## Bretting, Peter

**From:** Peter Wenzl <peter.wenzl@croptrust.org>
**Sent:** Saturday, December 05, 2015 8:26 AM
**To:** Andreas Graner; David Marshall; Elizabeth Arnaud (Bioversity-France); Emily Marden; Bretting, Peter; Rajeev Varshney (ICRISAT-IN); Ruaraidh Sackville Hamilton (IRRI); Sarah Ayling; Susan McCouch
**Cc:** Daniele Manzella; Ruth Bastow; Wayne Powell (CGIAR Consortium)
**Subject:** Re: DivSeek meeting in Bonn
**Attachments:** DS_SC-2_15_2 (Updates from JFU Organizations).pdf

Dear DivSeek Steering Committee members,

We're sending you an additional input document with **updates from JFU organizations** for your review.

We are looking forward to welcoming you in Bonn. I will be at the Stern Hotel's reception at 18:45 on Monday, 7 Dec, to walk with you to the restaurant where we will have dinner.

On behalf of the Joint Facilitation Unit,

Best regards, Peter

On Tue, Dec 1, 2015 at 11:44 AM, Peter Wenzl <peter.wenzl@divseek.org> wrote:
Dear DivSeek Steering Committee members,

Please find attached the Agenda and other Documents for the **DivSeek Steering Committee meeting on December 8th** next week. The meeting will start at **9:00am** at the premises of the Crop Trust (Platz der Vereinten Nationen 7, 53113 Bonn).

In a follow-up message later this week, we will send you two additional Documents (No. 2, 3) with notes and updates from Susan and the JFU organizations.

As a reminder, here's a quick overview of the logistics:

- **Accommodation:** Accommodation has been booked at the Hotel Stern, Markt 8, 53111 Bonn, Germany. Accommodation and breakfast have been pre-paid by the Crop Trust. All extras are to be paid by the guest upon check out.
- **Transfer:** please see attached sheet for directions from the airport to the hotel, and from the hotel to the Crop Trust offices.
- **Per diem:** we'll pay you a per diem to cover costs for meals not provided by us. It should also cover incidentals such as transport to/from the airport.
- **Dinner on 7 Dec:** you're invited for a joint dinner at 19:00 on the evening before the meeting at the Ruland Restaurant, Bischhofsplatz 1, Bonn. This is a 2-min walk from the hotel.

Please do not hesitate to contact us in case you've any questions in regard to the agenda, documents or the logistics of the meeting.

Looking forward to meeting you next week!

On behalf of the Joint Facilitation Unit,

Peter

--

Peter Wenzl
DivSeek Liaison
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 126
Mobile: [redacted]
www.croptrust.org

*Securing our Food, Forever*

DS/SC-2/15/2

**Updates since the last meeting of the Steering Committee**

*Information note by the Joint Facilitation Unit*

This document contains information on relevant developments within the various work tracks of the organizations that serve the Joint Facilitation Unit.

The Committee is invited to appraise the developments in the light of the vision, mission and principles of DivSeek, as well as its current status and expected progress.

## 1. Global Plant Council

The Global Plant Council undertook an initial landscape of projects of relevance for DivSeek, which included those characterizing genebank accessions, web-based portals to access crop diversity data, and projects developing software or data standards for sharing information about crop diversity. A summary of this was presented at the Steering Committee (SC) meeting in Rome in May.

At this meeting, it was recommended that the initial landscape study be extended to highlight less visible projects, and those areas where DivSeek might facilitate coordination and help to leverage the sharing of information and expertise among different projects. To achieve this, an email was sent to all DivSeek member organizations to request information on additional projects.

To date an additional 25 projects have been identified; further information about these, as well as the previous projects, are provided in *Appendix 1*.

In summary, 26 infrastructure, tools and standards projects; 12 crop/species databases and portals; 29 germplasm evaluation projects; 2 reference genome projects, and 4 other projects have been identified.

To help make progress in the general areas of data standards and data integration, the Global Plant Council's Executive Director Ruth Bastow has been working with the NSF, DoE, USDA, BBSRC and DFG, to organize three events at the Plant and Animal Genome (PAG) meeting in San Diego in January 2016. These events comprise two open forum workshops, followed by a brainstorming session that will bring together 50–75 invited scholars from diverse fields. During these events participants will seek to address how investments in genomics, phenomics, and information management can be leveraged to advance frontiers in basic, translational, and applied plant biological research by supporting data integration and re-use. The series of events seeks to encourage cross-disciplinary discussions to develop key

priorities for strategic national and international investments that advance the development of data-driven approaches to predictive phenomics in plants.

In addition to the activities listed above Ruth Bastow has also presented the concept of DivSeek at a number of events and meetings including:

- XV Brazilian Congress of Plant Physiology
- International Plant Molecular Biology Congress
- GPC AGM

### 2. International Treaty on Plant Genetic Resources for Food and Agriculture

The Governing Body of the International Treaty held its Sixth Session on 5-9 October 2015.

The Chairperson of the Sixth Session, the Secretary of the Treaty and the Executive Director of the Global Crop Diversity Trust referred to DivSeek in their written reports to the Governing Body, and in their interventions during the Session. Delegates of Contracting Parties and civil society organizations considered DivSeek, under the agenda item on the Global Information System and other agenda items, such as on the enhancement of the Multilateral System of Access and Benefit-Sharing. This witnesses the high level of attention that the initiative is receiving from the constituency of the International Treaty.

Two Resolutions of the Governing Body contain direct references to DivSeek and its Joint Facilitation Unit.

In Resolution 3/2015 on the Global Information System, the Governing Body:

- noted that the Secretariat participates in the Joint Facilitation Unit of the DivSeek initiative, with the view to enabling synergies with the Global Information System in full respect of the provisions of the Treaty, and requested the Secretary to continue doing it;

- requested the Secretary to invite DivSeek stakeholders to report on the implications for the objectives of the Treaty of the technologies underlying the DivSeek initiative and to compile a synthesis report on this for consideration by the Governing Body at its Seventh Session.

DivSeek is also referred to in the multi-year programme of work of the Global Information System. In relation to the promotion of transparency on the rights and obligations of users for accessing, sharing and using crop germplasm associated information, and ways to exercise those rights and obligations within the Global Information System, the programme of work is to develop a deeper understanding of the applicability and impacts of models developed in other parts of the Treaty, such as the Multilateral System of Access and Benefit Sharing, and initiatives such as DivSeek.

The programme of work contains several other activities such the development of software and technical infrastructure, standard setting and capacity development that could overlap with the first programme of work of DivSeek, based on the discussions that have already taken place during 2015, and could benefit from close collaboration.

The full text of Resolution 3/2015 is appended to this document (*Appendix 2*).

In Resolution 8/2015, on policy guidance to the Global Crop Diversity Trust, the Governing Body:

- commended the Trust for its close collaboration with the Treaty Secretariat on the Global Information System, *inter alia* through the joint development of Genesys and the

DivSeek Initiative, in accordance with the Treaty's objectives and provisions, especially Article 17;

- encouraged the Trust to support, in its activities, synergies and complementarities with the Programme of Work on the Global Information System in recognition of the mandate and standard setting capacity of the Governing Body and to continue close collaboration with the Treaty Secretariat on the joint facilitation of the DivSeek Initiative.

Another Resolution of the Governing Body refers to a theme that may be relevant to DivSeek in its further progress. Resolution 1/2015 on measures to enhance the functioning of the Multilateral System of Access and Benefit-Sharing, emphasized the importance of collections that are fully characterized and evaluated, including specific traits, genomics and phenotypic characterization and appealed to both Contracting Parties to the Treaty and natural and legal persons to make them available in the Multilateral system together with the relevant characterization information. The Governing Body re-convened the working group that has been negotiating enhancement measures for the Multilateral System since 2013 and requested such Group, *inter alia*, to consider the issues regarding genetic information associated with the material accessed from the Multilateral System.

The Treaty Secretariat also contributed extensively to the research project conducted by the Arizona State University (ASU) which has been jointly commissioned by the Treaty Secretariat and the Trust. An update and summary of this contribution is contained in the Memorandum on the DivSeek Governance Research Project (document DS/SC-2/15/5).

Finally, the Treaty Benefit-sharing Fund financed a workshop for the multi-country construction of a test platform for the development and allocation of globally unique identifiers for rice germplasm, linking the MLS information infrastructure and the DivSeek repository, as well as the Platform for Co-development and Transfer of Rice Technologies. The workshop was attended by scientists, researchers and expert of the Global Biodiversity Information Facility (GBIF), the DivSeek JFU, the African Agricultural Technology Foundation (AATF), the Chinese Academy of Agricultural Sciences (CAAS), IRRI, Collima Agro, the Agriculture Research Center of Lao, PhilRice and Indonesian Agency for Agricultural Research and Development (IAARD), inter alia. It contributed to the establishment of a multi-country platform for application of Digital Object Identifiers for genetic resources of rice. This platform will also be connected with the infrastructure of the MLS and with a number of rice databases chosen in the context of the GLIS. The workshop was held in September 2015 at IAARD.

### 3. Global Crop Diversity Trust

Since the meeting in May, the Crop Trust has presented DivSeek at the *Annual General Meeting of the Genebank CRP* (Turkey) and conferences of the *Crop Science Society of America* (USA), the *European Plant Phenotyping Network* (Spain), and the *COST Action on Phenotyping Platforms* (Germany). An increasing number of people in the audience seem to be wondering when DivSeek will start generating tangible outputs and what these outputs will be. As a result, the Crop Trust has been reflecting on the scope and role of DivSeek so that it can generate useful outputs for the scientific community in the near future.

*What is the scope of DivSeek?*

Genebanks have long waited for an opportunity to unlock the genetic potential of their holdings. Genomics researchers have started to deploy cutting-edge technologies to systematically characterize crop diversity. Breeders would like to more effectively utilize

genetic variation to speed up genetic gains. DivSeek could be seen as the area where the interests of these three stakeholder groups *intersect*. Clearly, research in this area is rapidly accelerating and will continue to do so, irrespective of DivSeek. DivSeek, however, could add value to otherwise uncoordinated efforts by establishing and advocating a common set of tools, approaches, and platforms.

The Crop Trust's initial role was to encourage and assist these communities to gather around the common goal of unlocking the value of crop diversity. The resulting DivSeek partnership was intended to be a research-driven initiative for and by scientists, as there is much that needs to be done within the scientific arena to best utilize genetic resources for the benefit of food security and diet quality. We therefore believe that DivSeek needs to maintain a research-driven focus.

At the recent Governing Body (GB) meeting of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) DivSeek stakeholders were asked to investigate the implications of technologies for the objectives of the ITPGRFA.[1] Although this is an important topic for the ITPGRFA, we believe it goes beyond the originally intended scope of DivSeek. The discussions at the GB meeting illustrated how much DivSeek has become a topic of the political discourse among ITPGRFA stakeholders.

Given these discussions, it is important to reflect on who the key constituencies of DivSeek are. Was DivSeek meant to embrace the *entire* spectrum of PGRFA stakeholders, including R&D organizations, governments, farmers, NGOs, and consumers? We believe not. Convincing independently managed research projects to agree on a common set of standards, approaches, and actions is a challenging- enough task on its own! Other, better-prepared fora exist to deal with policy-related topics at the science-society interface where some of the conflicts seem to arise. Instead, the 'policy' component of DivSeek should primarily focus on how researchers and organizations participate in the Initiative, interact with each other, and share the data generated.

*The need for an effective organizational structure*

The governance-related input documents for the SC meeting deal with questions of leadership in the context of operational effectiveness and consensus among stakeholders. In our opinion, consensus on the Initiative's *purpose*, *principles*, and *objectives* is critical. Seeking consensus at an *operational* level, however, simply doesn't seem to be feasible if the Initiative wishes to become an innovation platform and keep up with a rapidly evolving science and technology landscape.

*The Crop Trust's role in DivSeek*

The Crop Trust has contributed substantial resources to gather DivSeek stakeholders, help the Steering Committee form, and support its operations by hiring a staff to work for the Joint Facilitation Unit. We suggest the Steering Committee now clarifies the identity and scope of the Initiative and develops an effective and independently funded operational unit.

We are a deliberately nimble organization and continuously challenge ourselves to focus on the essence of our mission. Going forward, we intend to shift our focus from DivSeek governance towards taking concrete actions[2] to advance what we consider to be the original

[1] The Governing Body requested "... *the Secretary to invite DivSeek stakeholders to report on the implications for the objectives of the Treaty of the technologies underlying the DivSeek initiative and to compile a synthesis report on this for consideration by the Governing Body at its Seventh Session*".

purpose of the Initiative, which deeply resonates with us: to **re-imagine the way genebanks conserve and add value to crop diversity** in the form of freely accessible information linked to accessions.[3] This will include assisting genebanks in their efforts to capture and 'digest' the growing body of data generated by 'DivSeek projects' to tackle questions like:

- How will genomic data change the way genebanks manage and promote their collections?
- What combination of skills will the next generation of genebank managers require?
- How to use this information to accelerate the development and implementation of Global Crop Conservation Strategies?

### 4. CGIAR Consortium Office

*No update received.*

[2] See examples outlined in the Potential Elements for a DivSeek Strategy document prepared for the May meeting in Rome.
[3] We also refer back to some earlier documents including the DivSeek White Paper and the Report from the meeting in Jan 2015.

**Appendix 1**

## DRAFT LANDSCAPE STUDY SUMMARY

| Project | Mission/Aims/Goals |
|---|---|
| 1000 Plant Genomes Project | The 1000 plants (oneKP or 1KP) initiative is an international multidisciplinary consortium that has generated large-scale gene sequencing data for over 1000 species of plants. |
| 150 Tomato Genome ReSequencing project | The aim of the 150 Tomato Genome ReSequencing project is to reveal and explore the genetic variation available in tomato. |
| 3,000 Rice Genome Project | The 3,000 Rice Genomes Project (3K RGP) is a collaborative, international research program that has sequenced 3,024 rice varieties from 89 countries. |
| African Orphan Crop Consortium | The consortium's goal is to sequence, assemble and annotate the genomes of 100 traditional African food crops, which will enable higher nutritional content for society over the decades to come. |
| AKER - Improve the competitiveness of sugar beet | The AKER program will first expand the genetic variability of sugar beet by providing a collection of genes from resources around the world. |
| AMAIZING | AMAIZING is designed to support the competitiveness of the French maize breeding sector as well as to meet society demands for sustainability and quality. |
| Australian Plant Phenomics Facility (APPF) | The APPF is a national facility available to Australian and international plant scientists. The APPF is based around automated image analysis of the phenotypic characteristics of extensive germplasm collections, and large breeding, mapping and mutant populations. It exploits recent advances in robotics, imaging and computing to enable sensitive, high throughput analyses to be made of plant growth and function. |
| BamYIELD | BamYIELD is a multidisciplinary research program for underutilized legumes using Bambara groundnut as an exemplar species. |
| BEAN-ADAPT | This project seeks to understand the genetic architecture of adaptation of the common bean (*Phaseolus vulgaris*; *Pv*) and its cross-compatible sister species, the runner bean (*Phaseolus coccineus*; *Pc*), in their centers of origin in the Americas and following their dispersal to Europe, as a model for future major environmental and socio-economic changes, such as increases in temperature, variability in rainfall, and new consumer preferences. |
| Big Data Infrastructure for Crop Genomics | This is a project to develop a crop bioinformatics platform, which enables users to access information on genetic and phenotypic variation, and perform analyses to explore gene expression and associations between genetic variation and traits. |
| BreedWheat | Aims at strengthening the competitiveness of the French wheat breeding sector, as well as to address the societal demands for sustainability, quality, and safety in agricultural production. |
| BRIDGE | Aims at evolving the Federal ex situ Genebank of IPK Gatersleben from a 'storage facility' of genetic resources to an integrated resource and information center, representing a one stop shop for facilitated and informed utilization of crop plant biodiversity. |
| Cacao Genome Database | The genomics, genetics and breeding resource for cacao improvement |
| CerealsDB | Provides a range of online resources for the study of the wheat genome. |
| Conservation and Divergence in the Wild Common Bean (*Phaseolus vulgaris*) Genome | This projects seeks to facilitate the utilization of such extra diversity in the breeding of improved bean varieties, whether dry beans or green beans. |
| CottonGen | Cotton community genomics, genetics and breeding database being developed to enable basic, translational and applied research in cotton. |
| Crop Ontology | This project has two primary objectives for the Integrated Breeding Platform: (1) publish online fully documented lists of breeding traits used by the Breeding |

| | |
|---|---|
| | Management System (BMS) for producing standard fieldbooks, and (2) support data analysis and integration of genetic and phenotypic data through harmonized breeders' data annotation. The project also offers a forum for scientists to discuss their variables, methods and scales of measurement, and fieldbooks. |
| ELIXIR-EXCELERATE | EXCELERATE funding will help ELIXIR coordinate and extend national and international data resources to ensure the delivery of world-leading life-science data services. It will support a pan-European training program, anchored in national infrastructures, to increase bioinformatics capacity and competency. It will also provide efficiencies in management and operation throughout the infrastructure, which is distributed amongst 17 countries. |
| EMPHASIS | EMPHASIS is a European infrastructure project, which aims to develop and provide access to infrastructures addressing multi-scale phenotyping for the analysis of genotype performance under diverse environmental conditions, and to quantify the diversity of traits contributing to performance in diverse environmental scenarios (i.e. plant architecture, major physiological functions and output, yield components and quality). Specifically, EMPHASIS will address the technological and organizational limits of European phenotyping, for a full exploitation of genetic and genomic resources available for crop improvement in changing climate. |
| EURISCO | EURISCO is a search catalog providing information about *ex situ* plant collections maintained in Europe. It is based on a European network of *ex situ* National Inventories (NIs). Currently, EURISCO comprises passport data about 1.1 million samples. |
| EU FruitBreedomics - Bridging the gap between genomics and fruit breeding | The objective of FruitBreedomics is to improve the efficiency of fruit breeding (apple and peach, but many tools will also benefit other species of the Rosaceae family) by bridging the gap between scientific genetics research and application in breeding. |
| EU Tree4Future (2012-2016) | Trees4Future is an Integrative European Research Infrastructure project that aims to integrate, develop and improve major forest genetics and forestry research infrastructures. It will provide the wider European forestry research community with easy and comprehensive access to currently scattered sources of information (including genetic databanks, forest modeling tools and wood technology labs) and expertise. |
| Genesys | Genesys is a global portal to access information about Plant Genetic Resources for Food and Agriculture (PGRFA). It is a gateway from which germplasm accessions from genebanks around the world can be easily found and ordered. |
| Genome Database for Rosaceae | A curated and integrated web-based relational database providing centralized access to Rosaceae genomics, genetics and breeding data and analysis tools to facilitate basic, translational and applied Rosaceae research. |
| Genomes To Fields (G2F) | A public research initiative to enable trait prediction from genotype and environment, thereby leading to increased maize production. |
| Genomic Standards Consortium's Biodiversity Working Group | International community to promote mechanisms that standardize the description of genomes and the exchange and integration of genomic data. |
| Geospatial database of collected crop samples and repository of expedition fieldbooks | Access to original passport data of more than 220,000 samples collected around the world during Bioversity International supported missions. |
| Germinate | Germinate is a generic plant genetic resources database and offers facilities to store both standard collection information and passport data along with more advanced data types such as phenotypic, genotypic and field trial data. |
| Global Information System (GLIS) of the ITPGRFA | A global information system to facilitate the exchange of information, based on existing information systems, on scientific, technical and environmental matters related to plant genetic resources for food and agriculture. |
| GrainGenes | GrainGenes, a database for Triticeae and Avena, is a comprehensive resource for molecular and phenotypic information for wheat, barley, rye, and other related species, including oat. |
| Gramene | Gramene is a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species. |
| GRIN | The Germplasm Resources Information Network (GRIN) web server provides |

| | |
|---|---|
| | germplasm information about plants, animals, microbes and invertebrates. This program is within the U.S. Department of Agriculture's Agricultural Research Service. |
| GRIN-Global | The GRIN-Global project's mission is to create a new scalable, version of the Germplasm Resource Information Network (GRIN), used by the USDA/ARS National Plant Germplasm System. The GRIN-Global database platform will be suitable for use by any interested genebank in the world. |
| iPlant Collaborative | The iPlant Collaborative is where scientists in all domains of life sciences can connect to public datasets, manage and store their own data and experiments, access high-performance computing, and share results with colleagues. |
| IRIC | Aims to provide access to well organized information about rice, and to facilitate communication and collaboration for the rice community, having germplasm diversity as a focal entry point. |
| KBase | KBase is an open platform for comparative functional genomics and systems biology for microbes, plants and their communities, and for sharing results and methods with other scientists. |
| Legume Federation | Aims to build a federation of legume databases through data standards, distributed development and comparative analysis, to support research across the legume family, and to support robust agriculture for a world that is significantly 'legume-fed'. |
| Legume Information System | A collaborative, community resource to facilitate crop improvement by integrating genetic, genomic, and trait data across legume species. |
| Maize GDB | Maintains and delivers genome sequences, maps, genotypes, and other genomic information for maize. |
| NEXTGEN Cassava | Aims to significantly increase the rate of genetic improvement in cassava breeding and unlock the full potential of cassava, a staple crop central to food security and livelihoods across Africa. The project will implement and empirically test a new breeding method known as Genomic Selection that relies on statistical modeling to predict cassava performance before field-testing, and dramatically accelerates the breeding cycle. |
| PeanutBase | A collaborative, community resource to facilitate crop improvement by integrating genetic, genomic, and trait data for peanuts to accelerate genetic improvement. |
| PEAMUST - Pea Multi-Stress adaptation and biological regulations for yield improvement and stability | The project will provide innovative solutions for genomic assisted selection, thanks to the production of new tools and innovative results such as the cloning of resistance genes, better understanding of the impact of the interactions between plant architecture, symbionts and stress tolerance, and the identification of regions of the genome involved in yield stability. |
| Phenomics Ontology Driven Data (PODD) | The PODD project will deliver a data management service that can handle multiple phenotyping platforms and data formats (text, image, video). The project will also provide the ability to manage a repository of associated metadata based on standard ontologies. A range of tools and other features will be developed to provide web-based discovery interfaces for users, external repositories, and services and support for the automatic capture and annotation of data and metadata from instrumentation, when possible. |
| Planteome | The goals of the Planteome project are to provide researchers and agricultural breeding programs with a common semantic framework and a focused set of comparative analysis tools to leverage the scientific value of the ever-expanding array of sequenced plant genomes and phenotype data. |
| PLantGenIE | The Plant Genome Integrative Explorer is a collection of interoperable web resources for searching, visualizing and analyzing genomics and transcriptomics data for different plant species. Currently it includes dedicated web portals for enabling in-depth exploration of poplar, Norway spruce, and Arabidopsis. |
| PlantPhenomeNET/Plant Phenomics Working Group | PhenomeNET is a cross-species phenotype network of phenotypic similarity. The use of ontologies, annotation standards, shared formats, and best practices for cross-taxon phenotype data analyses represents a novel approach to plant phenomics that enhances the utility of model genetic organisms and can be readily applied to species with fewer genetic resources and less characterized genomes. |
| RAPSODYN - Optimisation of the | Improving breeding of adapted varieties using genetics and genomics. Improving oil yield per hectare and nutrient use efficiency are the main final targets of |

