



Proposal number ¹ : 2012-02
Date of reception ¹ : 12/10/2012

EXPERT WORKING GROUP TOPIC SUBMISSION²

Expert Working Groups (EWG) are established where a particular topic of direct relevance to the Wheat Initiative would benefit from bringing together experts in that specific field. The EWG would provide them with a platform for discussion, information sharing, consideration of specific problems, identification of research priorities and gaps. The EWG should have clear objectives and these could include (but are not limited to) specific activities such as contributing to the development of the Wheat Initiative Strategic Research Agenda, producing a position paper for publication or addressing a particular challenge through a research programme. An output should be annual reports to the Scientific Board for dissemination to the Research Committee, the Institutions' Coordination Committee and the wheat research community through the Wheat Initiative website.

Expert Working groups are established following the attached flow diagram. Organised consortia addressing wheat research challenges can be endorsed by the Wheat Initiative as EWGs following the same process.

Topic title		
Construction of a network of expert for the development of an integrated Wheat Information System		
Proposer(s) ³		
Proposer 1	Name	Hadi
	Surname	Quesneville
	Institution	INRA
	Address	URGI, Centre INRA de Versailles, RD10, route de St Cyr, 78026 Versailles Cedex, FRANCE
	Email	Hadi.quesneville@versailles.inra.fr
	Telephone	+33 1 30 83 30 08
Proposer 2	Name	Mario
	Surname	Caccamo
	Institution	TGAC
	Address	Norwich Research Park, Norwich NR4 7UH, UK
	Email	mario.caccamo@tgac.ac.uk
	Telephone	+44 1603 450898
Summary		
We propose here to gather an expert group that will work together to coordinate data standard and exchanges and develop a framework to support an integrated Wheat Information System.		

¹ For Secretariat use only, do not fill

² Please send the completed form to H el ene Lucas, Wheat.Initiative@versailles.inra.fr

³ Add lines for other proposers if needed.

Our aim is to provide the international wheat research community with an easy access to wheat genetic information, genomic data and bioinformatics tools.

Detailed description (5 pages maximum)

Rationale

Today, wheat genetic and genomic data are dispersed among several databases. They are scattered, difficult to access and not always in compatible formats. The need is for interoperability, data comparison at a global scale, with requirements of integration of data from a myriad of source (e.g. genetic, phenotypic, genomic, agronomic, etc...).

Many recently funded large collaborative projects will soon provide huge amounts of data (e.g. WISP, Breedwheat, Speed, T-CAP, SeeD). There is an urgent need for standardizing and sharing data. This is particularly timely as the wheat genome and many closely related species are being sequenced providing an important base for data integration.

Coordination and cooperation between wheat bioinformatics teams are needed today to manage and exploit the expected deluge of data. The final objective is to develop an International Wheat Information System that captures in a single-entry point all the available data.. The construction of such as systems is however ambitious. We have therefore design a long-term development plan with several intermediary milestones that will gradually provide the wheat community, easier access to the data.

A report was commissioned by the *Wheat Initiative* scientific board following the proposal adopted by the G20 Agriculture Ministers in June 2011. The aim of this report was to offer recommendations on the best strategies to follow to develop an integrated Wheat Information System (hereafter called WheatIS).

See <http://www.wheatinitiative.org/sites/default/files/docs/wheat-info-system-report.pdf>

This report will serve as a starting point to the construction of the WheatIS that will be constructed collectively.

Description of the EWG aims

Our aim is to create and animate an EWG that brings together people representing different stakeholder groups and countries , willing to coordinate their efforts for providing the wheat scientific community with integrated genetic and genomic data. Cooperation, data standardization, and data sharing will be the main drivers of this group working at defining a strategy to build a Wheat IS. This expert working group should be also considered as a network where data and expertise exchanges are facilitated and enhanced.

Bioinformatics platforms will easily exchange their data to provide their users with up-to-date top-quality information. Users will also have access to enriched integrated and curated data from a central repository.

