) rationalized recommendations about data format standards; (iii) rationalized recommendations about metadata

Figure 2. Main items in the menu of the wheat data interoperability guidelines.

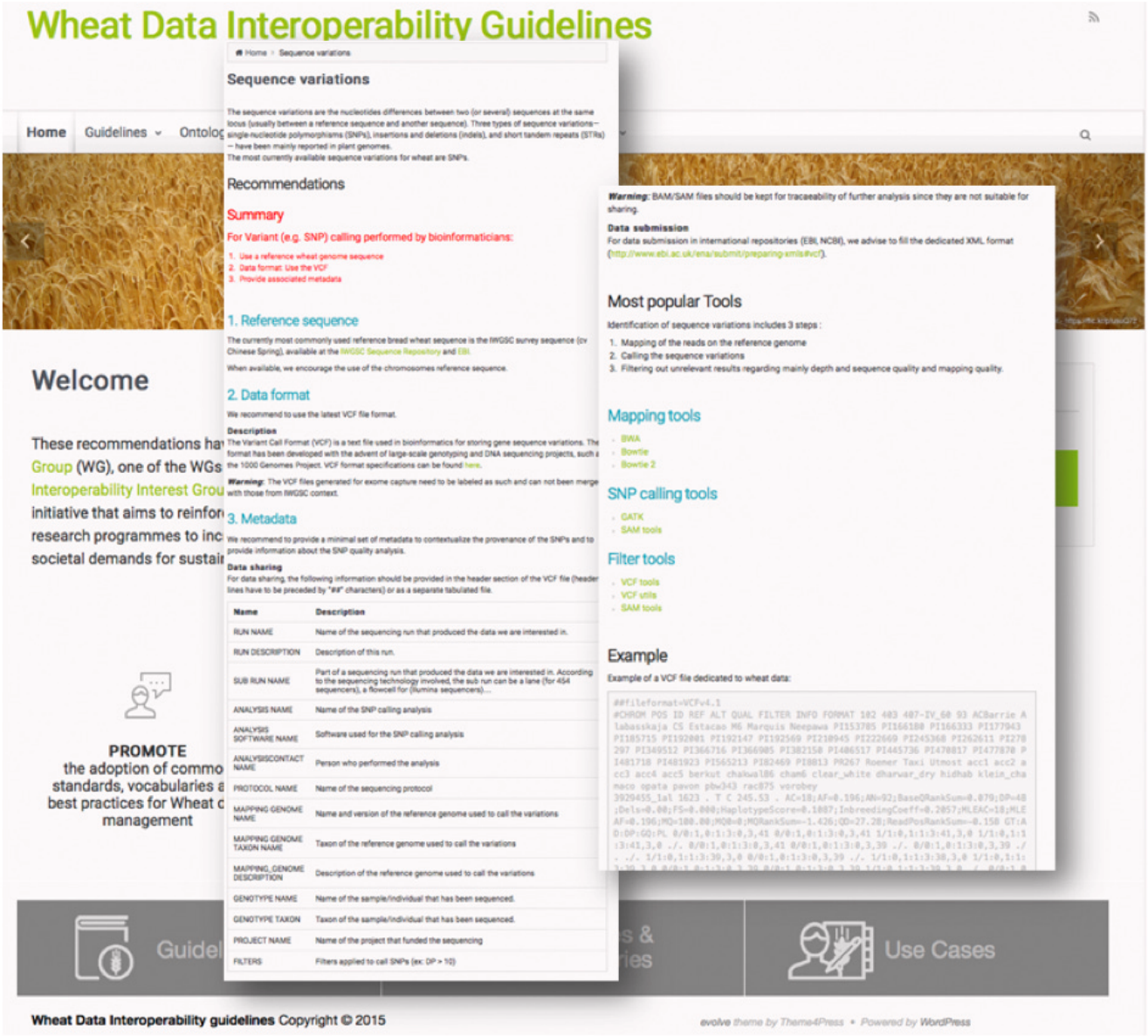

Figure 3. Example of data type specific page (sequence variations).

standards and ontologies; (iv) tools; (v) examples; and (vi) comments. The summary of the recommendations[13] and the http://data-standards.wheatis.org website provide detailed information for each data type covered by the guidelines.