| | |
|---|---|
| RAPeSeed Oil content anD Yield under low Nitrogen input | RAPSODYN, along with a better understanding of both hybrid value and genotype x N interaction. |
| Seeds of Discovery (SeeD) | The Seeds of Discovery (SeeD) project strives to invigorate maize and wheat genebanks to make them more attractive to breeders. We want to "see into seeds" (hence the SeeD acronym) to unlock the dormant genetic potential of maize and wheat genetic resources by providing breeders with a toolkit that enables their more targeted use in the development of high-yielding, climate-ready and resource-efficient cultivars. |
| Seeds4needs | Seeds for Needs' initiative works with >20,000 smallholder farmers in 11 countries to research how agricultural biodiversity can minimize the risks associated with climate change. |
| Smart tools for Prediction (and) Improvement of Crop Yield (SPICY) | Aims to develop a suite of tools based on molecular breeding to help breeders in predicting phenotypic response of genotypes for complex traits like yield under a range of environmental conditions. |
| SoyBase | Maintains and delivers genome sequences, maps, genotypes, and other genomic information for soybeans. |
| SUNRISE | SUNRISE aims to develop the economic competitiveness of the sunflower crop, an environmentally friendly agronomic solution in the context of climate and agricultural practices changes. It offers unique opportunities to accelerate the implementation of the genomic prediction of ideotypes. SUNRISE will combine this approach with the improvement of the sunflower hybrid breeding process: it aims to identify the loci which are the most involved in homeostasis, depending on the parental alleles, and then build new gene pools exhibiting between themselves the better specific combining ability for homeostasis, i.e. yield stability. |
| The 3000 Chickpea Genome Sequencing Initiative | An initiative to sequence the global chickpea collection to identify superior alleles and use them in breeding program for chickpea improvement. |
| The Biology of Rare Alleles in Maize and Its Wild Relatives (Panzea) | Investigating the connection between phenotype and genotype of complex traits in maize and its wild relative, teosinte. |
| CGIAR Research Program on Roots, Tubers and Bananas (RTB) | Joint initiative of CGIAR and partners to address global challenges for roots, tubers, and bananas. Among many other activities, it promotes the large-scale re-sequencing and high-throughput genotyping of genetic resources held in genebanks and breeding materials, as well as the development and interoperability of the underlying informatics infrastructure. |
| European Plant Phenotyping Network (EPPN) | EU-funded project with the goal to create structural and functional synergies between the leading plant phenotyping institutions in Europe. The project includes three pillars related to i) transnational access of the diverse user community to the existing European plant phenotyping facilities; ii) dedicated networking activities to provide a link between phenotyping experts, user communities and technology developers; iii) joint research activities to develop and adapt novel sensors and establish experimental as well as IT standards for application in plant phenotyping. |
| Genomic and Open-source Breeding Informatics Initiative (GOBII) | Aims to streamline the breeding of five staple crops – wheat, rice, maize, sorghum and chickpea, via the development of modular, open-source breeding software resources. |
| International Cocoa Germplasm Database (ICGD) | An information service for the cocoa research community. |
| International Plant Phenotyping Network (IPPN) | The International Plant Phenotyping Network (IPPN) is a non-profit association with large phenotyping hubs across the globe as members. The overarching goal is to increase the visibility and impact of plant phenotyping, and to use existing synergies by enabling communication and cooperation within the plant phenotyping community, from academia, industry, policy and general public. |
| Rice Diversity Project | Collaborative effort to explore the genetic basis of variation in rice and its wild ancestors. |
| TransPlant | A European Union-funded e-infrastructure to support the computational analysis of |

| | |
|---|---|
| | genomic data from crop and model plants. |
| Triticeae CAP (TCAP) | USDA's National Institute of Food and Agriculture (NIFA) has been awarded $25 million to develop new varieties of wheat and barley. |
| UK Barley Genome Sequencing | The UK Barley Genome Sequencing Consortium is part of a global effort to produce more resilient and higher yielding varieties of this important crop, resistant to pests and diseases. |
| USDA Pisum Diversity collection | The USDA Pisum Diversity Collection, UPDC, is a unique resource that represents the breadth of the genetic diversity of the genus in an inbred format that facilitates genetic study. The UPDC includes inbred accessions from the refined pea core collection, parent lines of USDA recombinant inbred (RIL) populations, accessions from U.S. pea breeding programs, *Pisum sativum* subspecies, and the wild species, *Pisum fulvum*, from diverse geographies worldwide. |
| W3B-PR-29-Indonesia | Multicountry construction of a test platform for the development and allocation of globally unique identifiers for rice germplasm, linking the MLS information infrastructure and the DivSeek repository |
| WHEALBI - Wheat and barley Legacy for Breeding Improvement | WHEALBI will develop and implement tools, methods and procedures to facilitate the characterization of wild relatives and local varieties of wheat and barley as sources of genes for use in crop improvement, exome capture and phenotyping. |
| Wheat Initiative | Provides a framework to establish strategic research and organization priorities for wheat research at the international level in both developed and developing countries. |
| WISP | The WISP consortium is a BBSRC-funded collaborative program for wheat improvement, which brings together experts from five UK institutions: John Innes Centre, National Institute for Agricultural Botany, University of Nottingham, University of Bristol and Rothamsted Research. |
| In situ conservation and use of crop wild relatives in three ACP countries of the SADC region | Aims to develop National Strategic Action Plans (NSAP) for the in situ conservation of crop wild relatives (CWR) in Mauritius, South Africa and Zambia, as a means to underpin regional food security and mitigate the adverse effects of climate change. |
| Data recording of coconut germplasm from CRI collection in the International Coconut Genetic Resources Database (CGRD | Aims to complete the characterization of data of the coconut accessions conserved by the Sri Lanka Coconut Genebank, and to test new methodologies to measure innovative criteria linked to the nutrition and quality of coconut fruit (sugar and fatty acid profiles). |
| G2P-SOL | More information to follow |
| REMBRANDT | More information to follow |

**Appendix 2**

## RESOLUTION 3/2015

## THE VISION AND THE PROGRAMME OF WORK ON THE GLOBAL INFORMATION SYSTEM

---

**THE GOVERNING BODY,**

***Recalling*** its previous decisions on the Global Information System and particularly Resolution 10/2013,

***Further recalling*** the interlinkages between the Global Information System of Article 17 and the Multilateral System of Access and Benefit-sharing as referred to in Article 13.2.a);

***Mindful of*** the role of the International Treaty and the Food and Agriculture Organization of the United Nations as caretakers of agro-biodiversity, including information related to PGRFA;

***Acknowledging*** the need to facilitate the documentation and dissemination of PGRFA available in the Multilateral System and its associated information to facilitate research, plant breeding and training subject to applicable law;

***Thanking*** the Government of Germany for the financial support provided for COGIS and the Government of Spain for their support to the PGRFA analysis tools during the 2014-2015 biennium;

1. ***Adopts*** the *Vision* and the *Programme of Work* contained in Annexes 1 and 2 respectively;
2. ***Decides*** to establish the Scientific Advisory Committee on the Global Information System of Article 17 with the terms of references contained in Annex 3. Subject to the availability of financial resources, the Committee will meet at least twice during the 2016-2017 biennium.
3. ***Requests*** the Secretary to implement the programme of work based on the recommendations of the Scientific Advisory Committee, and ***further requests*** the Secretary to submit to the Governing Body a progress report at each regular session and a proposal for review of the programme of work when appropriate, based on the recommendations of the Scientific Advisory Committee;
4. ***Invites*** Contracting Parties, other governments and stakeholders to provide the necessary resources to implement a set of pilot activities of the Programme of Work;
5. ***Notes that*** the Secretariat participates in the Joint Facilitation Unit of the DivSeek initiative, with the view to enabling synergies with the Global Information System in full respect of the provisions of the Treaty, and ***requests*** the Secretary to continue doing it;
6. ***Further requests*** the Secretary to invite DivSeek stakeholders to report on the implications for the objectives of the Treaty of the technologies underlying the DivSeek initiative and to compile a synthesis report on this for consideration by the Governing Body at its Seventh Session;

7. ***Requests*** the Secretary and the Contracting Parties to continue promoting initiatives to support national and regional programmes in the development and transfer of information technologies for, and data analysis of, PGRFA such as CAPFITOGEN and the Platform for Co-development and Transfer of Technologies in support of the programme of work as detailed in Annex 2.

8. ***Encourages*** the Secretary and the Contracting Parties to promote synergies and connections among existing information systems and national and regional information networks, as well as with the operations of the Multilateral System and the projects funded by the Benefit-sharing Fund for the documentation of PGRFA, as long as they operate in accordance with the provisions of the Treaty;

9. ***Requests*** the Secretary to design a monitoring and assessment mechanism on the usefulness and effectiveness of the Global Information System according to Article 17.1, subject to availability of resources, based on the recommendations of the Scientific Advisory Committee and present a draft to the next Session of the Governing Body;

10. ***Requests*** the Secretary to report on the above to the next Session of the Governing Body

*Annex 1*

## Vision for the Global Information System on PGRFA

***The Global Information System for PGRFA integrates and augments existing systems to create the global entry point to information and knowledge for strengthening the capacity for PGRFA conservation, management and utilization.***

The development of a truly effective Global Information System as foreseen in the International Treaty involves, *inter alia:* strengthening existing systems and, where gaps remain, establishing new systems and initiatives; promoting inter-connectivity among systems; and providing overarching mechanisms to ensure ready access to the information and services provided. This translates into the following objectives:

1-To create a web-based platform with use-oriented entry points to PGRFA information;

2-To provide a comprehensive overview and facilitate access to sources of PGRFA and associated information;

3-To promote and facilitate interoperability among existing systems by providing clear principles, technical standards and appropriate tools to support their operations in accordance to the principles and rules of the Treaty;

4-To promote transparency on the rights and obligations of users for accessing, sharing and using PGRFA associated information and to establish ways to exercise those rights and obligations within the Global Information System;

5-To create and enhance opportunities for communication and international and multidisciplinary collaboration to increase knowledge about and add value to PGRFA;

6-To provide capacity development and technology transfer opportunities for the conservation, management and use of PGRFA and associated information and knowledge paying special attention to the needs of developing countries;

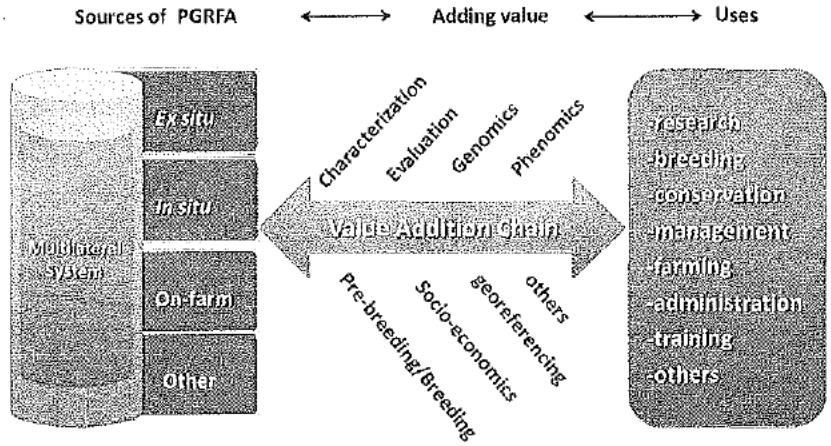
7-To create a mechanism to assess progress and monitor effectiveness of the Global Information System.

*Annex 2*

## Programme of Work on the Global Information System (2016-2022)

The Programme of Work will cover an initial period of six years. It will be implemented through a phased approach and funded through extra budgetary contributions.

The value chain of PGRFA, illustrating the components of and linkages between the three elements of the value chain, namely the sources of PGRFA, adding value, and using PGRFA.




**1-To create a web-based platform with use-oriented entry points to PGRFA information**

a. To set up the technical infrastructure needed for the development of the Platform;
b. To engage with FAO, the Convention on Biological Diversity and its Nagoya Protocol and other organizations to build on the experience accumulated on the development of global portals;
c. To define use case scenarios for target groups and set up mechanisms to get feedback from them;
d. To create a prototype of the global portal able to receive feedback from the users on the SMTA material;

**2-To provide a comprehensive overview and facilitate access to sources of PGRFA and associated information**

a. To create an index of sources of information, knowledge and other materials;
b. To strengthen the capacity of genebanks and other providers to document their holdings using traditional and modern methods and to facilitate access to that information and to the genetic resources in accordance with the provisions of the Treaty;
c. To enable recipients of PGRFA to make available to the MLS all non-confidential information according to applicable law that results from their research and development carried out on the material received, in compliance with their obligation under SMTA Art 6.9;
d. To allow quick access to information on the material available in the Multilateral System of Access and Benefit-sharing (MLS) at sample level;

e. To strengthen capacity to develop national and regional inventories and information systems and networks.

**3-To promote and facilitate interoperability among existing systems by providing clear principles, technical standards and appropriate tools to support their operations in accordance to the principles and rules of the Treaty;**

a. To develop a common standard for Permanent Unique Identifiers applied to PGRFA and an operational mechanism to promote the adoption of DOIs;
b. To develop further training and capacity development material, including e-learning material, in collaboration with other relevant organizations;
c. To recommend common standards for data and metadata and develop further standards (e.g. for phenotypic data) based on existing experiences in other sectors;
d. To establish functional connections with other initiatives relevant for the adoption of open data and standards to PGRFA;
e. To develop the technical standards required for interoperability between different PGRFA information management systems;

**4-To promote transparency on the rights and obligations of users for accessing, sharing and using PGRFA associated information and to establish ways to exercise those rights and obligations within the Global Information System;**

a. To analyse the institutional, organizational, policy and legal factors for PGRFA information access, sharing and use in the context of the Treaty's provisions, in particular Articles 12 and 13;
b. To understand the applicability and impacts of models developed in other parts of the Treaty, such as the Multilateral System of Access and Benefit Sharing, and initiatives such as DivSeek.

**5-To create and enhance opportunities for communication and international and multidisciplinary collaboration to increase knowledge about and add value to PGRFA**

a. To identify and create tools, mechanisms and opportunities for communication and collaboration with partners and users of the system (media, mailing lists, etc);
b. To conduct focused surveys with a wide range of users and validation methods;
c. To strengthen the linkages among stakeholders to concentrate research on high-priority germplasm jointly identified by them;

**6-To provide capacity development and technology transfer opportunities for the conservation, management and use of PGRFA and associated information and knowledge paying special attention to the needs of developing countries**

a. To convey and support regional meetings and scientific conferences related to new technologies and themes;
b. To provide access to training materials for capacity development;
c. To support the training of staff in areas such as taxonomy, information management and bioinformatics in collaboration with relevant partners;
d. To design mechanisms to promote training opportunities across institutions (training of trainers, match-making);
e. To provide training for the genebank managers of the future;

f. To facilitate transfer of relevant technologies to developing countries;
g. To raise awareness among stakeholders in the Global Information System on traditional knowledge relevant to PGRFA in accordance with the Treaty's provisions and in harmony with the Convention on Biological Diversity.

**7. To create a mechanism to assess progress and monitor effectiveness of the Global Information System**

a. To implement a feedback system for the portal in order to allow assessments of the usefulness and effectiveness of the Global Information System;
b. To promote periodic consultations among Contracting Parties, and stakeholders, users and providers, about the usefulness and effectiveness of the Global Information System.
   a.

*Annex 3*

**Terms of Reference for the Scientific Advisory Committee on Article 17**

Objectives

The Governing Body provides guidance on the development and strengthening of the Global Information System, to facilitate the exchange of information, based on existing information systems, on scientific, technical and environmental matters related to plant genetic resources for food and agriculture (PGRFA).

The Scientific Advisory Committee (Committee) shall advise the Secretary on:

- general recommendations on the development and deployment of the Global Information System and its components as adopted by the Governing Body;
- the discovery of new areas of work with potential impact on the System;
- the selection of pilot activities for the Global Information System and, upon request of the Secretary, other initiatives and actions to sustain the operation of the Global Information System, and the further update of the Programme of Work.

In particular, the Committee shall provide scientific advice to the Secretary on the following items:

1. the effectiveness and efficiency of the Global Information System as a mechanism to promote advice regarding scientific, technical and environmental cooperation on PGRFA matters;
2. the exchange of PGRFA-related information and the transfer of publicly available expertise, technology and scientific cooperation;
3. scientific and technical components of the Global Information Systems and its Programme of Work as recommended by the Governing Body;
4. cooperation with other relevant international and regional scientific and technical cooperation and technology transfer initiatives, including the Access and Benefit-sharing Clearing House Mechanism of the Nagoya Protocol;
5. means to facilitate the implementation of the Global Information System at the national level and the establishment of a Stakeholder Platform;
6. scientific, technical and environmental cooperation and benefits of all the Treaty activities, including the Multilateral System and Access and Benefit-sharing and the Programme of Work on Sustainable Use

Subject to the availability of financial resources, the Committee will hold two meetings per biennium.

Composition

The Committee is composed of:

- up to 2 scientific experts from each Region, nominated by the Vice-chairpersons of the Governing Body of each respective Region;
- 10 additional scientific and technical experts appointed by the Secretary including experts suggested by the regions and relevant stakeholders, taking into account the required range of technical expertise and regional balance as appropriate.

The members shall be selected for their scientific expertise and understanding of the Global Information System and the International Treaty, taking into account the need for specialized and in-depth expertise including: bioinformatics and molecular genetics; the 'omics', in particular genomics, phenomics and proteomics; management of environmental and geo-spatial data about plant genetic resources; scientific, taxonomy, crop wild relatives and genebank management and *ex situ*, *in situ* and on-farm conservation of PGRFA, technical cooperation; capacity-building; system integration, information exchange and data sharing; fair and equitable benefit sharing and legal expertise in the relevant international law and regulations; partnerships with other organizations, institutions and initiatives.

The Committee shall elect its Co-chairpersons from among the experts.

The Secretariat of the International Treaty will facilitate the work of the Committee and report to the Governing Body.

The mandate and the composition of the Committee may be renewed by the Governing Body at its Seventh Session.

**Bretting, Peter**

| | |
|---|---|
| **From:** | [redacted] on behalf of Susan McCouch <srm4@cornell.edu> |
| **Sent:** | Thursday, May 28, 2015 7:51 PM |
| **To:** | Andreas Graner (IPK); David Marshall (JHI); Elizabeth Arnaud (Bioversity); Emily Marden (UBC); Bretting, Peter; Rajeev Varshney (ICRISAT & GCP); Ruaraidh Sackville Hamilton (IRRI); Sarah Ayling (TGAC) |
| **Cc:** | Susan McCouch; Daniele Manzella (ITPGRFA); Peter Wenzl; Ruth Bastow (GPC) |
| **Subject:** | photo of DivSeek SC & JFU |
| **Attachments:** | DivSeek_SC & JFU_Rome_2015.JPG |

Dear All,

Thanks for all you contributed to our meeting today. Here is a photo of most of us, standing on the terrace of the FAO building at lunch time. We missed Sara, Emily and Ruaraidh, who stayed downstairs and did not make it into the photo.

Travel well,
Susan

--
Susan McCouch
Professor, Plant Breeding & Genetics
Cornell University
162 Emerson Hall
Ithaca, NY 14853-1901
Phone: +1 607-255-0420
Fax: +1-607-255-6683
Email: srm4@cornell.edu or mccouch@cornell.edu
Alternate Email: [redacted]

**Bretting, Peter**

| | |
|---|---|
| **From:** | Bretting, Peter |
| **Sent:** | Friday, May 29, 2015 1:17 AM |
| **To:** | Susan McCouch |
| **Cc:** | Andreas Graner (IPK); David Marshall (JHI); Elizabeth Arnaud (Bioversity); Emily Marden (UBC); Rajeev Varshney (ICRISAT & GCP); Ruaraidh Sackville Hamilton (IRRI); Sarah Ayling (TGAC); Susan McCouch; Daniele Manzella (ITPGRFA); Peter Wenzl; Ruth Bastow (GPC) |
| **Subject:** | Re: photo of DivSeek SC & JFU |

**Thanks Susan. Safe travels to all!**

**Peter**

**Peter Bretting**
**USDA/ARS, ONP**
**GWCC, 4-2212**
**5601 Sunnyside Avenue**
**Beltsville MD 20705**
**301-504-5541**
[redacted] **mobile**
**Peter.bretting@ars.usda.gov**

**> On May 29, 2015, at 1:50 AM, Susan McCouch <srm4@cornell.edu> wrote:**
**>**
**> Dear All,**
**>**
**> Thanks for all you contributed to our meeting today. Here is a photo of most of us, standing on the terrace of the FAO building at lunch time. We missed Sara, Emily and Ruaraidh, who stayed downstairs and did not make it into the photo.**
**>**
**> Travel well,**
**> Susan**
**>**
**> --**
**> Susan McCouch**
**> Professor, Plant Breeding & Genetics**
**> Cornell University**
**> 162 Emerson Hall**
**> Ithaca, NY 14853-1901**
**> Phone: +1 607-255-0420**
**> Fax: +1 607-255-6683**
**> Email: srm4@cornell.edu<mailto:srm4@cornell.edu> or mccouch@cornell.edu<mailto:mccouch@cornell.edu>**
**> Alternate Email:** [redacted] **mailto:** [redacted]
**> <DivSeek_SC & JFU_Rome_2015.JPG>**

**Bretting, Peter**

| | |
|---|---|
| **From:** | Bretting, Peter |
| **Sent:** | Sunday, May 31, 2015 3:39 PM |
| **To:** | Susan McCouch; Andreas Graner (IPK); David Marshall (JHI); Elizabeth Arnaud (Bioversity); Emily Marden (UBC); Rajeev Varshney (ICRISAT & GCP); Ruaraidh Sackville Hamilton (IRRI); Sarah Ayling (TGAC) |
| **Cc:** | Susan McCouch; Daniele Manzella (ITPGRFA); Peter Wenzl; Ruth Bastow (GPC) |
| **Subject:** | RE: photo of DivSeek SC & JFU |

**Thanks, Susan, my Steering Committee colleagues, the JHU, and FAO staff for a pleasant and productive meeting.**

**Peter**

**Peter Bretting**
**USDA/ARS Office of National Programs**
**Room 4-2212, Mailstop 5139**
**5601 Sunnyside Avenue**
**Beltsville, MD 20705-5139**
**Phone 1.301.504.5541**
**Fax 1.301.504.6191**
**Mobile Phone** [redacted]
**E-mail peter.bretting@ars.usda.gov**
**Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301**

**From:** [redacted] [mailto:[redacted] **On Behalf Of** Susan McCouch
**Sent:** Thursday, May 28, 2015 7:51 PM
**To:** Andreas Graner (IPK); David Marshall (JHI); Elizabeth Arnaud (Bioversity); Emily Marden (UBC); Bretting, Peter; Rajeev Varshney (ICRISAT & GCP); Ruaraidh Sackville Hamilton (IRRI); Sarah Ayling (TGAC)
**Cc:** Susan McCouch; Daniele Manzella (ITPGRFA); Peter Wenzl; Ruth Bastow (GPC)
**Subject:** photo of DivSeek SC & JFU

Dear All,

Thanks for all you contributed to our meeting today. Here is a photo of most of us, standing on the terrace of the FAO building at lunch time. We missed Sara, Emily and Ruaraidh, who stayed downstairs and did not make it into the photo.