One major aim of the EWG will be to write a number of research proposals to obtain funds for implementation.

Expected deliverables/outputs of the EWG

Deliverables of the EWG will be:

1-The creation of a network of experts, collaborating to provide the scientific community with wheat genetic and genomic data.

2-The writing of a research proposal to get funds for the implementation of the WheatIS

3-Every year, an activity report will be delivered to follows the EWG activities.

4-Standards, protocols, and processes for data sharing will be established. Guidelines and recommendations will be disseminated and shared between bioinformatics teams and the scientific community. These documents will be accessed through a web site and regularly updated according to the evolution of the scientific fields.

5-Provide a web platform allowing the exchange of standardized data files, built on a collaborative and interoperable network of platforms, working together.

6-Provide an entry point for the wheat community to find available data through a full text search engine, allowing to dynamically search the central repository and the databases of the platform network.

7-Provide integrated data focusing on relevant data sets, chosen with the wheat scientific community.

Timeline of Activities

Instead of proposing a static view of the infrastructure, we present a dynamic model that will be adapted as the system implementation progresses, and where each step builds optimally on the previous infrastructure. The rationale is that bearing in mind the complexity of building a WheatIS, it will be preferable to build it progressively, learning from each phase by the successes and the failures and correcting the strategy accordingly. Users needs will be considered as well, as they will evolve in parallel. This iterative process will have a better chance to succeed, as its evolution will benefit from the community feedback at each step.

This project will be driven by this EWG, a network of expert created from volunteers willing to participate to the WheatIS project. Once the group created (up to 20 persons), they will elect for 4 years, a chair and a co-chair to animate the group.

During all the phases of this project, annual meeting will be planned to discuss the progresses, the needs and the future orientations. They will be organized in coordination with other international meetings for facilitating logistics and limiting travels cost. Regular teleconference calls (at least 3 per year) will maintain the contact between the participants. Working groups will interact monthly to work on specific topics. Activity reports will be provided to follows EWG activities (one per year).

We therefore propose to build the WheatIS in three main steps starting from a low-technology, easy-to-achieve infrastructure, towards a more ambitious integrated system.

- Step1 : Network building. (year 1-2)

At this stage, the WheatIS will be a web platform allowing the exchange of standardized data files. It will be possible to search the WheatIS metadata using keywords or full text searches. Indexation through Google web and Google scholar search engines will be used to allow scientist to find data through the Google interface without necessarily knowing the WheatIS portal.

- Step2 : Integrated virtual portal (year 3-4)

A full text search engine will be set up on the WheatIS portal, allowing to dynamically search local platform databases. The users will be able to connect to the WheatIS portal and type a keyword or a term that will be searched remotely in each database. Results will be provided as a brief summary of the matching data (e.g. Identifier, Name, Short description) with links to access the remotely hosted data.

- Step3 : Integrated database (year 5-6)

The final step will be to integrate the data in one single, centralised information system. Previous steps will help to provide a broad view of the available data. Being a major task, integration will focus on relevant data sets, chosen with the wheat scientific community. It will not replace the tools built in Steps 1 and 2, but will add a new browsing functionality, allowing users to navigate through data and explore their relationship. It will also answer complex queries involving data that hosted initially in different location (files, databases). This system will also be able to produce integrated, consolidated and consistent information, which could be exported as data files to feed analysis pipelines or other information systems.

Tools and technologies will undoubtedly evolve rapidly in the next coming years. The strategy presented here will be re-evaluated each year and the WheatIS development strategy adapted accordingly. In particular, the WheatIS EWG will follow the work of other international bioinformatics initiatives (such as Ontology, IAIC, TransPLANT) by developing synergies and collaborations. Members of this EWG must be also involved in these other initiatives.