In addition to the data type-specific pages, a page dedicated to ontologies and vocabularies explains their benefits and current situation in the context of wheat research data. The use cases page describes examples of use cases with interoperability issues.

The AgroPortal repository for wheat-related vocabularies

In the context of research data, the use of common vocabularies or ontologies plays a key role in managing, publishing, and reusing data[6]. Words may have different meanings to different people, and standard definitions for these words are key to avoiding miscommunication and enabling good collaboration. Standardized vocabularies and ontologies enhance the efficiency of interoperability and the effectiveness of data exchange, thus facilitating the reuse of data by others[5,7]. A need to offer a common unique repository of standard vocabularies and ontologies relevant for wheat was identified, and the AgroPortal (http://agroportal.lirmm.fr/)[8], a starting project in 2015, was recognized as suitable solution.

AgroPortal is a collaborative initiative to build a repository of vocabularies and ontologies for agronomy, and related domains (plant sciences, biodiversity, and nutrition). By reusing the National Center for Biomedical Ontologies (NCBO) BioPortal technology[9], the portal features ontology hosting, search, versioning, visualization, comment, recommendation, and enables semantic annotation, as well as storing and exploiting ontology alignments, all within a semantic web compliant infrastructure. The AgroPortal specifically pays attention to respect the requirements of the agronomy community in terms of ontology formats (e.g., SKOS, trait dictionaries), or supported features (metadata, annotation). AgroPortal addresses the WDI-WG identified need, while offering a set of interesting features for the ontologies being hosted. Therefore, we have created and maintain an explicit group within AgroPortal and its corresponding slice (http://wheat.agroportal.lirmm.fr/). Slices are a mechanism supported by the platform to allow users to interact (both via user and application programming interfaces) only with a subset of ontologies in AgroPortal. If browsing the slice, all the portal features will be restricted to a subset, enabling users to focus on their specific use cases. As of today, AgroPortal's wheat group contains 20 ontologies of the 23 identified by the WDI-WG[10]. Each ontology has been carefully described (with licenses, authority, availability, etc.), and a new metadata property (omv:endorsedBy) is used to show the ontology's endorsement by the WDI-WG. The wheat slice in AgroPortal will allow the community to share common meanings of the words they utilize to describe and annotate data, which will in turn make the data more machine-readable and interoperable. Furthermore, the slice will enable wheat-related ontology developers to make their ontologies more visible to the agronomic research community, thus contributing to reduce the proliferation of concurrent ontologies on the Web. This slice has been reported in the WDI-WG guidelines web site (section "Ontologies and vocabularies"), and used as a reference resource to identify and select ontologies related to wheat since then.

Discussion

Validation issues

The WDI-WG's guidelines have been collaboratively built and validated under the umbrella of two authoritative organizations (the WheatIS expert working group and RDA, respectively). The Expert Working Groups of the Wheat Initiative have been instrumental in efficiently interacting with the wheat scientific community in order to take into account the needs from the different fields of biology working on this species. Consequently, the needs and the practices of this community were well-addressed. In addition, the WDI-WG took care to build on existing good practices and preserve prevailing strong synergies in the community, while proposing new standards and practices where relevant. This has been achieved by consulting the wheat research community and experts as frequently as needed. This strategic approach ensures a better adoption of the guidelines by the wheat research community.

Despite this initial strong validation process, the WDI-WG anticipates changes to the recommendations, especially due to an evolving landscape of standards and practices. The blog-like website that hosts the recommendations will facilitate rapid implementation of future changes.