Travel well,
Susan

--

Susan McCouch
Professor, Plant Breeding & Genetics
Cornell University
162 Emerson Hall
Ithaca, NY 14853-1901
Phone: +1 607-255-0420

Fax: +1 607-255-6683
Email: srm4@cornell.edu or mccouch@cornell.edu
Alternate Email: [redacted]

**Bretting, Peter**

**From:** Susan McCouch <susan.mccouch@divseek.org>
**Sent:** Friday, June 26, 2015 11:46 AM
**To:** srm4@cornell.edu
**Subject:** Note from the DivSeek Chair
**Attachments:** Note from the DivSeek Chair_150626.docx; Preliminary Landscape Study.xlsx

**The DivSeek Steering Committee gets going**

Dear DivSeek partners and colleagues interested in DivSeek,

On May 28th, the newly elected DivSeek Steering Committee (SC) gathered at the FAO premises in Rome. Members of the Joint Facilitation Unit (JFU) and additional staff of the International Treaty also participated in the first SC meeting.

As an initial step, the SC reviewed a draft that which takes stock of independently funded projects in areas of relevance for DivSeek. This 'project landscape' study identified around 50 (!) projects, including projects characterizing genebank accessions, web-based portals to access crop-diversity data, and projects developing software or data standards for sharing information about crop diversity.

The SC also began discussing components of a multi-year strategy and an initial work plan for DivSeek. This discussion was facilitated by a document containing a 'menu' of ideas and potential elements for such a strategy.

The current task of the SC is to identify a set of objectives and activities for the DivSeek initiative and a mechanism for funding and administering those activities. A major goal underlying DivSeek's strategic plan would be to augment the potential for many independent, stand-alone efforts to work together under a common umbrella to apply state-of-the-art genomic, phenomic, molecular and bioinformatics tools and strategies to characterize crop diversity and to integrate and share data and information. A second goal would be to enhance the utilization of crop diversity in plant breeding programs that seek to enhance local and global food and nutritional security.

Governance-related topics that are critical to success of the DivSeek initiative were also discussed, including private-sector engagement, recruitment of new members to expand DivSeek's constituency, and examination of the roles and responsibilities of the JFU, the SC and the Partners Assembly (PA) as the initiative evolves.

Among the next steps --

• The JFU was encouraged to expand and refine the 'project landscape' study, and to make results available online through the DivSeek website and as a peer-reviewed publication. Please see attached list of currently funded projects relevant to the DivSeek initiative and let us know of any others that you would like to see included in the survey (info@divseek.org).

• During the coming months, the SC, with support from the JFU, will elaborate a proposal for a multi-year DivSeek strategy. We expect to share this proposal and an initial work plan with DivSeek Partners at the next Assembly, in January 2016. The strategy will include ideas for working groups and workshops on key topics of broad interest.

• An independent governance-expert committee was convened to propose a framework for engaging with the private sector, outline guidelines for publishing DivSeek documents, clarify the governance structure of DivSeek and describe lines of communication and governance principles that will allow it to remain flexible and evolve in the future.

I take this moment to reflect on the mission of the DivSeek initiative, which is to help unlock the potential of crop diversity so it can be utilized to enhance the productivity, sustainability and resilience of crops and agricultural systems

throughout the world. The mission is multi-faceted and we count on the input and support of DivSeek's partners to help us move forward.

Susan McCouch

**Note from the DivSeek Chair: The DivSeek Steering Committee gets going**

Dear DivSeek partners and colleagues interested in DivSeek,

On May 28th, the newly elected DivSeek Steering Committee (SC) gathered at the FAO premises in Rome. Members of the Joint Facilitation Unit (JFU) and additional staff of the International Treaty also participated in the first SC meeting.

As an initial step, the SC reviewed a draft that which takes stock of independently funded projects in areas of relevance for DivSeek. This 'project landscape' study identified around 50 (!) projects, including projects characterizing genebank accessions, web-based portals to access crop-diversity data, and projects developing software or data standards for sharing information about crop diversity.

The SC also began discussing components of a multi-year strategy and an initial work plan for DivSeek. This discussion was facilitated by a document containing a 'menu' of ideas and potential elements for such a strategy.

The current task of the SC is to identify a set of objectives and activities for the DivSeek initiative and a mechanism for funding and administering those activities. A major goal underlying DivSeek's strategic plan would be to augment the potential for many independent, stand-alone efforts to work together under a common umbrella to apply state-of-the-art genomic, phenomic, molecular and bioinformatics tools and strategies to characterize crop diversity and to integrate and share data and information. A second goal would be to enhance the utilization of crop diversity in plant breeding programs that seek to enhance local and global food and nutritional security.

Governance-related topics that are critical to success of the DivSeek initiative were also discussed, including private-sector engagement, recruitment of new members to expand DivSeek's constituency, and examination of the roles and responsibilities of the JFU, the SC and the Partners Assembly (PA) as the initiative evolves.

Among the next steps --

• The JFU was encouraged to expand and refine the 'project landscape' study, and to make results available online through the DivSeek website and as a peer-reviewed publication. Please see attached list of currently funded projects relevant to the DivSeek initiative and let us know of any others that you would like to see included in the survey (info@divseek.org).

• During the coming months, the SC, with support from the JFU, will elaborate a proposal for a multi-year DivSeek strategy. We expect to share this proposal and an initial work plan with DivSeek Partners at the next Assembly, in January 2016. The strategy will include ideas for working groups and workshops on key topics of broad interest.

• An independent governance-expert committee was convened to propose a framework for engaging with the private sector, outline guidelines for publishing DivSeek documents, clarify the governance structure of DivSeek and describe lines of communication and governance principles that will allow it to remain flexible and evolve in the future.

I take this moment to reflect on the mission of the DivSeek initiative, which is to help unlock the potential of crop diversity so it can be utilized to enhance the productivity, sustainability and resilience of crops and agricultural systems throughout the world. The mission is multi-faceted and we count on the input and support of DivSeek's partners to help us move forward.

Susan McCouch

**Project Name**

Seeds of Discovery (SeeD)
3,000 Rice Genome Project
TransPlant
African Orpahn Crop Consortium
The 3000 Chickpea Genome Sequencing Initiative
The Genomic and Open-source Breeding Informatics Initiative (GOBII)
BRIDGE
Crop Ontology
Planteome
Crops for the Future - BamYIELD
Germinate
IRIC
WHEALBI - Wheat and barley Legacy for Breeding Improvement
WISP
CerealsDB
The CGIAR Research Program on Roots, Tubers and Bananas (RTB)
Seeds4needs
Geospatial database of collected crop samples and repository of expedition fieldbooks
iPlant Collaborative
NEXTGEN Cassava
GRIN-Global
Maize GDB
SoyBase
Legume Information System
PeanutBase
GrainGenes
Gramene
Cacao Genome Database
CottonDB
Genome Database for Rosaceae
KBase
BreedWheat
Wheat Initiative
The European Plant Phenotyping Network (EPPN)
The Biology of Rare Alleles in Maize and Its Wild Relatives (Panzea)
Phenomics Ontology Driven Data (PODD)
1000 Plant Genomes Project
Smart tools for Prediction (and) Improvement of Crop Yield (SPICY)
150 Tomato Genome ReSequencing project
The Rice Diversity Project
The International Cocoa Germplasm Database (ICGD)

W3B-PR-29-Indonesia, "Multicountry construction of a test platform for the development and allocation of globally unique identifiers for rice germplasm, linking the MLS information infrastructure and the DivSeek repository

Big Data Infrastructure for Crop Genomics

In situ conservation and use of crop wild relatives in three ACP countries of the SADC region

Data recording of coconut germplasm from CRI collection in the International Coconut Genetic Resources Amaizing

**Website/URL**

http://seedsofdiscovery.org/en/
http://gigadb.org/dataset/200001
http://www.transplantdb.eu/
http://africanorphancrops.org/
Not yet available
Not yet available
under development
www.cropontology.org
http://www.planteome.org/ (will be updated)
www.bambaragroundnut.org
http://ics.hutton.ac.uk/germinate/
http://iric.irri.org/home
http://www.whealbi.eu/
http://www.wheatisp.org/Consortium/WISP.php
http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php
http://www.rtb.cgiar.org/
http://www.bioversityinternational.org/seeds-for-needs/
http://bioversity.github.io/geosite/
http://www.iplantcollaborative.org/
http://www.nextgencassava.org/about.html
http://www.ars-grin.gov
http://www.maizegdb.org/
http://soybase.org/
http://legumeinfo.org/
http://peanutbase.org/
http://wheat.pw.usda.gov/GG3/
http://www.gramene.org/
http://www.cacaogenomedb.org/
http://www.cottondb.org/wwwroot/cdbhome.php
http://www.rosaceae.org/
http://kbase.us/
http://www.breedwheat.fr/
http://www.wheatinitiative.org/
http://www.plant-phenotyping-network.eu/eppn/home
http://www.panzea.org/
https://projects.ands.org.au/id/NEAT-PODD
https://sites.google.com/a/ualberta.ca/onekp/
http://www.spicyweb.eu/
http://www.tomatogenome.net/
http://ricediversity.org/
http://www.icgd.rdg.ac.uk/

http://www.planttreaty.org/sites/default/files/files/Third%20Call%20for%20Proposals-%20Projects%20approved%20for%20funding-for%20web.pdf

Not yet available
http://www.cropwildrelatives.org/sadc-cwr-project/
http://www.cogentnetwork.org/cgrd-version-6-0-test-version
http://www.amaizing.fr/index.php

**ıng, Peter**

**ım:** Bretting, Peter
**ent:** Friday, June 26, 2015 1:16 PM
**To:** Susan McCouch
**Subject:** RE: Note from the DivSeek Chair

**Thanks, Susan, for the carefully-written, informative summary of DivSeek progress,**

**Peter**

**Peter Bretting**
**USDA/ARS Office of National Programs**
**Room 4-2212, Mailstop 5139**
**5601 Sunnyside Avenue**
**Beltsville, MD 20705-5139**
**Phone 1.301.504.5541**
**Fax 1.301.504.6191**
**Mobile Phone** [redacted]
**E-mail peter.bretting@ars.usda.gov**
**Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301**

**From:** Susan McCouch [mailto:susan.mccouch@divseek.org]
**Sent:** Friday, June 26, 2015 11:46 AM
**To:** srm4@cornell.edu
**Subject:** Note from the DivSeek Chair

**The DivSeek Steering Committee gets going**

Dear DivSeek partners and colleagues interested in DivSeek,

On May 28th, the newly elected DivSeek Steering Committee (SC) gathered at the FAO premises in Rome. Members of the Joint Facilitation Unit (JFU) and additional staff of the International Treaty also participated in the first SC meeting.

As an initial step, the SC reviewed a draft that which takes stock of independently funded projects in areas of relevance for DivSeek. This 'project landscape' study identified around 50 (!) projects, including projects characterizing genebank accessions, web-based portals to access crop-diversity data, and projects developing software or data standards for sharing information about crop diversity.

The SC also began discussing components of a multi-year strategy and an initial work plan for DivSeek. This discussion was facilitated by a document containing a 'menu' of ideas and potential elements for such a strategy.

The current task of the SC is to identify a set of objectives and activities for the DivSeek initiative and a mechanism for funding and administering those activities. A major goal underlying DivSeek's strategic plan would be to augment the potential for many independent, stand-alone efforts to work together under a common umbrella to apply state-of-the-art genomic, phenomic, molecular and bioinformatics tools and strategies to characterize crop diversity and to integrate and share data and information. A second goal would be to enhance the utilization of crop diversity in plant breeding programs that seek to enhance local and global food and nutritional security.

Governance-related topics that are critical to success of the DivSeek initiative were also discussed, including private-sector engagement, recruitment of new members to expand DivSeek's constituency, and examination of the roles and responsibilities of the JFU, the SC and the Partners Assembly (PA) as the initiative evolves.

Among the next steps --

- The JFU was encouraged to expand and refine the 'project landscape' study, and to make results available online through the DivSeek website and as a peer-reviewed publication. Please see attached list of currently funded projects relevant to the DivSeek initiative and let us know of any others that you would like to see included in the survey (info@divseek.org).
- During the coming months, the SC, with support from the JFU, will elaborate a proposal for a multi-year DivSeek strategy. We expect to share this proposal and an initial work plan with DivSeek Partners at the next Assembly, in January 2016. The strategy will include ideas for working groups and workshops on key topics of broad interest.
- An independent governance-expert committee was convened to propose a framework for engaging with the private sector, outline guidelines for publishing DivSeek documents, clarify the governance structure of DivSeek and describe lines of communication and governance principles that will allow it to remain flexible and evolve in the future.

I take this moment to reflect on the mission of the DivSeek initiative, which is to help unlock the potential of crop diversity so it can be utilized to enhance the productivity, sustainability and resilience of crops and agricultural systems throughout the world. The mission is multi-faceted and we count on the input and support of DivSeek's partners to help us move forward.

Susan McCouch

**Bretting, Peter**

**From:** [redacted] on behalf of Susan McCouch <srm4@cornell.edu>
**Sent:** Thursday, July 09, 2015 12:43 PM
**To:** Bretting, Peter
**Cc:** Daniele Manzella (ITPGRFA); Peter Wenzl (GCDT); Ruth Bastow (GPC); Wayne Powell (CGIAR Consortium); Andreas Graner (IPK); David Marshall (JHI); Elizabeth Arnaud (Bioversity); Emily Marden (UBC); Rajeev Varshney (ICRISAT & GCP); Ruaraidh Sackville Hamilton (IRRI); Sarah Ayling (TGAC); Susan McCouch
**Subject:** article of interest to DivSeek
**Attachments:** Stephens_Big data_PloSBio_15.pdf

Hi Peter,

Thanks very much for bringing this article to my attention. I agree it will be of interest to DivSeek -- it certainly underscores many of the issues we are currently grappling with, both technically and sociologically.

I am cc'ing the SC and JFU on this msg, and suggest that we provide a link to the article on the DivSeek web site.

Thanks again,
Susan

On Thu, Jul 9, 2015 at 11:46 AM, Bretting, Peter <Peter.Bretting@ars.usda.gov> wrote:

**Hi Susan—might the attached article be of interest to the DivSeek Steering Committee?**

**Thanks,**

**Peter**

**Peter Bretting**

**USDA/ARS Office of National Programs**

**Room 4-2212, Mailstop 5139**

**5601 Sunnyside Avenue**

**Beltsville, MD 20705-5139**

**Phone 1.301.504.5541**

**Fax 1.301.504.6191**

**Mobile Phone** [redacted]

--
Susan McCouch
Professor, Plant Breeding & Genetics
Cornell University
162 Emerson Hall
Ithaca, NY 14853-1901
Phone: +1 607-255-0420
Fax: +1 607-255-6683
Email: srm4@cornell.edu or mccouch@cornell.edu
Alternate Email: [redacted]

PLOS | BIOLOGY

PERSPECTIVE

# Big Data: Astronomical or Genomical?

**Zachary D. Stephens[1], Skylar Y. Lee[1], Faraz Faghri[2], Roy H. Campbell[2], Chengxiang Zhai[3], Miles J. Efron[4], Ravishankar Iyer[1], Michael C. Schatz[5]*, Saurabh Sinha[3]*, Gene E. Robinson[6]***

1 Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, 2 Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, 3 Carl R. Woese Institute for Genomic Biology & Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, 4 School of Library and Information Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, 5 Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America, 6 Carl R. Woese Institute for Genomic Biology, Department of Entomology, and Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America

* mschatz@cshl.edu (MCS); sinhas@illinois.edu (SS); generobi@illinois.edu (GER)



OPEN ACCESS

**Citation:** Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. (2015) Big Data: Astronomical or Genomical?. PLoS Biol 13(7): e1002195. doi:10.1371/journal.pbio.1002195

**Published:** July 7, 2015



**Funding:** This research was supported, in part, by grant 1U54GM114838 awarded by NIGMS to SS through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative and by National Institutes of Health award (R01-HG006677) to MCS. ZDS and RKI were supported by NSF grant MRI13-37732 (S.S. Lumetta, PI). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** API, application programming interface; ASKAP, Australian Square Kilometre Array Pathfinder; CPU, central processing unit; ExAC, Exome Aggregation Consortium; ICGC, International Cancer Genome Consortium; I/O, input/output; NIH/NCBI, National Institutes of Health National Center for Biotechnology Information; SKA, Square Kilometre Array; SRA, Sequence Read Archive; TCGA, The

## Abstract

Genomics is a Big Data science and is going to get much bigger, very soon, but it is not known whether the needs of genomics will exceed other Big Data domains. Projecting to the year 2025, we compared genomics with three other major generators of Big Data: astronomy, YouTube, and Twitter. Our estimates show that genomics is a "four-headed beast"—it is either on par with or the most demanding of the domains analyzed here in terms of data acquisition, storage, distribution, and analysis. We discuss aspects of new technologies that will need to be developed to rise up and meet the computational challenges that genomics poses for the near future. Now is the time for concerted, community-wide planning for the "genomical" challenges of the next decade.

We compared genomics with three other major generators of Big Data: astronomy, YouTube, and Twitter. Astronomy has faced the challenges of Big Data for over 20 years and continues with ever-more ambitious studies of the universe. YouTube burst on the scene in 2005 and has sparked extraordinary worldwide interest in creating and sharing huge numbers of videos. Twitter, created in 2006, has become the poster child of the burgeoning movement in computational social science [6], with unprecedented opportunities for new insights by mining the enormous and ever-growing amount of textual data [7]. Particle physics also produces massive quantities of raw data, although the footprint is surprisingly limited since the vast majority of data are discarded soon after acquisition using the processing power that is coupled to the sensors [8]. Consequently, we do not include the domain in full detail here, although that model of rapid filtering and analysis will surely play an increasingly important role in genomics as the field matures.

Cancer Genome Atlas; WGS, whole genome shotgun.

To compare these four disparate domains, we considered the four components that comprise the "life cycle" of a dataset: acquisition, storage, distribution, and analysis (Table 1).

## Data Acquisition

The four Big Data domains differ sharply in how data are acquired. Most astronomy data are acquired from a few highly centralized facilities [9]. By contrast, YouTube and Twitter acquire data in a highly distributed manner, but under a few standardized protocols. Astronomy, YouTube, and Twitter are expected to show continued dramatic growth in the volume of data to be acquired. For example, the Australian Square Kilometre Array Pathfinder (ASKAP) project currently acquires 7.5 terabytes/second of sample image data, a rate projected to increase 100-fold to 750 terabytes/second (~25 zettabytes per year) by 2025 [9,10]. YouTube currently has 300 hours of video being uploaded every minute, and this could grow to 1,000–1,700 hours per minute (1–2 exabytes of video data per year) by 2025 if we extrapolate from current trends (S1 Note). Today, Twitter generates 500 million tweets/day, each about 3 kilobytes including metadata (S2 Note). While this figure is beginning to plateau, a projected logarithmic growth rate would suggest a 2.4-fold growth by 2025, to 1.2 billion tweets per day, 1.36 petabytes/year. In short, data acquisition in these domains is expected to grow by up to two orders of magnitude in the next decade.

For genomics, data acquisition is highly distributed and involves heterogeneous formats. The rate of growth over the last decade has also been truly astonishing, with the total amount of sequence data produced doubling approximately every seven months (Fig 1). The OmicsMaps catalog of all known sequencing instruments in the world [11] reports that currently there are more than 2,500 high-throughput instruments, manufactured by several different companies, located in nearly 1,000 sequencing centers in 55 countries in universities, hospitals, and other research laboratories. These centers range in size from small laboratories with a few instruments generating a few terabases per year to large dedicated facilities producing several petabases a year. (An approximate conversion factor to use in interpreting these numbers is 4 bases = 1 byte, though we will revisit this below.)

The raw sequencing reads used in most published studies are archived at either the Sequence Read Archive (SRA) maintained by the United States National Institutes of Health National Center for Biotechnology Information (NIH/NCBI) or one of the international counterparts. The SRA currently contains more than 3.6 petabases of raw sequence data (S1 Fig), which reflects the ~32,000 microbial genomes, ~5,000 plant and animal genomes, and ~250,000 individual human genomes that have been sequenced or are in progress thus far [12].