For year 1 we plan to:

- Organize the first annual meeting in 2013.
- Write a report of the planned activities.
- Submit a research proposal to get funds for the implementation of the WheatIS

Alignment with the Wheat Initiative objectives

This initiative is aligned to the *Wheat Initiative* objective of providing an easy access to standardized data through a shared wheat information system. It will strengthen coordination of world wheat research, and will improve the sharing of resources, methods and expertise to improve and stabilise yields.

Potential links with other Wheat Initiative activities

The work of the EWG has important links with genomic selection, breeding, phenotyping, and genetic resources

Supporting countries/institutes

The report was written with the collaboration of the following international experts that support this EWG view:

<i>Name</i>	<i>Affiliation</i>	<i>Country</i>
Catherine Feuillet	INRA	France
Cesar Martinez	CIMMYT	International
Dave Edwards	University of Queensland	Australia
David Marshall	James Hutton Institute	UK
Doreen Ware	Gramene	USA
Eva Huala	TAIR	USA
Hadi Quesneville	INRA	France

Hirokazu Handa	NIAS	Japan
Jizeng Jia	CAAS	China
Jorge Dubcovsky	UC Davis	USA
Keith Edwards	University of Bristol	UK
Klaus Mayer	MIPS	Germany
Mario Caccamo	TGAC	UK
Paul Kersey	EBI-EMBL	International
Peter Langridge	University of Adelaide	Australia
Potential participating countries⁴		
<p>Major Wheat databases experts around the world need to be included in the EWG. Potential countries identified from the list of expert above are: Australia, Argentina, China, France, Germany, Japan, Spain, India, UK and USA. Participants from any country with wheat bioinformatics database resources or contributors of data willing to engage will be welcome.</p>		
Resources (budget requirement, potential funders, etc.)		
<p>With no specific funding, the WheatIS will rely mainly on existing infrastructures and the contribution from the scientists and bioinformaticians from the community (e.g from wheat platforms that will directly participate in this effort).</p> <p>However, resources will be important to insure the success of this initiative. Consequently, one important aim of this group will be to write project proposals to be funded by national or international funding agencies. At the minimum, to start the project, we ask the Wheat initiative 50keuros to organize the first EWG annual meeting in 2013. This will cover mainly travels and accommodation expenses of the EWG members.</p> <p>The central repository should probably be hosted by one of the existing platforms. A key prerequisite will be to guarantee the sustainability of this platform, and therefore the long-term availability of the following resources:</p> <ul style="list-style-type: none"> • Hardware infrastructure: petabyte-scale storage must be available with support for backups (such as mirror servers). Note that raw and curated data submitted to NCBI or EBI will be transferred regularly. Servers must be securely kept in a dedicated computer room with fire and intrusion protection. • Personnel: Highly qualified staff will be required. The infrastructure will need database managers, system and network managers, developers and data managers. Around 10 full time engineers/researchers should work for the WheatIS. • Quality control procedures: A management quality system must guarantee the correct functioning of the hosting platform. It will insure effectiveness of the procedures and protocols used the availability of the computer infrastructure, and the quality of the services. 		

⁴ Not limited to current members of the Wheat Initiative

- **Financial sustainability:** Long term service preservation and data accessibility must be guaranteed. Staff salaries, computer licences and maintenance fees, hardware renewal costs, travel and accommodation expenses to coordinating and scientific meetings must be covered.

Other comments

The development of the WheatIS is a very challenging objective. This EWG will be fundamental for the success of this endeavour that will heavily rely on open collaborations and cooperations between good willing people. The incremental implementation is also an important approach, as it has a better chance to reach the goal and answer the needs.

This EWG initiative is also an excellent opportunity to improve collaborations between bioinformatics platforms at the international level to leverage progress in data management, define missing standards in bioinformatics, develop new tools for complex genomes, stimulate collaborations between scientists (public, private) by providing an infrastructure to share their results, and accelerate crop improvement by providing wealth of consistent data sets.

Date of submission to the International Scientific Coordinator

10/12/2012

NB: relevant accompanying papers (concept note, articles, research project,...) could be joined to the pro-forma.