One pitfall the WDI-WG managed to avoid is the quest for immediate comprehensiveness. We deliberated focused on the six data types that were considered most relevant by the wheat research community in the coming years. However, the recommendations can be extended to more data types in future.

Adoption issues

In order to maximize the adoption of the recommendations, the WDI-WG favors a bottom-up approach rather than enforcing the choice of particular standards. Consequently, to begin with, it is better that individual project initiatives develop their own usage of the proposed recommendations and standards, especially since there are some standards that share common concepts, but address different needs. We prefer the community to adopt at least some standards rather than none. We provide guidelines to facilitate the decision of standards suggesting the most widely adopted ones. By developing the tools required to map/convert from one standard to another, it should be possible to bridge data respecting different standards. The important point is that a standard is used to remove ambiguities in data semantics and representation to enable automated processing. At a later date, when several standards have converged or become widely adopted, it could be possible to enforce their usage. But the time needed to reach this second step will vary between the different fields of biology.

The WDI-WG will develop training programs to increase the adoption of the guidelines. In fact, the guidelines have already been adopted by a number of stakeholders (http://ist.blogs.inra.fr/wdi/adopters/). However, these are part of large institutions. This highlights the need to provide tools and training to facilitate the adoption of the guidelines within smaller organizations. Two kinds of training will be developed for two types of audience: data managers with technical skills on data management, and biologists

with data knowledge. Another target community of adopters of the WDI-WG's guidelines is software developers. The adoption of the guidelines by this community is essential to showcase the benefits of data interoperability. Therefore, there is a strong need to raise awareness in this community.

Finally, reengineering legacy data in accordance with the WDI-WG is an open question. Indeed, it requires from data producers and managers to convert legacy data in recommended data formats or learn how to annotate data with specific vocabularies, which is not trivial for anyone. Depending on the use case and/or the value of the data, it may or may not be worth making such efforts. The use of automated tools for the transformation of data in different formats (where applicable) is expected to minimize the human effort required for such processes.

Follow up and conclusions

As an RDA working group, the WDI-WG is now in an adoption and maintenance phase. Consequently, the WDI-WG will know focus on dissemination and maintenance activities. A steering group, including representatives of the adopters and the WDI-WG chairs, will drive these activities, taking into account the feedback and contributions of the wheat research community. The action plan of the WDI-WG includes: (i) the promotion of the guidelines via development of information material such as flyers or short videos; and (ii) technical and non-technical training for data managers and scientists, respectively. The WDI-WG will also consolidate the wheat vocabularies group and slice within AgroPortal (http://wheat.agroportal.lirmm.fr/ontologies). In addition to these projects, it is worth mentioning that the methodology and the results of the WDI-WG have inspired the creation of a rice data interoperability working group within the frame of RDA (https://www.rd-alliance.org/groups/rice-data-interoperability-wg.html).

The recommendations of the WDI-WG are intended for data producers, data managers, data consumers, and software developers. They constitute a key building block for FAIR data[1] sharing infrastructures (https://www.force11.org/fairprinciples). Indeed, the adoption of the recommendations will facilitate the depositing of data within well recognized repositories in addition to make them easily understandable and reusable.

Competing interests

No competing interests were disclosed.