**Table 1. Four domains of Big Data in 2025.** In each of the four domains, the projected annual storage and computing needs are presented across the data lifecycle.

| Data Phase | Astronomy | Twitter | YouTube | Genomics |
|---|---|---|---|---|
| **Acquisition** | 25 zetta-bytes/year | 0.5–15 billion tweets/year | 500–900 million hours/year | 1 zetta-bases/year |
| **Storage** | 1 EB/year | 1–17 PB/year | 1–2 EB/year | 2–40 EB/year |
| **Analysis** | In situ data reduction | Topic and sentiment mining | Limited requirements | Heterogeneous data and analysis |
| | Real-time processing | Metadata analysis | | Variant calling, ~2 trillion central processing unit (CPU) hours |
| | Massive volumes | | | All-pairs genome alignments, ~10,000 trillion CPU hours |
| **Distribution** | Dedicated lines from antennae to server (600 TB/s) | Small units of distribution | Major component of modern user's bandwidth (10 MB/s) | Many small (10 MB/s) and fewer massive (10 TB/s) data movement |

doi:10.1371/journal.pbio.1002195.t001

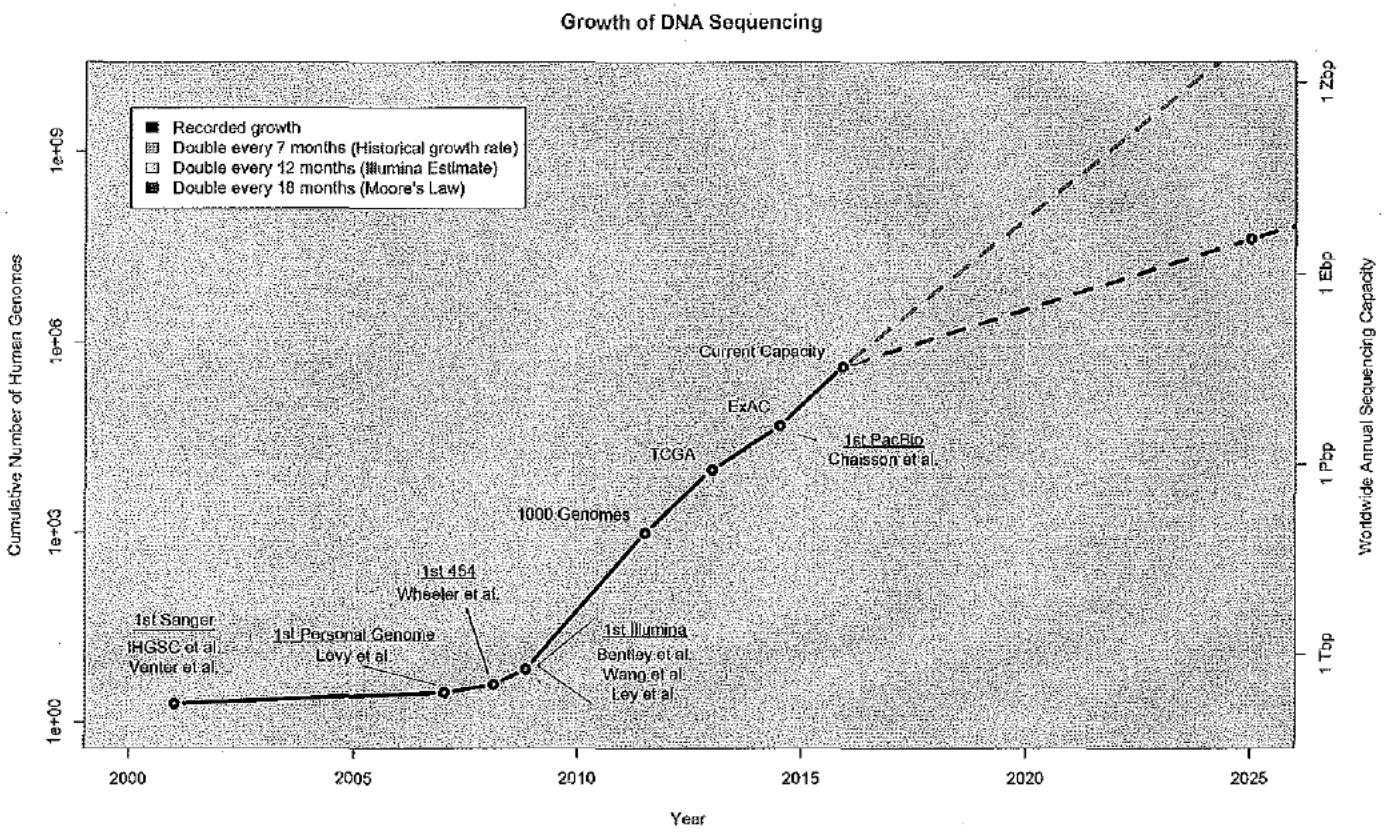



**Fig 1. Growth of DNA sequencing.** The plot shows the growth of DNA sequencing both in the total number of human genomes sequenced (left axis) as well as the worldwide annual sequencing capacity (right axis: Tera-basepairs (Tbp), Peta-basepairs (Pbp), Exa-basepairs (Ebp), Zetta-basepairs (Zbps)). The values through 2015 are based on the historical publication record, with selected milestones in sequencing (first Sanger through first PacBio human genome published) as well as three exemplar projects using large-scale sequencing: the 1000 Genomes Project, aggregating hundreds of human genomes by 2012 [3]; The Cancer Genome Atlas (TCGA), aggregating over several thousand tumor/normal genome pairs [4]; and the Exome Aggregation Consortium (ExAC), aggregating over 60,000 human exomes [5]. Many of the genomes sequenced to date have been whole exome rather than whole genome, but we expect the ratio to be increasingly favored towards whole genome in the future. The values beyond 2015 represent our projection under three possible growth curves as described in the main text.

doi:10.1371/journal.pbio.1002195.g001

However, the 3.6 petabases represent a small fraction of the total produced; most of it is not yet in these archives. Based on manufacturer specifications of the instruments, we estimate the current worldwide sequencing capacity to exceed 35 petabases per year, including the sixteen Illumina X-Ten systems that have been sold so far [13], each with a capacity of ~2 petabases per year [14].

Over the next ten years, we expect sequencing capacities will continue to grow very rapidly, although the project growth becomes more unpredictable the further out we consider. If the growth continues at the current rate by doubling every seven months, then we should reach more than one exabase of sequence per year in the next five years and approach one zettabase of sequence per year by 2025 (Fig 1, Table 1). Interestingly, even at the more conservative estimates of doubling every 12 months (Illumina's current own estimate [12]) or every 18 months (equivalent to Moore's law), we should reach exabase-scale genomics well within the next decade. We anticipate this sequencing will encompass genome sequences for most of the approximately 1.2 million described species of plants and animals [15]. With these genomes, plus those of thousands of individuals of "high value" species for energy, environmental, and agricultural reasons, we estimate that there will be at least 2.5 million plant and animal genome

sequences by 2025. For example, the genomics powerhouse BGI, in conjunction with the International Rice Research Institute and the Chinese Academy of Agricultural Sciences, has already sequenced 3,000 varieties of rice [16] and announced a massive project of their own to sequence one million plant and animal genomes [17]. The Smithsonian Institute also has similar plans to "capture and catalog all the DNA from the world's flora and fauna." There also will be genomes for several millions of microbes, with explosive growth projected for both medical and environmental microbe metagenomic sequencing [18,19].

These estimates, however, are dwarfed by the very reasonable possibility that a significant fraction of the world's human population will have their genomes sequenced. The leading driver of this trend is the promise of genomic medicine to revolutionize the diagnosis and treatment of disease, with some countries contemplating sequencing large portions of their populations: both England [20] and Saudi Arabia [21] have announced plans to sequence 100,000 of their citizens, one-third of Iceland's 320,000 citizens have donated blood for genetic testing [22], and researchers in both the US [23] and China [17] both aim to sequence 1 million genomes in the next few years. With the world's population projected to top 8 billion by 2025, it is possible that as many as 25% of the population in developed nations and half of that in less-developed nations will have their genomes sequenced (comparable to the current worldwide distribution of Internet users [24]).

We therefore estimate between 100 million and as many as 2 billion human genomes could be sequenced by 2025, representing four to five orders of magnitude growth in ten years and far exceeding the growth for the three other Big Data domains. Indeed, this number could grow even larger, especially since new single-cell genome sequencing technologies are starting to reveal previously unimagined levels of variation, especially in cancers, necessitating sequencing the genomes of thousands of separate cells in a single tumor [10].

Moreover, the technology used to sequence DNA is deployed creatively for other applications (e.g., transcriptome, epigenome, proteome, metabolome, and microbiome sequencing) necessitating generating new sequencing data multiple times per person to monitor molecular activity [25]. These applications require precise quantitative counts of sequencing reads to capture diversity of expression or diversity of abundances, thus requiring millions of reads to accurately estimate underlying distributions as they change over time. For medicine, just having the genome will not be sufficient: for each individual, it will need to be coupled with other relevant 'omics data sets, some collected periodically and from different tissues, to compare healthy and diseased states [26]. Computational challenges will increase because of dramatic increases in the total volume of genomic data per person, as will the complexities of integrating these diverse data sources to improve health and cure diseases. Genomics thus appears to pose the greatest challenges for data acquisition of the four Big Data domains.

## Data Storage

Data storage requirements for all four domains are projected to be enormous. Today, the largest astronomy data center devotes ~100 petabytes to storage, and the completion of the Square Kilometre Array (SKA) project is expected to lead to a storage demand of 1 exabyte per year. YouTube currently requires from 100 petabytes to 1 exabyte for storage and may be projected to require between 1 and 2 exabytes additional storage per year by 2025. Twitter's storage needs today are estimated at 0.5 petabytes per year, which may increase to 1.5 petabytes in the next ten years. (Our estimates here ignore the "replication factor" that multiplies storage needs by ~4, for redundancy.) For genomics, we have determined more than 100 petabytes of storage are currently used by only 20 of the largest institutions (S1 Table).

Projections of storage requirements for sequence data depend on the accuracy and application of the sequencing. For every 3 billion bases of human genome sequence, 30-fold more data (~100 gigabases) must be collected because of errors in sequencing, base calling, and genome alignment. This means that as much as 2–40 exabytes of storage capacity will be needed by 2025 just for the human genomes (S3 Note). These needs can be diminished with effective data compression [27], but decompression times and fidelity are a major concern in compressive genomics [28].

Are the emerging "third-generation" single-molecule sequencing technologies with much longer reads, such as those from Pacific Biosciences and Oxford Nanopore, a computational panacea? Though error rates currently are higher and throughput lower than short-read technologies, as they mature, these technologies are starting to be used to sequence and assemble nearly entire chromosomes [29]. This will minimize the need to oversample as much, and eventually, the raw sequence data may not need to be stored at all. However, eliminating the need to store raw sequence data and only retaining complete genomes will have relatively little impact overall—perhaps one or two orders of magnitude less data storage. More significant reductions in storage demand will come when improvements in sequencing accuracy and database comprehensiveness reach the point at which genome sequences themselves do not need to be stored, just the list of variants relative to a reference collection ("delta encoding") [30]. This works well for cataloging the simplest variants in a human genome, but it may not be as useful for complex samples, such as cancer genomes, that have many novel rearrangements and mutations. While certainly helpful, we thus do not expect long-read sequencing technology or delta encoding to solve the storage challenges for genome sequencing in 2025.

In contrast, we do see great opportunities for data reduction and real-time analysis of other 'omics analysis. For example, once sequencing becomes fast enough and the methods mature enough to correctly infer transcript expression levels in real time, we anticipate that raw RNA-seq reads will no longer be stored, except for specific research purposes. Already several such "streaming" algorithms have been published for this purpose, performing as well as or superior to their nonstreaming counterparts [31]. For RNA-seq and other 'omics applications, genomics will benefit greatly from the lessons learned in particle physics, in which in most cases raw data are discarded almost as fast as they are generated in favor of higher level and greatly compressed summaries.

Altogether, we anticipate the development of huge genomics archives used for storing millions of genomes along with the associated 'omics measurements over time. Ideally, these archives will also collect or be linked to the patient phenotypic data, especially disease outcomes and treatments provided to support retrospective analysis as new relationships are discovered. To make it practical to search and query through such vast collections, the data will be stored in hierarchical systems that make data and their statistical summaries available at different levels of compression and latency, as used in astronomy [32] and text analysis [33]. Thus, although total genomic data could far exceed the demands for the others, with the right new innovations the net requirements could be similar to the domains of astronomy and YouTube.

## Data Distribution

Astronomy, YouTube, Twitter, and genomics also differ greatly in data distribution patterns. The major bandwidth requirement of the SKA project is to get data from its 3,000 antennae to a central server, requiring as much as 600 terabytes/second [34]. The bandwidth usage of YouTube is relatively small for a single download and well supported by the average consumer's 10 Mbps connection, but aggregate needs worldwide are enormous, with estimates up to 240

petabytes/day (S4 Note) [35]. The distribution patterns of genomics data are much more heterogeneous, involving elements of both situations [36].

Genomic data are distributed in units spanning a wide range of sizes, from comparisons of a few bases or gene sequences to large multiterabyte bulk downloads from central repositories. For large-scale analysis, cloud computing is particularly suited to decreasing the bandwidth for distribution of genomic data [37] so that applications can run on remote machines that already have data [38]. Only small segments of code are uploaded and highly processed outputs are downloaded, thus significantly reducing the computing resources necessary for distribution.

But in addition to tailoring genomics applications for the cloud, new methods of data reliability and security are required to ensure privacy, much more so than for the other three domains. A serious breach of medically sensitive genomic data would have permanent consequences and could seriously hinder the development of genomic medicine. Homomorphic encryption systems, in which encrypted data can be analyzed and manipulated for certain controlled queries without disclosing the raw data, are currently too computationally expensive for widespread use, but these and related cryptographic techniques are promising areas of research [39].

## Data Analysis

Astronomy, YouTube, Twitter, and genomics differ most in computational requirements for data analysis. Astronomy data require extensive specialized analysis, but the bulk of this requirement is for in situ processing and reduction of data by computers located near the telescopes [40]. This initial analysis is daunting because of its real-time nature and huge data volumes but can often be effectively performed in parallel on thousands of cores. YouTube videos are primarily meant to be viewed, along with some automated analysis for advertisements or copyright infringements. Twitter data are the subject of intense research in the social sciences [41], especially for topic and sentiment mining, which is performed chiefly on textual "tweets" in the context of associated metadata (e.g., user demographics and temporal information).

Analysis of genomic data involves a more diverse range of approaches because of the variety of steps involved in reading a genome sequence and deriving useful information from it. For population and medical genomics, identifying the genomic variants in each individual genome is currently one of the most computationally complex phases. Variant calling on 2 billion genomes per year, with 100,000 CPUs in parallel, would require methods that process 2 genomes per CPU-hour, three-to-four orders of magnitude faster than current capabilities [42]. Whole genome alignment is another important form of genomic data analysis, used for a variety of goals, from phylogeny reconstruction to genome annotation via comparative methodologies. Just a single whole genome alignment between human and mouse consumes ~100 CPU hours [43]. Aligning all pairs of the ~2.5 million species expected to be available by 2025 amounts to 50–100 trillion such whole genome alignments, which would need to be six orders of magnitude faster than possible today.

Improvements to CPU capabilities, as anticipated by Moore's Law, should help close the gap, but trends in computing power are often geared towards floating point operations and do not necessarily provide improvements in genome analysis, in which string operations and memory management often pose the most significant challenges. Moreover, the bigger bottleneck of Big Data analysis in the future may not be in CPU capabilities but in the input/output (I/O) hardware that shuttles data between storage and processors [44], a problem requiring research into new parallel I/O hardware and algorithms that can effectively utilize them.

## The Long Road Ahead

Genomics clearly poses some of the most severe computational challenges facing us in the next decade. Genomics is a "four-headed beast"; considering the computational demands across the lifecycle of a dataset—acquisition, storage, distribution, and analysis—genomics is either on par with or the most demanding of the Big Data domains. New integrative approaches need to be developed that take into account the challenges in all four aspects: it is unlikely that a single advance or technology will solve the genomics data problem. Several key technologies that are most critically needed to support future solutions are discussed in Box 1.

In human health, the major needs are driven by the realization that for precision medicine and similar efforts to be most effective, genomes and related 'omics data need to be shared and compared in huge numbers. If we do not commit as a scientific community to sharing now, we run the risk of establishing thousands of isolated, private data collections, each too underpowered to allow subtle signals to be extracted. More than anything else, connecting these resources requires trust among institutions, scientists, and the public to ensure the collections will be used for medical purposes and not to discriminate or penalize individuals because of their genetic makeup.

Finally, the exascale data and computing centers that are emerging today to meet Big Data challenges in several domains (YouTube [50], Google [51], Facebook [51], and the National Security Agency [52]), are the result of far-sighted planning and commitment by the respective organizations. Now is the time for concerted, community-wide planning for the "genomical" challenges of the next decade.

### Box 1. Key Technological Needs for Big Data Genomics

#### (1) Acquisition

The most important need to sustain the explosive growth in genomic data acquisition is continued advances in sequencing technologies to reduce costs, improve throughput, and achieve very high accuracy. The current costs of ~US$1,000 per human genome begin to make it practical to sequence human genomes in large numbers, especially for critical medical treatments, but to scale to populations of hundreds of millions to billions of genomes, costs must be reduced by at least another one to two orders of magnitude or more. For many medical applications, the time for sequencing must also be reduced so that it can be completed in near real time, especially to rapidly diagnosis acute infections and conditions. Finally, to make a genome sequence most useful, it must be paired with automated methods to collect metadata and phenotype data, all according to appropriate standards so that data collected in one environment can be compared to those collected in another.

#### (2) Storage

The community needs to start designing and constructing data centers with fast, tiered storage systems to query and aggregate over large collections of genomes and 'omics data. There are new technologies on the horizon that will help support these needs, including 3-D memory, integrated computing technologies that overcome the I/O bottleneck, and networks that are two-to-five orders of magnitude faster because of optical switching [45,46]. Similarly, efficient compression and indexing systems are critical to make the best use out of each available byte while making the data highly accessible. We

also expect algorithmic developments that can represent large collections of personal genomes as a compact graph, making it more efficient and robust to compare one genome to many others. Beyond these approaches, we see the rise of streaming approaches to make on-the-fly comparisons that will allow us to rapidly discard data, especially for sequencing applications that use the sequence data as a means to infer abundances or other molecular activity.

## (3) Distribution

The most practical, and perhaps only, solution for distributing genome sequences at a population scale is to use cloud-computing systems that minimize data movement and maximize code federation [47]. New developments from companies such as Google, Amazon, and Facebook that include applications built to fit the frameworks of distributed computing efficient data centers and distributed storage and cloud computing paradigms will be part of the solution. Already, large cloud-based genomic resources are being developed using these technologies, especially to support the needs of the largest sequencing centers or to support the needs of large communities (BGI-cloud, TCGA, the International Cancer Genome Consortium [ICGC], etc.). To make these online systems most useful, the community needs to develop application programming interfaces (APIs) for discovering and querying large datasets on remote systems. The Global Alliance for Genomics and Health [48] and others are beginning to develop such standards for human genomic data, and we expect other communities to follow. Finally, authentication, encryption, and other security safeguards must be developed to ensure that genomic data remain private.

## (4) Analysis

Our ultimate goal is to be able to interpret genomic sequences and answer how DNA mutations, expression changes, or other molecular measurements relate to disease, development, behavior, or evolution. Accomplishing this goal will clearly require integration of biological domain expertise, large-scale machine learning systems, and a computing infrastructure that can support flexible and dynamic queries to search for patterns over very large collections in very high dimensions. A number of "data science technologies," including R, Mahout, and other machine learning systems powered by Hadoop and other highly scalable systems, are a start, but the current offerings are still difficult and expensive to use. The community would also benefit from libraries of highly optimized algorithms within a simple interface that can be combined and reused in many contexts as the problems emerge. Data science companies as well as open-source initiatives are already starting to develop such components, such as Amazon's recent "Amazon Machine Learning" prediction system. But because genomics poses unique challenges in terms of data acquisition, distribution, storage, and especially analysis, waiting for innovations from outside our field is unlikely to be sufficient. We must face these challenges ourselves, starting with integrating data science into graduate, undergraduate, and high-school curricula to train the next generations of quantitative biologists, bioinformaticians, and computer scientists and engineers [49].

## Supporting Information

**S1 Fig. Growth of GenBank.** The *y*-axis shows the total sequence in bp. (Blue = GenBank, red = whole genome shotgun [WGS] sequences.) Each line is double of the previous. The *x*-axis indicates time. Each line is 6 months after the previous. Source: http://www.ncbi.nlm.nih.gov/genbank/statistics.
(TIF)

**S1 Note. YouTube data estimates.**
(DOCX)

**S2 Note. Twitter data estimates.**
(DOCX)

**S3 Note. Human genomic data storage estimates for 2025.**
(DOCX)

**S4 Note. YouTube distribution statistics (current).**
(DOCX)

**S1 Table. Capacities of 20 major genomics institutions.** The number of sequencers as listed from OmicsMaps.com and their storage capacities from the listed citation. These 20 institutions alone collectively have more than 100 PB of storage available.
(DOCX)

## Acknowledgments

We thank V. Jongeneel, M. Baker, and the anonymous reviewers for comments that improved the manuscript and the Illinois CompGen Initiative for providing the collaborative structure that fostered the idea for this analysis. We would also like to thank all of the participants of the 2014 Keystone Symposium on "Big Data in Biology" (organized by Lincoln D. Stein, Doreen Ware, and MCS) for their inspiration for this analysis.

## References

1. Mole, B. The gene sequencing future is here. 2014; http://www.sciencenews.org/article/gene-sequencing-future-here.
2. Robinson G.E., et al., Creating a Buzz About Insect Genomes. Science, 2011. 18: 1386.
3. Abecasis G.R., et al., An integrated map of genetic variation from 1,092 human genomes. Nature, 2012. 491(7422): 56–65. doi: 10.1038/nature11632 PMID: 23128226
4. Chin L., Andersen J.N., and Futreal P.A., Cancer genomics: from discovery science to personalized medicine. Nature medicine, 2011. 17(3): 297–303. doi: 10.1038/nm.2323 PMID: 21383744
5. Exome Aggregation Consortium. Exome Aggregation Consortium ExAC Browser. 2015; http://exac.broadinstitute.org/.
6. Giles J., Computational social science: Making the links. Nature, 2012. 488(7412): 448–50. doi: 10.1038/488448a PMID: 22914149
7. Pak, A. and P. Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining, in LREC. 2010. p. 1320–1326.
8. O'Luanaigh, C. Animation shows LHC data processing. 2013; http://home.web.cern.ch/about/updates/2013/04/animation-shows-lhc-data-processing.
9. Newman, R. and J. Tseng. Cloud Computing and the Square Kilometre Array. 2011; http://www.skatelescope.org/uploaded/8762_134_Memo_Newman.pdf.
10. IBM Research. Square Kilometer Array: Ultimate Big Data Challenge. 2013; http://www.skatelescope.org/uploaded/8762_134_Memo_Newman.pdf.

11. Omicsmap. Next Generation Genomics: World Map of High-throughput Sequencers. 2015; http://omicsmaps.com/.
12. Regalado, A. EmTech: Illumina Says 228,000 Human Genomes Will Be Sequenced This Year. MIT Technology Review 2014 [cited 2015 April 28, 2015]; http://www.technologyreview.com/news/531091/emtech-illumina-says-228000-human-genomes-will-be-sequenced-this-year/.
13. AllSeq. HiSeq X Ten. 2015 [cited 2015 April 1, 2015]; http://allseq.com/x-ten.
14. Illumina. HiSeq X Series of Sequencing Systems. 2015 [cited April 28, 2015]; http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet-hiseq-x-ten.pdf.
15. Mora C., et al., How Many Species Are There on Earth and in the Ocean? PLoS Biology, 2011; 9: e1001127. doi: 10.1371/journal.pbio.1001127 PMID: 21886479
16. Li J.Y., Wang J., and Zeigler R.S., The 3,000 rice genomes project: new opportunities and challenges for future rice research. GigaScience, 2014. 3: 8. doi: 10.1186/2047-217X-3-8 PMID: 24872878
17. Zhu J., A year of great leaps in genome research. Genome medicine, 2012. 4(1): 4. doi: 10.1186/gm303 PMID: 22293069
18. Eisen J.A., Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. PLoS Biology, 2007. 5(3): e82. PMID: 17355177
19. Gevers D., et al., The Human Microbiome Project: a community resource for the healthy human microbiome. PLoS Biology, 2012. 10(8): e1001377. doi: 10.1371/journal.pbio.1001377 PMID: 22904687
20. Genomics England. The 100,000 Genomes Project. 2015; http://www.genomicsengland.co.uk/the-100000-genomes-project/.
21. Briggs, H (2013) Hundred thousand genomes to be mapped in Saudi Arabia. BBC News. http://www.bbc.com/news/health-25216135
22. Sulem P., et al., Identification of a large set of rare complete human knockouts. Nature genetics, 2015; 47: 448–452. doi: 10.1038/ng.3243 PMID: 25807282
23. Kaiser, J. White House fleshes out Obama's $215 million plan for precision medicine. Science Insider 2015; http://news.sciencemag.org/biology/2015/01/white-house-fleshes-out-obama-s-215-million-plan-precision-medicine.
24. Internet World Stats. 2015; http://www.internetworldstats.com/stats.htm.
25. Soon W.W., Hariharan M., and Snyder M.P., High-throughput sequencing for biology and medicine. Molecular systems biology, 2013. 9: 640. doi: 10.1038/msb.2012.61 PMID: 23340846
26. Chen R., et al., Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell, 2012. 148(6): 1293–307. doi: 10.1016/j.cell.2012.02.009 PMID: 22424236
27. Hsi-Yang Fritz M., et al., Efficient storage of high throughput DNA sequencing data using reference-based compression. Genome research, 2011. 21(5): 734–40. doi: 10.1101/gr.114819.110 PMID: 21245279
28. Loh P.-R., Baym M., and Berger B., Compressive genomics. Nature Biotechnology, 2012. 30: 627–630. doi: 10.1038/nbt.2241 PMID: 22781691
29. Koren S. and Phillippy A.M., One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. Current opinion in microbiology, 2015. 23C: 110–120.
30. Christley S., et al., Human genomes as email attachments. Bioinformatics, 2009. 25(2): 274–5. doi: 10.1093/bioinformatics/btn582 PMID: 18996942
31. Patro R., Mount S.M., and Kingsford C., Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nature biotechnology, 2014. 32(5): 462–4. doi: 10.1038/nbt.2862 PMID: 24752080
32. Golden A., Djorgovski S., and Greally J., Astrogenomics: big data, old problems, old solutions? Genome Biology, 2013. 8: 129.
33. Djoerd Hiemstra C.H., Brute Force Information Retrieval Experiments using MapReduce'. ERCIM News, 2012. 89: 31–32.
34. Smith, B., Data Transport for the Square Kilometre Array, in UbuntuNet Alliance Annual Conference. 2012. 15–22.
35. Global Internet Phenomena Report. 2013; http://www.sandvine.com/downloads/general/global-internet-phenomena/2013/2h-2013-global-internet-phenomena-report.pdf.
36. Sboner A., et al., The real cost of sequencing: higher than you think! Genome Biology, 2011. 12: 125. doi: 10.1186/gb-2011-12-8-125 PMID: 21867570
37. Marx, V., Drilling into big cancer-genome data. Nature Methods, 2013. 10: p. 293–297.
38. Baker M., Next-generation sequencing: adjusting to data overload. Nature Methods, 2010. 7: 495–499.

39. Erlich Y. and Narayanan A., Routes for breaching and protecting genetic privacy. Nature reviews. Genetics, 2014. 15(6): 409–21. doi: 10.1038/nrg3723 PMID: 24805122

40. Norris, R.P., Data Challenges for Next-generation Radio Telescopes. 2011.

41. Kumar S., Morstatter F., and Liu H., Twitter Data Analytics. 2014: Springer.

42. Langmead B., et al., Searching for SNPs with cloud computing. Genome Biol, 2009. 10(11): R134. doi: 10.1186/gb-2009-10-11-r134 PMID: 19930550

43. Kurtz S., et al., Versatile and open software for comparing large genomes. Genome Biol, 2004. 5(2): R12. PMID: 14759262

44. Trelles O., et al., Big data, but are we ready? Nature reviews. Genetics, 2011. 12(3): 224.

45. Loh, G.H. 3D-stacked memory architectures for multi-core processors. in ACM SIGARCH. 2008.

46. Chan, V.W., et al. Optical flow switching: An end-to-end "UltraFlow" architecture. in 15th International Conference on Transparent Optical Networks (ICTON). 2013. IEEE. doi: 10.1109/ICTON.2013.6602704

47. Schatz M.C., Langmead B., and Salzberg S.L., Cloud computing and the DNA data race. Nature biotechnology, 2010. 28(7): 691–3. doi: 10.1038/nbt0710-691 PMID: 20622843

48. Global Alliance for Genomics and Health. Global Alliance for Genomics and Health. 2015 April 28, 2015]; http://genomicsandhealth.org/.

49. Schatz, M.C., Computational thinking in the era of big data biology. Genome Biology, 2012. 13(11): 177.

50. Hollis, C. EMC's Record Breaking Product Launch. 2011 April 28, 2015]; http://chucksblog.emc.com/chucks_blog/2011/01/emcs-record-breaking-product-launch.html.

51. Hoff, T. How Google Backs up the Internet along with Exabytes of other data. 2014; http://highscalability.com/blog/2014/2/3/how-google-backs-up-the-internet-along-with-exabytes-of-othe.html.

52. Daily Kos. Utah Data Center stores data between 1 exabyte and 1 yottabyte. 2013; http://www.dailykos.com/story/2013/08/05/1228923/-Utah-Data-Center-stores-data-between-1-exabyte-and-1-yottabyte.

**Bretting, Peter**

**From:** [redacted] on behalf of Susan McCouch <srm4@cornell.edu>
**Sent:** Tuesday, November 03, 2015 10:03 AM
**To:** Andreas Graner (IPK); David Marshall (JHI); Elizabeth Arnaud (Bioversity); Emily Marden (UBC); Bretting, Peter; Rajeev Varshney (ICRISAT & GCP); Ruaraidh Sackville Hamilton (IRRI); Sarah Ayling (TGAC); Susan McCouch
**Cc:** Peter Wenzl; Ruth Bastow; Powell, Wayne (CGIAR Consortium); Daniele Manzella
**Subject:** Please review: Note to DivSeek Partners re January gathering in San Diego
**Attachments:** Draft re DivSeek Meeting in Jan_2016_151103.docx

Dear SC members,

After much deliberation about whether to convene a DivSeek Partner's Assembly (PA) in San Diego during January 2016, the JFU and I have reached the conclusion that it would be more constructive to invite PA members to participate in a workshop to discuss the goals and identity of the DivSeek initiative during this formative stage.

Please review the attached draft letter and get back to me (with a cc to the entire JFU) with your comments by Friday, Nov. 6.

We will do our best to accommodate your comments and suggestions, and plan to send a revised letter to the entire PA early on Saturday, Nov. 7.

We look forward to further discussions about DivSeek goals and identity, and plans for a PA later in the year during our SC meeting in Bonn on Dec. 8.

Thanks and best wishes,
Susan

--
Susan McCouch
Professor, Plant Breeding & Genetics
Cornell University
162 Emerson Hall
Ithaca, NY 14853-1901
Phone: +1 607-255-0420
Fax: +1 607-255-6683
Email: srm4@cornell.edu or mccouch@cornell.edu
Alternate Email: [redacted]

Dear Partners,

This year, the DivSeek intiative will host a round-table discussion about data-sharing and governance on the **afternoon of January 8, 2016 in San Diego**, and we invite you to join. Please hold the date, and we will follow up shortly with details about the venue.

In the mean time, I provide some updates on DivSeek activities. The DivSeek Charter has now been endorsed by 59 organizations, and since the Partner's Assembly (PA) meeting in January 2015 in San Diego, eight Steering Committee (SC) members were elected through electronic voting and the first SC meeting was held in Rome in May 2015.

According to the Charter, the DivSeek initiative aims to (1) facilitate networking among partners interested in the application of state-of-the-art genotyping and phenotyping technologies to deepen our understanding of crop diversity, (2) promote the development of common standards to link germplasm with characterization data and to enable interoperability among information systems, (3) develop an annual work plan with a budget and a resource mobilization plan, and (4) organize capacity-building workshops.

In keeping with these stated goals and principles, DivSeek has undertaken a landscaping study to take stock of ongoing research that is relevant to the initiative, endorsed the use of permanent unique identifiers as standards for tracking germplasm, and convened two studies to help identify an appropriate governance structure that will permit the initiative to define and execute a work plan and seek appropriate levels of external funding.

**Landscaping Study.** A Landscaping Study has documented ~50 funded initiatives related to the DivSeek agenda currently underway around the world <divseek.org/landscape>. These projects focus on evaluating over a dozen crops at the genotypic and phenotypic levels, developing information systems, software tools and data standards. Hosting links to these projects will help network researchers and facilitate data exchange and sharing of lessons learned.

**Permanent Unique Identifiers (PUID).** To help establish a set of core standards to facilitate data exchange and enable data integration and interoperability among diverse information systems, DivSeek endorses the use of PUID for crop germplasm; PUID will also enable phenotypic and genotypic datasets to be linked back to specific sources of germplasm.

**Governance.** A team at Arizona State University (ASU) is exploring lessons learned about governance and information-sharing from various community-driven initiatives involving genetics research. An SC subcommittee on governance is exploring guidelines to improve the operational efficiency of the DivSeek initiative that will enable it to be more inclusive, responsive, and to launch externally funded projects.

**Capacity Building.** Members of the SC have been exploring the possibility of capacity building workshops that would be endorsed by DivSeek, including training in the use of ontologies and standards for integration of phenotyping data at the lab and field scale.

We hope to see you at the Divseek workshop on Jan. 8 in San Diego and look forward to your input.

Regards and best wishes,
Susan McCouch

## Bretting, Peter

**From:** Bretting, Peter
**Sent:** Tuesday, November 03, 2015 2:27 PM
**To:** 'Susan McCouch'
**Cc:** Peter Wenzl; Ruth Bastow; Daniele Manzella; Wayne Powell (CGIAR Consortium)
**Subject:** RE: Please review: Note to DivSeek Partners re January gathering in San Diego
**Attachments:** 2015 Draft re DivSeek Meeting in Jan_2016_151103 PKB edits.docx

**Hi Susan—attached are some suggested edits.**

**Thanks, much appreciated!**

**Peter**

**Peter Bretting**
**USDA/ARS Office of National Programs**
**Room 4-2212, Mailstop 5139**
**5601 Sunnyside Avenue**
**Beltsville, MD 20705-5139**
**Phone 1.301.504.5541**
**Fax 1.301.504.6191**
**Mobile Phone** [redacted]
**E-mail peter.bretting@ars.usda.gov**
**Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301**

**From:** [redacted] [mailto [redacted] **On Behalf Of** Susan McCouch
**Sent:** Tuesday, November 03, 2015 10:03 AM
**To:** Andreas Graner (IPK); David Marshall (JHI); Elizabeth Arnaud (Bioversity); Emily Marden (UBC); Bretting, Peter; Rajeev Varshney (ICRISAT & GCP); Ruaraidh Sackville Hamilton (IRRI); Sarah Ayling (TGAC); Susan McCouch
**Cc:** Peter Wenzl; Ruth Bastow; Powell, Wayne (CGIAR Consortium); Daniele Manzella
**Subject:** Please review: Note to DivSeek Partners re January gathering in San Diego

Dear SC members,

After much deliberation about whether to convene a DivSeek Partner's Assembly (PA) in San Diego during January 2016, the JFU and I have reached the conclusion that it would be more constructive to invite PA members to participate in a workshop to discuss the goals and identity of the DivSeek initiative during this formative stage.

Please review the attached draft letter and get back to me (with a cc to the entire JFU) with your comments by Friday, Nov. 6.

We will do our best to accommodate your comments and suggestions, and plan to send a revised letter to the entire PA early on Saturday, Nov. 7.

We look forward to further discussions about DivSeek goals and identity, and plans for a PA later in the year during our SC meeting in Bonn on Dec. 8.

Thanks and best wishes,
Susan

--
Susan McCouch
Professor, Plant Breeding & Genetics
Cornell University
162 Emerson Hall
Ithaca, NY 14853-1901
Phone: +1 607-255-0420
Fax: +1 607-255-6683
Email: srm4@cornell.edu or mccouch@cornell.edu
Alternate Email: [redacted]

Dear Partners,

The DivSeek initiative will convene a roundtable discussion and workshop about genetic diversity data-sharing, and the initiative's governance on the **afternoon of January 8, 2016 in San Diego.** We invite you to attend. Please reserve that date. We will follow up shortly with details about the venue.

In the mean time, I provide some updates on DivSeek activities. The DivSeek Charter has now been endorsed by 59 organizations. Since the Partner's Assembly (PA) meeting in January 2015 in San Diego, eight Steering Committee (SC) members were elected through electronic voting and the first SC meeting was held in Rome in May 2015.

According to the Charter, the DivSeek initiative aims to (1) facilitate networking among partners interested in the application of state-of-the-art genotyping and phenotyping technologies to deepen our understanding of crop diversity, (2) promote the development of common standards to link germplasm with characterization data and to enable interoperability among information systems, (3) develop an annual work plan with a budget and a resource mobilization plan, and (4) organize capacity-building workshops.

In keeping with these stated goals and principles, DivSeek has undertaken a landscaping study to take stock of ongoing research that is relevant to the initiative, endorsed the use of permanent unique identifiers as standards for tracking germplasm, and convened two studies to help identify an appropriate governance structure that will enable the initiative to define and execute a work plan and seek appropriate levels of external funding.

**Landscaping Study.** A Landscaping Study has documented ~50 funded initiatives, relevant to the DivSeek goals and principles, currently underway around the world (see <divseek.org/landscape>). These projects focus on evaluating over a dozen crops at the genotypic and phenotypic levels, developing information systems, software tools and data standards. Hosting links to these projects could facilitate data exchange and help researchers to "network" and share information about "lessons learned."

**Permanent Unique Identifiers (PUID).** To help establish a set of core standards to facilitate data exchange and enable data integration and interoperability among diverse information systems, DivSeek endorses the use of PUID for crop germplasm; PUID will also enable phenotypic and genotypic datasets to be linked to specific sources of germplasm.

**Governance.** A team at Arizona State University (ASU) is exploring "lessons learned" about governance and information-sharing from various community-driven initiatives involving genetics research. An SC subcommittee on governance is exploring guidelines to improve the operational efficiency of the DivSeek initiative, so that it becomes more inclusive, responsive, and capable of launching externally-funded projects.

**Capacity Building.** Members of the SC have been exploring the possibility of DivSeek endorsing capacity-building workshops, including training in applying ontologies and standards for integrating phenotyping data at the lab and field scale.

We hope to see you at the Divseek roundtable discussion and workshop on Jan. 8 in San Diego and look forward to your input.

Regards and best wishes,
Susan McCouch

**Commented [BP1]:** Should initiative be capitalized throughout?

**Deleted:** his year, t

**Commented [BP2]:** By terming the meeting a discussion and workshop rather than a Partners Assembly, presumably we are indicating that the attendees will not decide, vote on or adopt any measures or actions? Otherwise we'd call it a "Partners Assembly?"

**Deleted:** host

**Deleted:** -

**Deleted:** about

**Deleted:** , and w

**Deleted:** join

**Deleted:** hold

**Deleted:** e

**Deleted:** , and w

**Deleted:** , and s

**Deleted:** permit

**Deleted:** ated

**Deleted:** agenda

**Deleted:** will

**Deleted:** help network researchers and

**Deleted:** ing of

**Deleted:** back

**Deleted:** that will enable it to be

**Deleted:** to

**Deleted:**

**Deleted:**

**Deleted:** that would be endorsed by DivSeek

**Deleted:** the use of

**Commented [BP3]:** With genotyping data?

**Deleted:** on of

## Bretting, Peter

**From:** Susan McCouch <srm4@cornell.edu>
**Sent:** Thursday, November 12, 2015 6:20 PM
**To:** Arnaud, Elizabeth (Bioversity-France); Graner, Andreas (GCP); David Marshall (JHI); Emily Marden (UBC); Bretting, Peter; Varshney, Rajeev (ICRISAT-IN); Hamilton, Ruaraidh Sackville (IRRI); Sarah Ayling (TGAC)
**Cc:** Peter Wenzl; Ruth Bastow; Powell, Wayne (CGIAR Consortium); Daniele Manzella; Susan McCouch
**Subject:** Re: Invitation to a DivSeek roundtable discussion

Dear Elizabeth et al,

Yes, it will be a lot to discuss in the time we have. We will discuss this more at our SC meeting on Dec. 8.

If we were to try to put together a survey, what questions would you want to pose such that we get useful feedback to help us focus on key points? Your input here would be most welcome.

A few suggestions that occur to me are:

1. Ask whether people feel that DivSeek should limit its activities to organizing meetings, helping develop standards for data-sharing and providing letters of support for initiatives that are generating genotypic and phenotypic information about crop germ plasm (such as those outlined in the Landscaping study) **OR** take a more active role and get involved in helping construct interoperable systems, I.e., by develping API's, or by extending existing data management and data analysis systems so they can be used by gene bank managers, pre-breeding programs, and others managing germplasm repositories or conservation efforts (I.e., in situ).
2. Ask our partners to each outline 1 or 2 workshops they would like to prioritize.
3. Should we convene a working group to develop a plan for bringing private sector colleagues into the DivSeek initiative? If so, what kinds of organizations or expertise should be represented in the working group?

Thanks
Susan

---

**From:** "Arnaud, Elizabeth (Bioversity-France)" <e.arnaud@cgiar.org>
**Date:** Thursday, November 12, 2015 at 4:18 AM
**To:** "Graner, Andreas (GCP)" <graner@ipk-gatersleben.de>, "David Marshall (JHI)" <David.Marshall@hutton.ac.uk>, "Arnaud, Elizabeth (Bioversity-France)" <e.arnaud@cgiar.org>, "Emily Marden (UBC)" [redacted], "Peter Bretting (USDA-ARS)" <Peter.Bretting@ars.usda.gov>, "Varshney, Rajeev (ICRISAT-IN)" <R.K.Varshney@cgiar.org>, "Hamilton, Ruaraidh Sackville (IRRI)" <r.hamilton@irri.org>, "Sarah Ayling (TGAC)" <sarah.ayling@tgac.ac.uk>, Susan McCouch <srm4@cornell.edu>
**Cc:** Peter Wenzl <peter.wenzl@croptrust.org>, Ruth Bastow <ruth.bastow@divseek.org>, "Powell, Wayne (CGIAR Consortium)" <w.powell@cgiar.org>, Daniele Manzella <daniele.manzella@divseek.org>
**Subject:** FW: Invitation to a DivSeek roundtable discussion

Dear Susan,

I see that you have extended the round table topic to much larger questions on DivSeek. I hope that we can cover all aspects in 5 hours. Wouldn't it be useful to have an online survey before to start collecting feedback and provide some key points for discussion ?

Kind regards

Elizabeth

---

**From:** Divseek Meetings <meetings@divseek.org>
**Date:** jeudi 12 novembre 2015 07:55
**To:** Divseek Meetings <meetings@divseek.org>
**Subject:** Invitation to a DivSeek roundtable discussion

Dear DivSeek partners and colleagues interested in DivSeek,

The DivSeek Initiative will convene a **roundtable discussion between 13:00 and 17:00 on January 8, 2016, in San Diego**. The focus will be to solicit input about DivSeek's role and identity in the larger landscape, its organizational structure, and its role in promoting access to and exchange of genetic diversity data.

We **invite 1-2 members from your organization** to attend and will follow up shortly with details about the venue.

In the meantime, we provide some updates on DivSeek activities. The DivSeek Charter has now been endorsed by 59 organizations. Since the Partners' Assembly (PA) meeting in January 2015 in San Diego, eight Steering Committee (SC) members were elected through electronic voting; the first SC meeting was held in Rome in May 2015 and the second is to be held in Bonn in December 2015. The 2016 Partners' Assembly will take place during spring-summer and will be announced as soon as that date is set.