Grant information

The WDI-WG was partially funded by IWI funding received by the WheatIS Expert Working Group.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci Data.* 2016; **3**: 160018.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Wegner P: **Interoperability.** *ACM Comput Surv.* 1996; **28**(1): 285–287.
Publisher Full Text
3. Aubin S, Alaux M, Baumann U, *et al.*: **Data standards in the wheat research community.** *Zenodo.* 2014.
Publisher Full Text
4. Ćwiek-Kupczyńska H, Altmann T, Arend D, *et al.*: **Measures for interoperability of phenotypic data: minimum information requirements and formatting.** *Plant Methods.* 2016; **12**: 44.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Shrestha R, Arnaud E, Mauleon R, *et al.*: **Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature.** *AoB Plants.* 2010; **2010**: plq008.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Rubin DL, Shah NH, Noy NF: **Biomedical ontologies: a functional perspective.** *Brief Bioinform.* 2008; **9**(1): 75–90.
PubMed Abstract | Publisher Full Text
7. Jaiswal P, Cooper L, Elser JL, *et al.*: **Planteome: A resource for Common Reference Ontologies and Applications for Plant Biology.** In *Plant and Animal Genome XXIV Conference.* Plant and Animal Genome. 2016.
Reference Source
8. Jonquet C, Toulet A, Arnaud E, *et al.*: **Reusing the NCBO BioPortal technology for agronomy to build AgroPortal.** *7th Int Conf Biomed Ontol.* 2016.
Reference Source
9. Noy NF, Shah NH, Whetzel PL, *et al.*: **BioPortal: ontologies and integrated data resources at the click of a mouse.** *Nucleic Acids Res.* 2009; **37**(Web Server issue): W170–W173.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Subirats I, Cooper L, Shrestha R, *et al.*: **Towards a Comprehensive Overview of Ontologies and Vocabularies for Research on Wheat.** *Zenodo.* 2015.
Publisher Full Text
11. Aubin S, Alaux M, Baumann U, *et al.*: **Data standards in the wheat research community.** *Zenodo.* 2014.
Publisher Full Text
12. Subirats I, Cooper L, Shrestha R, *et al.*: **Towards a Comprehensive Overview of Ontologies and Vocabularies for Research on Wheat.** *Zenodo.* 2015.
Publisher Full Text
13. Dzalé Yeumo E, Fulss R, Alaux M, *et al.*: **Wheat Data Interoperability Guidelines, Ontologies and User Cases. Recommendations from the RDA Wheat Data Interoperability Working Group.** EUDAT B2Share.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 30 October 2017

https://doi.org/10.5256/f1000research.13244.r27011

✔ **Ramil Mauleon**
Strategic Innovation Platform, International Rice Research Institute, Metro Manila, Philippines

The opinion article is well written, and addresses a challenge that plant/crop science researchers face: how to systematically manage data from high throughput technologies (ie. next-gen sequencing, high density genotyping, standardized phenotype measurements/observations)

The inventory made of data standards dealing with wheat research (with greater focus on genomics/sequences and lesser on phenotype and germplasm, as acknowledged by authors) is very comprehensive and I believe covers what most of the wheat community is using.

Some minor improvements I see that could be done on the main paper itself is to mention directly some important summaries /findings of the survey results, without having to open the links to the results of the survey. For example, a reader might wish to know how many (or what proportion) of the wheat research institutions surveyed have data standards of what kind (making the paper citable directly and showing the importance of the WDI-WG paper). This is one of the few summaries that could be directly shown. Another is the mention of data standards for genotyping data (especially high density ones), directly in the paper, this is very important data type , I had to open the survey to know more about this.

Readers who wish to use the recommendations would also likely benefit from having concrete examples of documents that implement the recommendations of the paper directly available (again without having to navigate the external website of the cited resources) within the paper itself. Example, a direct example of snippet of GFF3 for a particular genome annotation would be nice. Can you also mention the most recent resource for wheat genome build, the most authoritative (or most widely used) gene naming of wheat genome (as of writing)? This is very important info this type of data. Another example could be a snippet of a phenotyping experiment result (a field book table, for example), wherein the MIAPPE terms, and the ontology terms tagging the phenotyping data observed/measured appropriate for the experiment can be seen? This gives the reader/researcher ideas on how the data standards/guidelines are used in real-world applications. I went to the http://datastandards.wheatis.org/ externally referred site and I did not easily see a sample dataset that could be used as template.

In summary, the paper is already in a very mature and good state, just having some important summaries of the survey directly mentioned and providing examples of applications of the standards in easily accessible sample documents would be a welcome addition.

Best wishes to the authors!