According to the Charter, the DivSeek Initiative aims to (1) facilitate networking among partners interested in the application of state-of-the-art genotyping and phenotyping technologies to deepen our understanding of crop diversity, (2) promote the development of common standards to link germplasm with characterization data and to enable interoperability among information systems, (3) develop an annual work plan with a budget and a resource mobilization plan, and (4) organize capacity-building workshops.

In keeping with these stated goals and principles, DivSeek has been undertaking a landscaping study to take stock of ongoing research that is relevant to the Initiative, providing the basis for discussing DivSeek's identity in a larger context.

In addition, over the course of the year, an SC subcommittee was tasked to help identify an organizational (governance) structure that will enable DivSeek to define and execute a work plan and seek external funding. DivSeek has supported the use of permanent unique identifiers (PUIDs) for tracking germplasm and facilitating data integration. A study funded by JFU member organizations at Arizona State University (ASU) was asked to examine strategies for data sharing in other community-driven initiatives to provide a foundation for a discussion about DivSeek policies. Finally, several capacity building workshops are being organized to promote the adoption of standards and sharing of data and information.

The **topics for discussion** at the January roundtable include:

- **The identity of DivSeek**. The landscaping study documents more than 50 projects relevant to DivSeek goals and principles that are currently underway around the world (see http://www.divseek.org/landscape for a draft list). These projects either focus on evaluating over a dozen crops at the genotypic and phenotypic levels or the development of information systems, software tools and data standards. How should DivSeek interact with these projects to help researchers network, develop and adopt common standards, improve interoperability among information systems, and promote information sharing?

- **Data sharing and private-sector involvement**. The team at ASU will summarize its findings from a study on data-sharing in genetics research based on a review of eight community-driven initiatives. Lessons learned will be presented in the context of DivSeek to seek input from public and private sector colleagues.

- **DivSeek organizational structure**. A report from the SC will outline a refined organizational (governance) structure that should enable DivSeek to become more inclusive, responsive, and capable of launching externally-funded projects.

- **Permanent Unique Identifiers (PUID)**. DivSeek supports the use of PUID for crop germplasm as a way to track samples, link them to genotypic and phenotypic information, enable data integration, and promote interoperability among diverse information systems. We will update you on progress in this area.

- **Capacity Building**. Members of the SC are exploring the possibility that DivSeek co-organize capacity-building workshops, including training in ontologies and developing standards for integrating data across species, geographies and laboratories. We are seeking your guidance on areas to prioritize.

We hope to see you at the DivSeek roundtable discussion on Jan. 8 in San Diego, and we look forward to your input!

Regards and best wishes,

*Susan McCouch*

--

Chair, DivSeek Partners' Assembly

**Bretting, Peter**

| | |
|---|---|
| **From:** | Peter Wenzl <peter.wenzl@divseek.org> |
| **Sent:** | Tuesday, December 01, 2015 5:44 AM |
| **To:** | Andreas Graner; David Marshall; Elizabeth Arnaud (Bioversity-France); Emily Marden; Bretting, Peter; Rajeev Varshney (ICRISAT-IN); Ruaraidh Sackville Hamilton (IRRI); Sarah Ayling; Susan McCouch |
| **Cc:** | Daniele Manzella; Ruth Bastow; Wayne Powell (CGIAR Consortium); Dagny Poser |
| **Subject:** | DivSeek meeting in Bonn |
| **Attachments:** | DS_SC-2_15_1 (Meeting Agenda).pdf; DS_SC-2_15_4 (Report of Governance Expert Group).pdf; DS_SC-2_15_4 (Report of Governance Expert Group - Appendix 2).pdf; DS_SC-2_15_5 (Update on ASU study).pdf; DS_SC-2_15_6 (Tentative Agenda for DivSeek Meeting in January).pdf; DS_SC-2_15_7 (Proposal for a Bellagio Meeting).pdf; DS_SC-2_15_8 (Membership Applications).pdf; Hotel and transfer.docx |

Dear DivSeek Steering Committee members,

Please find attached the Agenda and other Documents for the **DivSeek Steering Committee meeting on December 8th** next week. The meeting will start at **9:00am** at the premises of the Crop Trust (Platz der Vereinten Nationen 7, 53113 Bonn).

In a follow-up message later this week, we will send you two additional Documents (No. 2, 3) with notes and updates from Susan and the JFU organizations.

As a reminder, here's a quick overview of the logistics:

- **Accommodation:** Accommodation has been booked at the Hotel Stern, Markt 8, 53111 Bonn, Germany. Accommodation and breakfast have been pre-paid by the Crop Trust. All extras are to be paid by the guest upon check out.

- **Transfer:** please see attached sheet for directions from the airport to the hotel, and from the hotel to the Crop Trust offices.

- **Per diem:** we'll pay you a per diem to cover costs for meals not provided by us. It should also cover incidentals such as transport to/from the airport.

- **Dinner on 7 Dec:** you're invited for a joint dinner at 19:00 on the evening before the meeting at the Ruland Restaurant, Bischhofsplatz 1, Bonn. This is a 2-min walk from the hotel.

Please do not hesitate to contact us in case you've any questions in regard to the agenda, documents or the logistics of the meeting.

Looking forward to meeting you next week!

On behalf of the Joint Facilitation Unit,

Peter

--

Peter Wenzl

DivSeek Liaison
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 126
Mobile: [redacted]
www.croptrust.org

*Securing our Food, Forever*

DS/SC-2/15/1

# Second meeting of the Steering Committee

# 8 December 2015

**GCDT Headquarters**

**Bonn, Germany**

## Draft Provisional Agenda

1. Welcome and opening of the meeting
2. Approval of the agenda
3. Updates since the last Steering Committee meeting in May 2015
4. Identity and goals of DivSeek
5. DivSeek governance
6. Preparation of DivSeek workshop in January 2016
7. Plans for a Bellagio meeting to discuss the science/policy interface in relation to crop germplasm
8. Review of new membership applications
9. Other business
10. Preparation of the report

## Indicative Timetable

| Time | Agenda item | Title | Working documents |
|---|---|---|---|
| 09:00-09:15 | 1 | Welcome | |
| 09:15-9:30 | 2 | Approval of the agenda | DS/SC-2/15/1 |
| 9:30-10:30 | 3 | Updates since the last Steering Committee meeting in May 2015 | DS/SC-2/15/2 |
| 10:30-11:00 | 4 | Identity and goals of DivSeek | DS/SC-2/15/3 |
| 11:00-11:30 | *Coffee break* | | |
| 11:30-12:30 | 4 | Identity and goals of DivSeek (*cont.*) | |
| 12:30-13:00 | 5 | DivSeek governance | DS/SC-2/15/4<br>DS/SC-2/15/5 |
| 13:00-14:00 | *Lunch* | | |
| 14:00-15:00 | 5 | DivSeek governance (*cont.*) | |
| 15:00-16:00 | 6 | Preparation of DivSeek workshop in January 2016 | DS/SC-2/15/6 |
| 16:00-16:30 | *Coffee break* | | |
| 16:30-17:00 | 7 | Plans for a Bellagio meeting to discuss the science/policy interface in relation to crop germplasm | DS/SC-2/15/7 |
| 17:00-17:15 | 8 | Review of new membership applications | DS/SC-2/15/8 |
| 17:15-17:30 | 9 | Other business | |
| 17:30-17:45 | 10 | Preparation of the report | |

# DIVSEEK Roundtable Discussion

**8 January 2016**
**13:00 – 17:00**
**Venue to be confirmed**

## Draft Agenda

Welcome and introductions

**What is the goal of DivSeek?**

- DivSeek landscape
  - Discussion
- The role and identity of DivSeek
  - Discussion

**Governance**

- Future model for DivSeek governance
  - Discussion
- Institutional and organizational factors (ASU governance study)
  - Discussion

**Start-up Activities**

- Permanent Unique Identifiers for germplasm
  - Discussion
- Proposed DivSeek workshops
  - Discussion

DS/SC-2/15/8

# Applications for membership

This document contains the two (2) letters of interest that the Joint Facilitation Unit has received in the period in between the first and second meetings of the Steering Committee, namely from:

i) The University of Bologna (Italy), Department of Agricultural Sciences, Prof. Roberto Tuberosa – see Appendix 1;

ii) the Center for Desert Agriculture (Saudi Arabia), KAUST, Prof. Mark Tester – see Appendix 2.[1]

The two organizations have applied by filling out the form that is available on-line at http://www.divseek.org/apply-to-join-divseek/.

The Committee is invited to review the two applications with the respect to the relevance of the applicant organization to the mission and principles of DivSeek.

[1] The signature on this application is missing; the JFU will present a signed copy of this document at the Steering Committee meeting.


# LETTER OF EXPRESSION OF INTEREST

To: the Joint Facilitation Unit of DivSeek

## Purpose of this letter

This letter is to express interest in joining DivSeek as a Partner organization.

DivSeek is a community-driven initiative that aims to cross-link, support and add value to individual projects that deepen our understanding of crop diversity and stimulate efforts to mobilize natural genetic variation to accelerate crop improvement and enhance food and nutritional security.

The completed form (provided below) will be forwarded to the Steering Committee of DivSeek, who will review the request to join DivSeek with respect to the relevance of the organization to the mission and principles of DivSeek, as expressed in the DivSeek Charter. The DivSeek Charter is available at: http://www.divseek.org/documents

Upon review of the information provided in the completed form, the Steering Committee may request more information.

The Steering Committee of DivSeek meets twice per calendar year and will review expressions of interest during those meetings. The decision of the Steering Committee will be communicated to the requesting organization by email.

If the Steering Committee confirms acceptance of the request to join, the requesting organization will formally be invited to join DivSeek via acceptance of the DivSeek Charter in writing by a representative of the requesting organization.

The DivSeek Charter defines the general conditions for the operation of DivSeek and sets forth the governance structure for voluntary cooperation by Partner organizations. The Charter does not create any legally binding obligation between or among Partner organizations.

Partner organizations support DivSeek by voluntarily associating specific activities with DivSeek and by providing advice and support. Partner organizations individually determine the nature and extent of their participation in DivSeek.

*Please fill the sections below with information on your organization and return an electronic copy of this letter, and attachments thereto, to:* **membership@divseek.org**

## Contact details

***Name of the organization:***

Department of Agricultural Sciences, University of Bologna

***Institutional website:***

www.scienzeagrarie.unibo.it/en and also www.distagenomics.unibo.it

***Country:***

Italy

***Address:***

Department of Agricultural Sciences
Viale Fanin 44
40127 - Bologna
Italy

***Name and contact details of the person filling this letter on behalf of the organization:***

Prof. Roberto Tuberosa

## Mission and activities

***Please describe the mission and activities of the organization, as they relate to the mission and principles of DivSeek. Please include links to information available on the web and, if necessary, attach files.***

The Department of Agricultural Sciences (DipSA) has sound expertise across a diverse range of research activities related to the management of the agri-environment. The Department seeks to benefit society by advancing knowledge and creating new agri-environmental technologies that foster food security and sustainability of agricultural practices. We share this knowledge and seek its transfer to agriculture through publication, teaching and collaboration. DipSA offers a diverse range of scientific disciplines that conduct their research activities in the agri-environment. Therefore, DipSA fully covers the area of interest of DivSeek.
For more details, please see http://www.scienzeagrarie.unibo.it/en

## Anticipated contributions

***Please describe the specific activities and/or projects that the organization would like to associate with DivSeek.***

The research activities of the group led by Prof. Roberto Tuberosa deal with the application of genomics approches toward the improvement of durum wheat, barley and maize. Thede activities are often carried out in collaboration with seed companies from Italy and worldwide. A fundamental aspect of these activities relates to the assembly and characterization of genetic materials (e.g. mapping populations, panels of accessions suitable for association mapping, introgression library lines, near-isogenic lines, mutants, etc.) that allow us to dissect the genetic make-up of important agronomic traits while fostering a better understanding of the functional networs underlining the phenotype. The ultimte goal is to implement marker-assisted selection and to clone QTLs or mendelian loci in order to gain a better

## Other information

***Please insert any other information to support your expression of interest.***

We are very interested in DivSeek activities related to genetic resources of durum wheat, barley and maize, i.e. the three crops that we have been working on starting from 1970. We have valuable genetic materials that we will be glad to share with the cereal community and are keen to access additional biodiversity that would allow us to sample a broader range of diversity as compared to that we usually deal with.

Bologna 2/9/2015

Date

Signature

Prof. Roberto Tuberosa

Mobile: [redacted]
E-mail: roberto.tuberosa@unibo.it

# LETTER OF EXPRESSION OF INTEREST

To: the Joint Facilitation Unit of DivSeek

## Purpose of this letter

This letter is to express interest in joining DivSeek as a Partner organization.

DivSeek is a community-driven initiative that aims to cross-link, support and add value to individual projects that deepen our understanding of crop diversity and stimulate efforts to mobilize natural genetic variation to accelerate crop improvement and enhance food and nutritional security.

The completed form (provided below) will be forwarded to the Steering Committee of DivSeek, who will review the request to join DivSeek with respect to the relevance of the organization to the mission and principles of DivSeek, as expressed in the DivSeek Charter. The DivSeek Charter is available at: http://www.divseek.org/documents

Upon review of the information provided in the completed form, the Steering Committee may request more information.

The Steering Committee of DivSeek meets twice per calendar year and will review expressions of interest during those meetings. The decision of the Steering Committee will be communicated to the requesting organization by email.

If the Steering Committee confirms acceptance of the request to join, the requesting organization will formally be invited to join DivSeek via acceptance of the DivSeek Charter in writing by a representative of the requesting organization.

The DivSeek Charter defines the general conditions for the operation of DivSeek and sets forth the governance structure for voluntary cooperation by Partner organizations. The Charter does not create any legally binding obligation between or among Partner organizations.

Partner organizations support DivSeek by voluntarily associating specific activities with DivSeek and by providing advice and support. Partner organizations individually determine the nature and extent of their participation in DivSeek.

*Please fill the sections below with information on your organization and return an electronic copy of this letter, and attachments thereto, to:* **membership@divseek.org**

## Contact details

***Name of the organization:***

Department of Agricultural Sciences, University of Bologna

***Institutional website:***

www.scienzeagrarie.unibo.it/en and also www.distagenomics.unibo.it

***Country:***

Italy

***Address:***

Department of Agricultural Sciences
Viale Fanin 44
40127 - Bologna
Italy

***Name and contact details of the person filling this letter on behalf of the organization:***

Prof. Roberto Tuberosa

## Mission and activities

***Please describe the mission and activities of the organization, as they relate to the mission and principles of DivSeek. Please include links to information available on the web and, if necessary, attach files.***

The Department of Agricultural Sciences (DipSA) has sound expertise across a diverse range of research activities related to the management of the agri-environment. The Department seeks to benefit society by advancing knowledge and creating new agri-environmental technologies that foster food security and sustainability of agricultural practices. We share this knowledge and seek its transfer to agriculture through publication, teaching and collaboration. DipSA offers a diverse range of scientific disciplines that conduct their research activities in the agri-environment. Therefore, DipSA fully covers the area of interest of DivSeek.
For more details, please see http://www.scienzeagrarie.unibo.it/en

## Anticipated contributions

***Please describe the specific activities and/or projects that the organization would like to associate with DivSeek.***

The research activities of the group led by Prof. Roberto Tuberosa deal with the application of genomics approches toward the improvement of durum wheat, barley and maize. Thede activities are often carried out in collaboration with seed companies from Italy and worldwide. A fundamental aspect of these activities relates to the assembly and characterization of genetic materials (e.g. mapping populations, panels of accessions suitable for association mapping, introgression library lines, near-isogenic lines, mutants, etc.) that allow us to dissect the genetic make-up of important agronomic traits while fostering a better understanding of the functional networs underlining the phenotype. The ultimte goal is to implement marker-assisted selection and to clone QTLs or mendelian loci in order to gain a better

## Other information

***Please insert any other information to support your expression of interest.***

We are very interested in DivSeek activities related to genetic resources of durum wheat, barley and maize, i.e. the three crops that we have been working on starting from 1970. We have valuable genetic materials that we will be glad to share with the cereal community and are keen to access additional biodiversity that would allow us to sample a broader range of diversity as compared to that we usually deal with.

Bologna 2/9/2015

Date

Signature


DS/SC-2/15/4

# Report from the governance expert group

This document contains the report of the governance expert group that Prof. Emily Marden convened and chaired at the request of the other members of the Steering Committee. At its first meeting, the group was tasked with elaborating operational guidelines on multiple topics, including membership in the initiative and the role of the Joint Facilitation Unit.

The Committee is invited to consider the report and advise on the next steps.

To: DivSeek Steering Committee

From: DivSeek Governance Expert Group

Re: Report of Governance Expert Group

Date: December 8, 2015

---

## SUMMARY

DivSeek is strongly advised to:

1. Modify the current organizational structure to include an Executive Director with executive/operational function. In this revised model, the JFU members would become advisory, or could be seconded for specific functions under the direction of the Executive Director. The Steering Committee would maintain the same role and would be important for setting the ground rules and objectives. The Assembly's role would continue unchanged.

2. Develop (a) a five-year strategic plan that sets out operational policies to define the range of projects and partnerships to pursue and the key goals and objectives, and (b) an annual workplan to realize the goals of DivSeek.

3. Empower the Executive Director to execute the workplan and five year strategic plan.

If this recommendation is accepted, there are three options for implementation:

1. The Steering Committee recommends the appointment of an Executive Director to the Assembly. In its recommendation, the Steering Committee identifies one of the current JFU partner organizations as the host institution for the executive function (i.e. administrative capacity). The Assembly votes on the recommendation;

2. The Steering Committee recommends the appointment of an Executive Director to the Assembly. In its recommendation, the Steering Committee designates the executive function to be located within an existing organization that is engaged in similar ventures as DivSeek to deliver the programming (e.g. CIAT). The Assembly votes on the recommendation; or

3. The Steering Committee recommends to the Assembly that DivSeek contract with a third party organization with recognized executive and management capacity that can deliver the programming under contract (e.g. CABI). The Assembly votes on the recommendation.

## DISCUSSION

### 1. Background

At the first meeting of the DivSeek Steering Committee in May, 2015, certain issues were referred for further consideration to a "governance expert group" to be convened by Steering Committee member, Emily Marden. The list of issues identified by the Steering Committee for the governance expert group were identified in the Report of the Steering Committee Meeting; the relevant excerpts from the Report are included in Appendix 1 to this document.

Pursuant to this request, Emily Marden convened an expert group consisting of (in addition to herself): Bill Boland (U. Saskatchewan), Peter Bretting (USDA), Regiane Garcia (U. British Columbia), and Peter Phillips (U. Saskatchewan). Collectively, the group has extensive experience with public and private agriculture governance issues, including experience with organizations around the world. The expert group held meetings by teleconference in September and October, as well as discussions via email.

In addition to bringing their relevant expertise to bear on the questions presented, the expert group considered the following DivSeek documents: (1) the May, 2015 Steering Committee report; (2) the DivSeek Charter, adopted in January, 2015; and (3) the Operation of the Joint Facilitation Unit (2015) DS/SC---1/15/4 document.

The expert group deemed the governance questions to have priority and so focused mainly on these. Recommendations on publication and the private sector were also considered and are summarized at the end of this report.

### 2. Governance Issues

DivSeek was formed at the first Assembly of the Partners in January 2015 in San Diego. As a part of that formation, a Charter was approved. The DivSeek Charter identified roles for a Joint Facilitation Unit (JFU), a Steering Committee (SC) and the Assembly of Partner Organizations (Assembly). The DivSeek JFU currently consists of a single representative from each of: the Secretariat of the International Treaty (Treaty), the Global Crop Diversity Trust (GCDT), the Consortium Office of CGIAR, and the Global Plant Council (GPC).

As is often the case in the first year of a new organization, governance challenges have arisen. To some extent, DivSeek was conceived with both too much and too little governance: that is, DivSeek has a JFU, SC and Assembly, but lacks both clear operational leadership and a team to deliver the work of the organization.

#### a. Importance of Operational Leadership and Function

The JFU developed an Operational Document (DS/SC-1/15/4)(OD) to help guide DivSeek and to clarify roles. However, while this OD lays out what could be an appropriately

aggressive initial mandate and set of activities, it does not provide for an operational structure to advance the work.

The OD suggests that all activities of DivSeek are decided, supported and implemented collectively (§2.2). At the same time, the JFU members ultimately answer to their organizations rather than DivSeek. Such an approach is certainly appropriate for the development and founding of an organization. However, this structure is problematic for operation, especially for an organization such as DivSeek that aims to engage flexibly with a variety of actors, including international organizations, NGOs, universities, NARS, farmer organizations, producer organizations and the private sector. Consensus at the operational level, while representing individual organizations, is not feasible particularly in the face of pressures to be adaptive and responsive.

The attempt to embrace consensus, even while representing divergent views, is a common problem within agricultural research partnerships. In general, key contributing organizations want to position their own personnel within the decision-making process to observe developments and to protect their investment and interests. Moreover, agricultural research partnerships can be difficult to organize efficiently as they often consist of a variety of dissimilar organizations with different values and organizational objectives. Extensive research on agricultural-related partnerships by Phillips, Boland and Ryan (2013) (attached at Appendix 2) suggests that, outside of funding issues, a lack of feasible operational principles is the greatest threat to the survival of these partnerships. We mention a few cases from this research below to demonstrate the significant impact of operational principles on outcomes:

- Vineland Research and Innovation Center: One model of success is the Vineland Research and Innovation Center in Ontario, Canada. This is a large and complex partnership that evolved from a former public research institute into a partnership that consists of over 30 upstream and downstream organizations. Vineland is governed by a board of 12 directors, and one CEO, who has full control of operations and finance. The office of the CEO retains operational control and streamlines decision making into a single authoritative system. The Board approves annual work plans and all key operational policies. The Board's input is relied on as important – given that members are experts drawn from many of Vineland's partners. As such, while the CEO maintains operational authority, the Board provides input and review, and critically, links together a large number of diverse organizations into network and sets the tone for shared interests and investments.

- Molecular Plant Breeding Cooperative Research Centre (MPBCRC): In contrast, the MPBCRC in Australia failed despite having a sound business plan and being well-financed. At one time it was one of the largest agribiotechnology ventures in the Southern Hemisphere. MPBCRC used a distributed model of governance and lacked a central decision making capabilities, relying instead on consensus based decision making. MPBCRC is no longer operational. MPBCRC suffered from a lack of a clear board vision (they agreed on the general direction but could not distill it to

instructions to their operational team) and ineffective leadership. This was compounded by conflicts emanating from the different sectors, as public and private employees use different values that were operationally incompatible, providing grounds for conflict. Failure could not be attributed to their output: the ROI was 300% on technology investments and 700% on educational outreach. The business fundamentals were sound, but the governance structure was not capable of sustaining the partnership.

An array of other examples along the spectrum of success to failure is included in the attached report.

**b. Elements of Successful Ag Organizations**

The Phillips, Boland and Ryan (2013) study identifies a set of considerations necessary to a demand-driven research partnership capable of operational success. These are worth considering *in toto* as DivSeek moves forward. We acknowledge that some of the issues below have already been addressed by DivSeek.

The factors are:

1. Initial Identification of the Common Interest Driving the Organization

   - Formation of a Committee to oversee the planning of the partnership;
   - Mapping the research network, identifying and convening potential partners and key actors in the research network;
   - Determining the common interest shared by the potential partners.
   - Developing a clear and concise strategic vision to guide the participants and to empower an operational mandate; and
   - Defining the time commitment, nature of the work and extent of in kind services to the partnership so that the results and operations of the partnership do not elicit conflict with the individual partners.

2. Core Elements:

   - Organization: includes a description of the roles and responsibilities of each partner organization, the governing body the board (Steering Committee) and the executive (Executive Director);
   - Activities: includes a description of each partner's activities and responsibilities as well as the mechanisms of interaction among partners;
   - Budget: includes the total cost of partnership, joint financing requirements, and the specification of each partner's contributions—in cash and in-kind—or at least principles and practices that will enable future contributions; and

- Monitoring and evaluation mechanisms: include an examination not only of the results of the partnership, but also of the collaboration itself, including an analysis of the partners' commitments and the overall synergistic effects.

3. Common Clauses in a Organization Formation Agreement

- Identification of the partners
- Subject of the contract: the partnership
- Objectives of the partnership
- Organizational design
- Duration and termination
- Obligations and commitments of the partners
- Means of contributing resources (financial and in kind)
- Dates of payment
- Types of activities
- Evaluation and monitoring mechanisms
- Mechanisms for conflict resolution

### c. Recommendations for DivSeek

Further defining the operating principles is a necessary next step. Specifically, there is a need to define both: (1) the existence of and parameters for executive action and (2) the nature of partnerships to be encouraged through DivSeek.

#### i. Establish Operational Leadership

Based on the Phillips, Boland and Ryan study, it is clear that an empowered executive is necessary to allow DivSeek to engage and leverage opportunities.

There are two possible paths, any of which could be implemented via one of the three specific approaches discussed below:

- "Top Down": An executive could be established and given a set of guiding principles that set the outer bounds of the allowable partnerships and activities; or
- "Bottom Up": An executive could be established and given the authority to engage with any and all current and future partners on projects to advance

their DivSeek related activities, thereby building through custom and precedent the range and scope of allowable partnerships.

Neither is unambiguously preferred: top down definition of principles in absence of any practical examples can be slow and/or self-limiting while the bottom-up approach is highly enabling but can lead to an excessive diffusion of models and simply put the onus on the putative partners to define their principles. Over time, each model is likely to converge on a common set of principles.

Depending on the institutional approach chosen, an effective administration is needed to implement the strategies and plans. This would necessarily include development of a budget for the executive function and recruitment of necessary staff either by temporary/permanent staffing and/or secondments is necessary.

### ii. Further Define Operating Principles

In either case, an empowered executive needs clear parameters for operating (i.e. what can be done by the executive, and where must additional consultation with the Steering Committee take place?). There are models for such principles that can be provided; these could be reviewed and modified by the Steering Committee to delegate appropriate amounts of authority. In this context, the Steering Committee must decide what types of actions the executive is authorized to take without SC review, and which activities require notification to the SC, or authorization by the SC.

An executive for a multi-faceted organization, such as DivSeek, must also be able to draw upon and engage experts in relevant fields. Such expertise can be gained by seconding members of the JFU or other partner organizations, as appropriate.

Importantly, the move to adopt an executive function requires only moderate revision of the Charter. It can be revised to incorporate an executive function to undertake the operations of DivSeek. The JFU can remain in an advisory capacity. The SC and Assembly remain largely unchanged.

From a governance perspective, DivSeek currently lacks clear workable operating principles. The OD §3.5 addresses management of the Steering Committee and Assembly. However, while these work items are necessary, they are not sufficient to ensure that DivSeek itself operates. There is a need to provide a sharper focus on piloting or advancing the practical data sharing platforms envisaged in DivSeek.

Given the many distinct stakeholders and broad goals of DivSeek, it is not reasonable to expect the organization itself to have immediate or near-term access to adequate internal resources to deliver the new platforms by its own initiative. Instead, DivSeek will need to draw on experts and to work cohesively and effectively with other organizations. In this context, all of the early priority opportunities discussed at the past meetings are in areas where there are established actors, a few investments and some action consistent with DivSeek. For this reason, it will be necessary to develop an operational model that works with, rather than competes with, these other actors and ventures.

### iii. Medium Term Plan

While there is pressure for immediate action, a 5 year plan (say 1-3 pages) should be developed which lays out medium term expectations and goals. Most of the projects are unlikely to fit in one-year increments, so having a longer term vision and set of goals would help to guide the development and implementation of those projects.

### iv. Annual Work Plan

At present, DivSeek is long on principles and short on actionable activities. An annual workplan with priorities for the next calendar year must be developed for immediate action. The workplan should identify a range of specific activities that assign responsibility and motivate action.

## 3. Specific Options for Establishing an Executive Function

**Option 1: Build an executive function at one of the existing JFU partners**

- **Synopsis**: The JFU could be restructured into an operating unit rather than a secretariat. This would require the core partners to the JFU to decide on how they will transfer control of their staff and assets to one entity and then step back. This could be done quickly and cleanly if there is agreement among the JFU members.

- **Transition considerations**: Establishing an executive function within one of the existing JFU organizations would be the simplest to effect if the four JFU founding members support this approach. The advantage is that all four partners to the JFU have been involved from the start and have a good sense of the opportunities and implications. However, this approach may not be a simple matter to execute.

**Option 2: Partner with another organization, which is engaged in ventures similar to DivSeek, to provide the executive function**

- **Synopsis**: A number of organizations around the world are engaged directly in activities consistent with DivSeek. It could be possible, given the right circumstances, to negotiate a partnership whereby the responsibility for DivSeek's executive function is transferred to a third party. This could involve full devolution of the venture or the transfer of the venture as a new 'business-line' for the organization. One option floated as a for-instance was CIAT, which has recently received new funding for a DivSeek-like venture. Combining resources could motivate the DivSeek venture and accelerate new projects. Depending on the terms of the transfer, there would be different impacts on the JFU, the founding partners and the Steering Committee. The main challenge of this option is that any organization taking this on would likely want to ensure the venture adds value to their mission—if

their mission changes, it could pull DivSeek in directions other than intended by the Charter.

- **Transition considerations**:
    - This option would likely require (a) the partners of the JFU to agree to transfer authority and likely some funding/staff to support such a venture, (b) flexibility on the part of the recipient organization to accommodate the goals of DivSeek and the role of the Assembly (and possibly even the Steering Committee).
    - If this option is considered, the Steering Committee could either issue a call for expressions of interest, proactively identify and approach obvious partners to explore this option, or do both.

**Option 3: Contracting with a third party organization with recognized executive and management capacity that can deliver the programming under contract**

- **Synopsis**: One strategy would be to essentially contract out the management function, either to an international or not-for-profit organization (or even to a for-profit management firm). This would create the cleanest break between the executive function and board oversight, as the relationship would be governed by a contract, which would help to focus the efforts of the charter signatories and the other partners to identifying strategic direction. Sometimes the intervention of an arms-length disinterested manager can help the partners and projects be developed efficiently and effectively.
- Transition considerations:
    - This option would require the partners of the JFU to transfer funding to support such a venture.
    - This option would allow the Partners' Assembly (and possibly even the Steering Committee) to continue to function as envisaged in the Charter.
    - If this option is considered, the Steering Committee must issue a call for expressions of interest. There might be a few obvious partners to proactively approach and invite to bid on the contact but it would be ill advised to sole-source this contract.

## 4. Publication of DivSeek Meeting Documents

For an organization that places a priority on transparency, the common practice is to document meetings by reporting topics discussed, but omitting identification of individuals' positions or disagreements. Thus, a meeting report can identify the agenda, including issues

discussed, and report the discussion ensued. Where necessary different perspectives can be reported with the ultimate decision reached.

This kind of approach serves the purpose of transparency and communication while still ensuring space for free and open discussion.

## 5. Engagement with the Private Sector

The governance expert group feels strongly that open discussion with the private sector is important as a first step for gauging the degree to which the private sector is interested in participating in Divseek, and the terms they seek. The expert group received one unsolicited statement from Syngenta expressing interest and desired terms. However, the group also awaits the results of the ASU study which looked specifically at the terms and successes of private sector engagement in a number of analogous organizations.

# Appendix 1

**Summary of Requests to Expert Committee**

**(arising from May 28, 2015 Steering Committee Meeting)**

1. Governance Issues

*22. The Committee decided to request one of its members, namely Ms. Emily Marden, to convene, under her chairmanship, a governance expert group, in accordance with the Charter's provision to elaborate operational guidelines through expert consultations, in order to:*

*i) validate the Committee's provisional opinion about membership at the level of organizations/institutions, and/or clarify alternative options and implications;*

*ii) advise the Committee on possible steps towards private sector membership or other engagement, including an assessment of the implications on the implementation of DivSeek's principles as stated in the Charter.*

*23. In conjunction with the decision to convene a governance expert group, the Committee was informed about an on-going research project by Arizona State University (ASU) on institutional and organizational factors for enabling data access, exchange and use, which the Global Crop Diversity Trust and the Secretariat of the International Treaty were co-funding. Mr. Manzella, of the Joint Facilitation Unit and the International Treaty, informed the Committee of the preliminary research activities conducted by the ASU research team for the project, and distributed a progress report. The Committee invited Ms. Marden to coordinate with the ASU research team to obtain early access to the results of the study for consideration as part of the work of the governance expert group.*

*32. [The Steering Committee] considered a number of potential issues in relation to the role of the Joint Facilitation Unit within DivSeek, as follows:*

*i) modalities for expansion or contraction of the Joint Facilitation Unit, e.g. in cases where one organization is inactive or becomes unable to serve, or where a Partner organization expresses interest in joining the Unit;*

*ii) the roles and responsibilities of individual representatives of the organizations that serve the Unit;*

*iii) the modalities of representation by the respective organizations within the Unit;*

*iv) the modalities for decision-making within the Unit;*

*v) the relationship between the Unit and the other elements of DivSeek's governance structure (i.e. the Assembly and its Chairperson and the Steering Committee) with respect to communication lines and providing guidance and direction.*

*33. The Committee requested the governance expert group to be convened by Ms. Emily Marden to prepare a document for the consideration of the Committee, based on the provisions of the DivSeek Charter, to explain the governance structure*

*of DivSeek, to describe mechanisms that would allow it to evolve in the future, and to present options for clarifying the above issues.*

*37. The Committee requested the governance expert group to be convened by Ms. Emily Marden to elaborate a policy on the publication of DivSeek meeting documents and reports, for the consideration of the Committee. Pending the development of such a policy, the Committee decided not to publish this report online.*

2. Membership Issues

   a. Organizational Level

   *18. Regarding a) and b), the Committee agreed to provisionally keep the current membership at the level of organizations/institutions, as this aligned with the current governance settings of the Charter. It considered membership tiers as a possible future solution to reflect different interest groups (e.g. donors, communities of practice, advisors and service providers).*

   b. Private Sector

   *21. Regarding e), the Committee was alerted by the Joint Facilitation Unit to the opportunity to keep an active line of communication with the private sector representatives who were at the first Partner Assembly. The Committee highlighted the potential of private sector engagement for DivSeek funding of future training and capacity building programs, as well as for expanding the range of expertise and knowledge within DivSeek. It also discussed some of the systemic and practical implications of private sector membership, with particular attention to a balanced relationship among different DivSeek constituencies and the need to promote equitable data sharing policies. It also recalled the annotation in the Charter, which referred to observer status for private sector, pending the development of operational guidelines for private sector engagement.*

3. Publication Issue

   *37. The Committee requested the governance expert group to be convened by Ms. Emily Marden to elaborate a policy on the publication of DivSeek meeting documents and reports, for the consideration of the Committee. Pending the development of such a policy, the Committee decided not to publish this report online.*

DS/SC-2/15/5

# Report from the governance research project

This document contains a memorandum developed under the governance research project that Arizona State University (ASU), in cooperation with CIRAD, is conducting to present DivSeek with options for governance, data sharing and membership.

The Committee received a first progress report by the research team at its first meeting.

The Committee is invited to appraise the memorandum and advise on the next steps, including for the preparation of the DivSeek workshop/roundtable in January 2016.

**DivSeek Governance Research Project**
**Memorandum for the Steering Committee**
*Eric Welch, Selim Louafi, Federica Fusi and Daniele Manzella*
**8 December 2015**

## Purpose of the memorandum

Following the progress report to the first meeting of the Steering Committee in May 2015, this memorandum is to update the Committee at its second meeting, on progress made with the governance research project. The memorandum also contains preliminary observations derived from the research, on themes that are on the agenda of the Steering Committee.

This *interim* memorandum does not explicitly address the broader policy context in which the initiative takes place. Constraints and sensitivities deriving from such a context are likely to impact goal setting, data sharing, resources and membership, and the DivSeek community is encouraged to consider them in conjunction with the findings of the research project, for instance when addressing developing country involvement and capacity development. This memorandum focuses on general governance aspects that are applicable across the case studies analyzed in the research project, independently from their policy context.

## Methodology and case study selection

As outlined briefly in the first progress report of May 2015, the research team has organized the work in three main phases:

(1) Literature review and sample frame development. In line with the objectives of the study, nineteen cases in the area of food and agriculture and seven cases in the human health sector were identified for possible inclusion in the study. Appendix 1 contains the criteria used for evaluating those cases.

(2) Initial selection of ten cases most relevant for DivSeek. Initial selection was followed by 45-60 minute interviews of the ten Project Managers and Executive Directors to identify six cases for in-depth data collection and analysis. Appendix 2 contains the initial list of the ten cases and the final group of six cases.

(3) All six Project Managers and Executive Directors were sent invitations asking them to participate in the in-depth analysis of their organizations. The research team has conducted 45-75 minute interviews with staff, partners and clients of the six organizations, but has not completed all interviews.

This memorandum provides a second report on initial findings to date. The full report, to be delivered by the end of January 2016, will elaborate on the findings presented here.

## Preliminary Findings

Since data collection and analysis under the project are still taking place, the following observations should be considered preliminary. Nevertheless, the research team hopes that the findings and observations will be of use to the DivSeek Committee.

*Goal setting*
Based on the case analyses to date, it is clear that governance structures have strong accordance with program goals. Case study interviewees consistently explain that clear

specification of program goals is an essential first step for program success. All interviewees, whether leadership or staff, were able to articulate program goals and explain how their goals differ from others or fill a particular research or service gap. Cases show that program goals are the substance around which governance structures are built as they facilitate accomplishment of the goals.

Program goals vary across the cases. In some cases, program goals are mainly technical. In others, goals are broader and seek to integrate the values and perspectives of heterogeneous groups through some set of agreed-upon rules for access and behavior of members. Cases demonstrate that specification of the goals is often a slow process requiring two or more years. Some of the interviewees have noted that their projects or programs began small, with only a few actors who were able to achieve consensus to set program goals. Consensus on goals is much more difficult when a large more heterogeneous group of actors is involved. However, homogeneity and small initial size can also reduce inclusivity, ignore obvious complexities and work against long-term buy-in.

These observations are significant for DivSeek. For example, alternative strategies that the DivSeek community might pursue, could 1) aim to establish technical standards critical for advancement of cross-crop genomics research; or 2) aim to establish a heterogeneous community – industry, OECD country universities and national agriculture research organizations in developing countries – around technical standards, capacity development and sharing norms. Governance structures for the former will certainly be less complex than the latter. Both strategies require a reflection about which actors to include in the goal setting process and which actors might be invited later.

*Defined resources*

Case analyses to date show that governance structure also depends on the resource(s) that the program/project seeks to develop and make available to the community. The choice of resources is usually dependent on the identification of a demand that arises from the field. All interviewees to date were able to identify specific human, software, data, standardization, negotiation/brokering, material, computational or storage resources that are not offered (or are of low quality, or are poorly distributed) by other programs. They talked about how their organizations developed strategies to fill the identified gap. DivSeek should consider specifying the resources that 1) are not provided elsewhere; 2) are in high demand; and 3) may stimulate convergence across different sub-communities. Subsequently, governance structure should establish who has access to the common resources and under what conditions.

*Sharing policies*

There appear to be two main approaches to data sharing: centrally determined or locally autonomous. Centrally determined rules make products/data/findings 1) completely open such that all products/data/findings are publicly available; 2) partially open such that there is an embargo on products/data/findings of one to two years before making products publicly available; or 3) closed such that only members share among themselves (This last one is a theoretical category. None of the cases are completely closed). Organizations with centralized sharing policies can design monitoring and controlling mechanisms to prevent opportunistic behaviors, such as unbalanced contributions of resources to the common pool. Thus far, cases show that imposing centralized rules for resource sharing does not always encourage sharing behavior; actors may withdraw from organizations where there are sharing requirements.

Locally autonomous systems allow actors to decide whether or not to share products/findings/data publicly. Locally autonomous systems often include flexibility that leaves actors free to decide with whom and under what conditions they will share. Some of these projects create and manage formal and informal services such as match-making, brokerage activities, communication or technical support that build trust among members, with the ultimate aim of encouraging sharing.

*Membership, including by private sector*

The interplay between resources and sharing policies influences membership. Different types of resources and governance structures provide alternate incentives for participation. The influence of such an interplay on the distribution of membership is relevant to multiple categories, e.g. activity and branch of knowledge (molecular biologists, breeders, genebank managers, bio-informatics), geographical (developing and OECD country), sector (public and private). In keeping with the scope of this memorandum, we consider private sector involvement to illustrate the argument. According to interviewees, industry requires clear incentives to participate in any type of collaborative arrangement. Research programs that provide exclusive access to precompetitive knowledge or results can induce private sector involvement. Research that pools scientific and technical human resources to produce products and processes that are precompetitive, publically available and of high value can attract private sector involvement. Pooling of data, materials and information that have precompetitive value may also be an inducement. In none of our cases to date, interviewees have affirmed that private sector companies voluntarily contribute internally developed data or information into a pool.

Membership defines the boundaries of the initiative. Although several cases include a heterogeneous set of actors, interviewees also admit that not all needs can be accommodated within the same initiative. Whether or not programs include private sector actors provides an example of what we mean. The cases indicate that private sector involvement proceeds down one of three tracks. In one, the program works with large private sector companies that pool substantial financial resources from industry for the development of new publically available products that are useful to all members but too expensive for one company to produce independently. In another, programs seek out small and medium sized firms, not large firms. SMEs often provide better partners for data sharing initiatives as they are often more resource limited and more willing to pool resources. Finally, some programs make no effort to appeal to integrate the private sector. The exclusion of the private sector is not always related to property rights considerations – some interviewees have noted that universities often drive harder bargains than companies. Rather, exclusion of the private sector is often related to choices the program makes about the membership, resources and access.

*Leadership structure and attributes*

Case studies have not demonstrated that a single CEO form (or strong leader model) is consistently appropriate. Leadership structure depends on the goals, membership and resources embedded in the program. In most of the cases studied, leaders are members of a small team of highly respected individuals who have demonstrated competencies in the field and/or represent the key membership constituencies. Yet, leadership qualities also depend on the political context within which the program operates. Highly politicized contexts require a leader who is above all highly respected by the key constituencies. Leaders and leadership teams are only effective when they have sufficient financial resources that can be applied to accomplish goals and demonstrate the value of common and shared resources from initial stage.

The food and agriculture domain is populated by diverse array of genetic resources related initiatives. It can be characterized by having 'moderate to high' levels of political and legal uncertainty. The domain also includes a heterogeneous set of actors. This complexity is likely to impact goal setting, data sharing, resources and membership, and calls for a leadership model that is multifaceted and incorporates the range expertise that instills confidence among all key constituencies.

*Management capacity*

Management emerges as an important aspect across all analyzed cases. All initiatives have invested considerably in a reliable and efficient management structure/team/system. Management generally includes scientific support, community building and outreach activities, and technical expertise. Interviewees have highlighted how management is fundamental to create value and build trust towards the capacity of the initiative to achieve its goals. Actors will not consider engaging in any project, if they do not recognize its technical, financial and operational capacity. Management activities are necessary to ensure a high quality product that respond and quickly adapt to the needs of its members and users. Management also guarantees the quality and reliability of internal processes that address sensitive issues such as security and intellectual property.

**Next steps**

For the project report, the research team will create a template for the presentation of the in-depth case studies, reflecting metrics of case study evaluation, and synthesize research findings in both descriptive and analytical modalities.

The DivSeek January workshop will be an important moment to present the findings of the research project and facilitate the discussion of governance options for DivSeek. The research team stands ready to contribute to the workshop based on the guidance by the Steering Committee and in cooperation with the Joint Facilitation Unit.

*Appendix 1*

**First evaluation of case studies: criteria**

## 1. Resources exchanges

**Criteria:** Programs in which scientists and other actors exchange and use both data and genetic materials, under different legal constraints.

**Rationality:** While both data and material sharing might be restricted by legal constraints (Chokshi, Parker, & Kwiatkowski, 2006; Contreras, 2014) , material sharing is more exposed to legal burdens at the national and international level (Eric Welch & Selim Louafi, n.d.).

**Metrics:**

| Flows | |
|---|---|
| Data | Does the project provide access to data? |
| Material | Does the project provide access to material? |

## 2. Access terms of contributions to the pool

**Criteria:** Programs in which individuals and organizations freely access and use data and materials, but are also encouraged to contribute back to the common pool by sharing their data and materials. Programs with a good and clear data use and contribution policy have been preferred.