Is the topic of the opinion article discussed accurately in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Yes

Are arguments sufficiently supported by evidence from the published literature?
Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, data standards, genetics, genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 23 Nov 2017

Hadi Quesneville, Centre de recherche Versailles-Grignon, Versailles, France

Dear Ramil,

Thank you for this positive review and your useful comments. We have modified the manuscript taking into account your suggestions as follows.

We added a summary of the survey results as supplementary material to facilitate the reading of our article. In particular to provide the user with the current usage of the standard by the wheat community.

Concerning examples of documents that implement the recommendations, we consider that this is out of the scope of our paper which describes the methodology we followed to propose standards and guidelines, and not what they are. Because of the nature of the recommendation that would evolve with time, we preferred to implement a website that could be updated according to the community usage as explained in the paper. Writing them in a paper would freeze them and this is contrary to our philosophy. The examples can be found on our guideline website.

Best regards,

Competing Interests: No competing interests were disclosed.

Comments on this article

Version 1

Author Response 23 Nov 2017

Hadi Quesneville, Centre de recherche Versailles-Grignon, Versailles, France

Dear Mark,

Thank you for your interesting comment. We understand your concern about the sentence " *individual project initiatives develop their own usage*"! We are not encouraging that each group develops their own standards, but rather they choose those that correspond best to their usage among the existing ones. What we absolutely want to avoid, is that no standard is used because recommended ones would be too complicated or not adapted to the need. In this case, we prefer to suggest several alternatives and leave the decision to the group, rather than imposing only one standard.

We take your suggestion about the format validator tools. Generally, when submitters deposit their files, we check with our tools if the standards are respected. That could be by inserting the data in databases with well written ETL tools. We will provide those tools to the data submitters in order to make them able to check by themselves the formats.

We are indeed very interested by the FAIR metrics. This is a good way to monitor the progress made by our community to share their data. The wheat community can be one use case for the tools that are developed by the FAIR Data Metrics working group.

Best regards,

Competing Interests: No competing interests were disclosed.

Reader Comment 23 Oct 2017

Mark Wilkinson, Medical Genetics, Centro de Biotecnología y Genómica de Plantas, CBGP (UPM-INIA), Spain

Dear Authors,

I enjoyed your article! One idea that came to-mind while I was reading it was that you may need some form of validation beyond what you describe in this manuscript. Effectively, validation of the implementation/adoption, not just the recommendation.

In your text you say:

"the WDI-WG favors a bottom-up approach rather than enforcing the choice of particular standards. Consequently, to begin with, it is better that individual project initiatives develop their own usage of the proposed recommendations and standards,"

As someone who also works with standards and their implementation, the phrase "*individual project initiatives develop their own usage"* makes me very nervous!! I know that free-for-all implementation choices are probably not what you meant by that phrase, but I bet that's what you will get, if you're encouraging bottom-up decisions on usage!

One possibility to avoid faulty implementation of standards is to adopt, as part of your mandate, the creation of lightweight validation tools. In some cases, this has been done for you (e.g. http://genometools.org/cgi-bin/gff3validator.cgi), but no doubt there will be other cases where you will have to build your own. If you require that your community members use these validation tools, before claiming that they are standards-compliant, I bet you will end up with a higher level of both quality and interoperability! Win Win! I think some indication that you are going to run QC on your community member's adherence to the proposed standards would strengthen this article.

One final note - The FAIR Data Metrics working group will soon be publishing a set of Metrics for testing the FAIRness of data (and an automated testing tool will soon follow). One of the ideas that came out of this working group was that individual communities would have their own standards and expectations - beyond the minimal and generic standards of FAIR - that they expect their community members to adhere-to. Maybe your standards adoption process described here could be used as an exemplar of this kind of FAIR Metrics extension? If you're interested, I would welcome one of the authors to contact me! (many of you know me already)

Thank you again for this nice commentary! Good luck!

Mark Wilkinson

Competing Interests: No competing interests were disclosed.