**Rationality:**

- **Access policies:** Different legal contexts might affect the contracts that actors can stipulate among each other and might limited the opportunities for data and material exchange. For instance, the existence of IP rights might limited the use and diffusion of data and might prevent actors from freely sharing their resources (Chokshi et al., 2006, 2006; Contreras, 2014; Eckersley et al., 2003; Eric Welch & Selim Louafi, n.d.; Kosseim et al., 2014)
- **Presence of data policies:** As legal issues might affect the use and contribution of data and material, data-sharing initiatives have to develop their own policies to facilitate the exchanges among involved actors (Elta Smith, n.d.; Knoppers, 2009; Kosseim et al., 2014).
- **Data production**: Whether resources are collectively produced, sharing and governance are easier than when resources are individually collected and produced, as researchers might seek for more recognition and reward (Chokshi et al., 2006)

**Metrics:**

| Access policies | |
|---|---|
| Open | The access is open and free, even if users may be required to register or agree with the use/contribution policy |
| Specified access | Access is reversed to specific groups (i.e. members) |
| **Presence of data policies** | |

| | |
|---|---|
| Users | Rules that regulate the use of data and material |
| Contributors | Rules that regulate users' contribution |
| **Data production** | |
| Internal | Data are produced by the partners / members of the projects |
| External contribution | Data can be uploaded by anyone who respects the contribution policy |
| Aggregator of public data | Data are collected through publicly available datasets |

## 3. GOALS

**Criteria:** Programs that aim at integrating multiple goals such as fundamental research, scientific and technical innovation and sharing or whose goals have evolved over time to integrate multiple interests.

**Rationality:** Differences in goals might negatively affect individuals and organizations motivation for sharing and collaboration, while overlapping goals might foster interactions among involved actors and increase the motivation to participate in the initiative (Foster & Sharp, 2007; Powell, White, Koput, & Owen - Smith, 2005; Strandburg, Frischmann, & Cui, 2014).

**Metrics:**

| **Goals typology** | |
|---|---|
| Database | The project offers access to data and material produced by members, other institutions or upload by contributors. The project does not have any research goal on its own. |
| Research project | The project aims at pursuing its own agenda or supporting research initiatives |
| Platform | The project provides an open platform for the collection, management and sharing of data and material. |

## 4. TRANSDISCIPLINARITY

**Criteria:** Programs that include a diversity of stakeholders, including from public, non-profit and private sector, and from different regions in the world present one of the many dimensions of heterogeneity. Heterogeneity across actors is likely to increase the numbers of barriers that data sharing projects should have to overcome (Dedeurwaerdere, 2006).

**Rationality:** Whereas stakeholders belong to the same institutional background or the same professional / scientific field, shared codes and paradigms facilitate knowledge exchange and collaboration towards common goals, (Dove, Faraj, Kolker, & Özdemir, 2012; Möllering, 2005; Tasselli, Kilduff, & Menges, 2015; Tsai & Ghoshal, 1998). Different codes and paradigms might stress differences might create conflicts and tensions. Diversity might reduce the level of trust embedded in the relationships among actors which is positive related to the exchange of knowledge (Knoppers, 2009; Kosseim et al., 2014; Tsai & Ghoshal, 1998).

**Metrics:**

| **Partners** | |
|---|---|
| Public | Public institutions are part of the project |

| | |
|---|---|
| Private | Private institutions are part of the project |
| Non profit | Non-profit institutions are part of the project |
| Research | Universities and research centers are part of the project |
| Main partners | Funders |
| **Geographical characteristics** | |
| Focus of the project | Geographic area to which the project aims at providing benefits |
| Stakeholders nationality | Main stakeholders nationality |
| **Developing countries & Breeders** | |
| Developing countries | The project collaborates or involves actors from developing countries |
| Breeders | The project collaborates or involves breeders in training and education activities |

**Criteria:** Programs that offer innovative tools for the analysis, storage and management of data will be preferred, assuming equal conditions in the other criteria.

**Rationality:** Genomics research is becoming increasingly data-driven. Scientists need instruments for the analysis, storage and management of data (Bouffard et al., 2010; Schroeder, Gonzalez-Perez, & Lopez-Bigas, 2013). The possibility to access to technical tools is one of the reason that might incentive or reduce the propensity to share.

**Metrics:**

| **Platform characteristics** | |
|---|---|
| Access to germplasm data | The platform provides access to public datasets |
| Tools for data management | The platform provides tools to manage data |
| Web based / downloadable | The platform is available only online / The platform may be downloaded |
| Upload or no of own data | Users can upload and manage their own data |
| Private / public options | Users are allowed to choose with whom and to what extent share their data |
| Storage facilities | The platform provides storage facilities for users |
| Visualization tools | The platform provides visualization tools |
| Open source code | The code of the platform is open source |

***Appendix 2***

**Final selection of case studies**

| *Project name* | *Private actor engagement* | *Heterogeneity of actors involved* | *Developing country involvement* | *Involvement of members in decision making* | *Data sharing policies* | *IT infrastructure complexity* |
|---|---|---|---|---|---|---|
| **Integrated Breeding Platform** | Medium | Medium | Yes | Low | Decentralized | High |
| **iPlant** | Low | Medium/High | No | Low | Decentralized | High |
| **Open Science Grid** | None | Low | No | Medium | Decentralized | High |
| **Structural Genomics Consortium** | Very high | Medium | No | High | Centralized | Low |
| **Global Alliance for Genetics and Health** | High | High | Yes | High | Centralized and decentralized | Low |
| **IRIC** | Low | | Yes | Medium | | Medium |
| Seeds of discovery | Low | Low/Medium | Yes | Low | Centralized | Medium |
| Cacao Genome Database | High | Low | Yes | Low/High | Centralized | Medium |
| Cassavabase NextGen Cassava | Low | Low | Yes | Low | Centralized | Medium / High |
| SciCrunch | Low | Low | No | Low | Decentralized | Medium |

*Note: the table represents the evaluation of the case studies **before** completing all the interviews. Thus, errors in the evaluation of the criteria might be due to the low availability of data (one interview and collection of information through the website).*

DS/SC-2/15/7

# Proposal for a Bellagio conference on the science/policy interface in relation to crop germplasm

*Information note by the Joint Facilitation Unit*

This document contains a concept note for a Bellagio conference. The CGIAR Consortium Office introduced the idea in consultation with the Chairperson of the DivSeek Assembly and prepared the note, with inputs from other members of the Joint Facilitation Unit.

The Committee is invited to consider the proposal and advise on the merits.

THE ROCKEFELLER FOUNDATION

**Conference, Practitioner, and Academic Writing Proposal Questions**

*Please answer the questions below or in a separate document. Please include your name and contact information on all submitted materials. After answering the questions, please upload it to your online application as an attachment.*

Name: TBD

What is the purpose and goal of your project? (250 words)

The conference will explore the role of plant genetic resources in the context of the United Nation's Sustainable Development Goals (SDG). Achieving these SDGs will require producing enough food over the next 15 years to feed the planet's 9 billion people, including the ~800 million people who currently go to bed hungry every night. The largely untapped potential of plant genetic resources, which is explicitly acknowledged in the SDG, can contribute to enhancing agricultural productivity and climate resilience in an environmentally and socially sustainable manner. Global thought leaders in crop diversity and its role in agriculture will work with other experts to produce a position paper to advance an ongoing science-based global dialogue about how the informed use of plant genetic resources can help adapt agriculture to a changing climate and augment food and nutritional security, particularly in the developing world. Specifically, discussions will consider interactions between emerging scientific opportunities and policy challenges related to securing, managing and using plant genetic resources, with the ultimate goal of maximizing the contribution of crop diversity to the

resilience and productivity of agri-food systems. The *Diversity Seek* initiative (DivSeek) (www.divseek.org) is a new voluntary initiative of global players dealing with conservation, research and regulation of crop germplasm for food and agriculture. DivSeek partner organizations will convene the Bellagio meeting.

What forces are at play now that create an opportunity for positive change on the issue? (150 words)

Game-changing and potentially disruptive DNA-sequencing technologies, big-data platforms, high performance computing, image-based phenotyping methods, gene-editing techniques, and synthetic biology concepts are revolutionizing the way crop diversity is managed and used. Taken together, these technologies have the potential to promote a 'decoupling' between the biological samples conserved in genebanks and the exploitation of a rapidly growing informational domain associated with these samples in crop and animal breeding programs. This trend, sometimes referred to as the 'dematerialization of genetic resources', has led to renewed conversations about global stewardship and equity, both at the level of individual governments and in international fora. There is now a window of opportunity to propose and advocate innovative solutions to global commons policy issues to contribute to a food-secure future.

How does your work relate to the Rockefeller Foundation's Focus Areas—advancing health, transforming cities, securing livelihoods, and revaluing ecosystems—or the wellbeing of humanity more generally? (150 words)

DivSeek is supported by a coalition of the *Global Crop Diversity Trust* (GCDT; www.croptrust.org), the *International Treaty on Plant Genetic Resources for Food and Agriculture* (ITPGRFA; www.planttreaty.org), the Consortium Office of the *Consultative Group on International Agricultural Research* (CGIAR; www.cgiar.org), and the *Global Plant Council* (GPC; http://globalplantcouncil.org). All these organizations contribute to securing farmers' livelihoods and the wellbeing of humanity in general, by enhancing the productivity, resilience, and sustainability of global agriculture through a greater and more informed use of crop diversity. GCDT is raising an endowment fund to secure the conservation of humanity's most important crop genetic resources in perpetuity. ITPGRFA regulates the exchange of and benefit sharing from crop germplasm, and the Global Information System on crop germplasm. CGIAR genebanks are the most significant global custodians, users and dissemination centers of crop diversity for agriculture in developing countries. GPC is a coalition of national, regional and international plant, crop, agricultural, and environmental scientific societies. It represents over 55,00 plant and crops experts across the globe with a mission to promote plant research, teaching and training around the world.

What is the impact on poor or vulnerable populations? (150 words)

DivSeek's mission is to support "*...activities that harness the power of crop diversity for food and nutritional security and societal and economic benefits...*" The DivSeek Steering Committee specifically promotes the participation of gene bank managers, scientists working with natural variation, and policy makers from developing countries in the initiative. The purpose of the endowment fund being raised by GCDT is to support genebanks in developing countries. Ongoing work of the CGIAR targets poor small- and medium-holder farmers and both rural and

urban consumers, particularly women and children, in Africa, Asia, and poverty hotspots in Latin America. ITPGRFA manages the Benefit Sharing Fund, which targets crop-diversity management in developing countries. The majority of the conference participants will be from developing countries. Together these factors will ensure that the impact of the proposed project will be primarily directed towards farmers in developing countries.

What is the specific phase of the project you will work on while at the Center? (150 words)

The goals of the workshop will be to identify a research agenda for collaborative actions that can shape new scientific policy, training and communication initiatives to accelerate the responsible exploration, characterization and utilization of plant genetic resources to augment the productivity and climate resilience of agri-food systems in the developing world. We will bring together experts from diverse intellectual communities, including policy makers, biological scientists, genebank managers, information technologists and education/communication specialists to craft the research agenda for a more efficient and equitable functioning of the existing global commons policy, which unleashes the potential of world's germplasm holdings to catalyze a new era of rapid crop improvement designed to enhance food and nutritional security throughout the developing world.

What outputs/products do you anticipate will result from your conference/residency (Please check all that apply)

☐ Book Chapter

☐ Article for Peer Reviewed Publication

☐ Article for General Publication

☐ Strategic / Organizational Plan

☑ White Paper

☐ Grant Proposal

☑ Policy Initiative

☐ Private Report

☐ Artwork

☐ Poetry

☐ Musical Composition

☐ Translation

☐ Other (please explain)

What are the goals for your conference/residency that will result from the outputs listed above? (Please check all that apply)

☑ Generate/disseminate new knowledge/research

☑ Develop or disseminate solutions to existing problems

☐Support creation of a new organization

☑ Form new networks/alliances for problem solving

☑ Enable emergence of new global initiatives through existing organizations

☐Inspire thought or action through artwork or literature

☐Other (please explain)

How will the outputs be disseminated and to whom? (150 words)

The conference outcomes will be published in peer-reviewed articles in authoritative journals in the field of Genetic Resource Policy, e.g. Nature and Science. The white paper will be distributed throughout CGIAR, Crop Trust, Global Plant Council, ITPGRFA and their partners. It will inform an emerging global research agenda on plant genetic resources policy and formulate a coordination mechanism for taking up such an agenda.

## General Information

**Contents:**

1. **ARRIVAL/TRANSFER**
2. **LOCAL TRANSPORTATION/HOTEL – GCDT OFFICE**
3. **IMPORTANT INFORMATION**

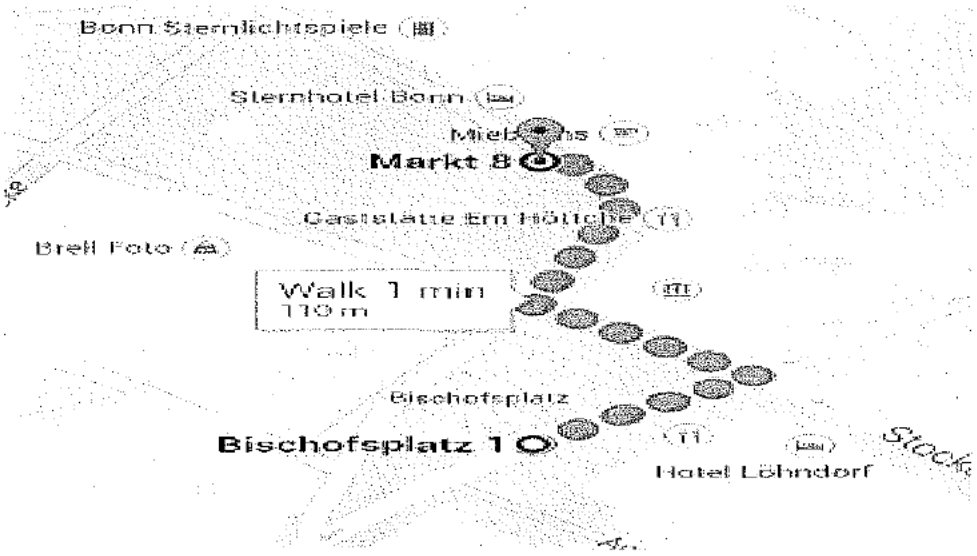
### 1. ARRIVAL/TRANSFER

Directions from various airports to Bonn City Centre and to Sternhotel Bonn

**Stern Hotel Bonn, Markt 8, 53111 Bonn**

**A. Cologne/Bonn Airport – Bonn Center**

Take the express bus SB 60, Cologne/ Bonn Airport – towards Main Station Bonn (Approx. 28 minutes travel time). Get off at Bonn Markt (one station before Hauptbahnhof/ main station). From there, it's just a minute walk to the Sternhotel Bonn. Cost per one-way travel: EUR 7.50 (please make sure you have coins to pay for your ticket, since machines sometimes do not accept notes). Buses depart twice an hour from 4am till 11 pm.

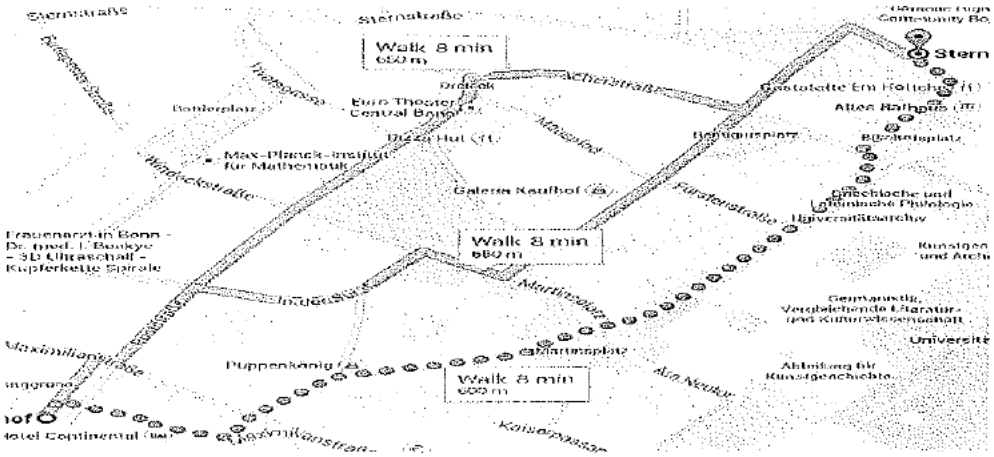


**B. Frankfurt Airport – Bonn**

Upon arrival in Frankfurt please make your way to the railway station (for long-distance trains), which is located within the airport. You have two options.

**Option 1**: Frankfurt Airport – Bonn Hbf

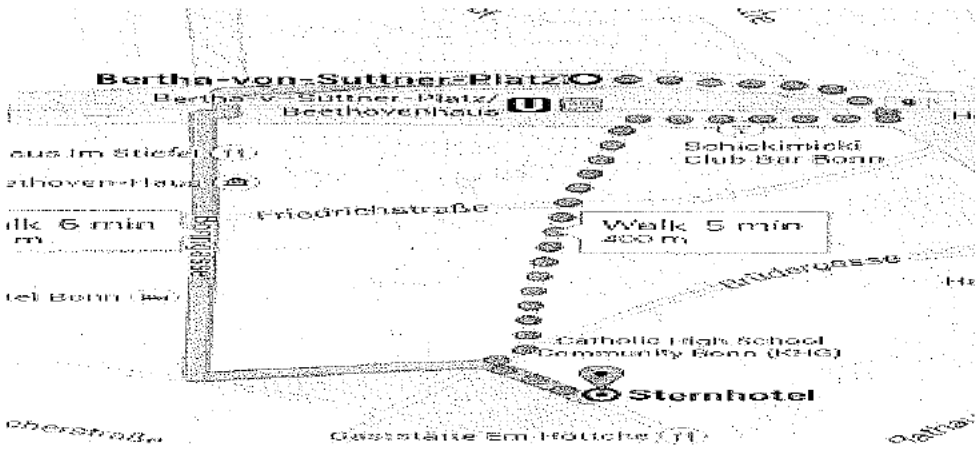
There are direct trains (IC-trains) from Frankfurt Airport to Bonn Main station. Trains go about every two hours and take 2.30 hrs. You will arrive at Bonn Main Station. From Bonn Main Station to the hotel is about 8 min by foot.



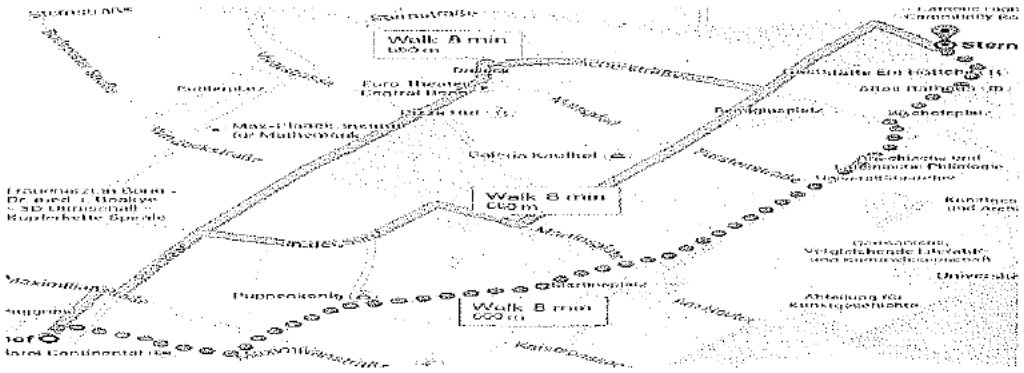

**Option 2**: Frankfurt Airport – Siegburg - Bonn

A slightly faster option is to take the ICE to Bonn/Siegburg. Journey time is about 45 minutes. Get off at Bonn/Siegburg. From the platform take the elevator down one level. You will reach the Tram station. Depending on the ticket you are holding you might need to buy a separate tram ticket (approx. EUR 4.60, remember to have cash and/ or change!). Get on the tram. As they are only going one direction from Siegburg, you can't go wrong. The journey to Bonn Centre takes about 20 min. Get off at Bertha-von-Suttner-Platz. From there it is a 5 minute walk to the hotel.

For timetables please visit: http://www.bahn.com/i/view/GBR/en/



**C. Dusseldorf Airport –**
From the terminal make your way to the airport railway station. Trains from Dusseldorf Airport to Bonn Hauptbahnhof depart every two hours. The journey takes about 1 hour and the ticket cost approx. EUR 20.00 one way. For timetables please visit: http://www.bahn.com/i/view/GBR/en/
Get off at Bonn Hauptbahnhof/ Bonn Main station. From there it is about a 8 min walk to the hotel.

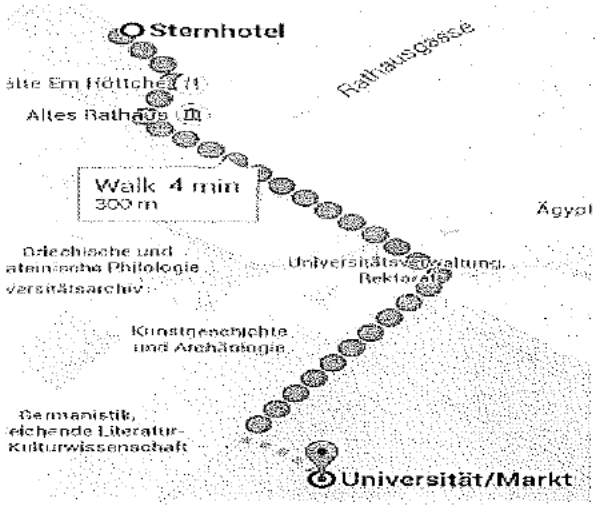



**TICKETS:** Sometimes the rail tickets are included in the airfare. Your itinerary will say 'Frankfurt Airport – Railway Germany' or 'Cologne Station'. Operated by 9B. In this case you will go to the Deutsche Bahn ticket machine, type in the number indicated on your ticket and collect your rail ticket.
Some receive an "electronic rail ticket" with a reservation number. Same procedure here: go to the ticket machine and collect your ticket. There is DB-staff to help you out. Just ask for help.

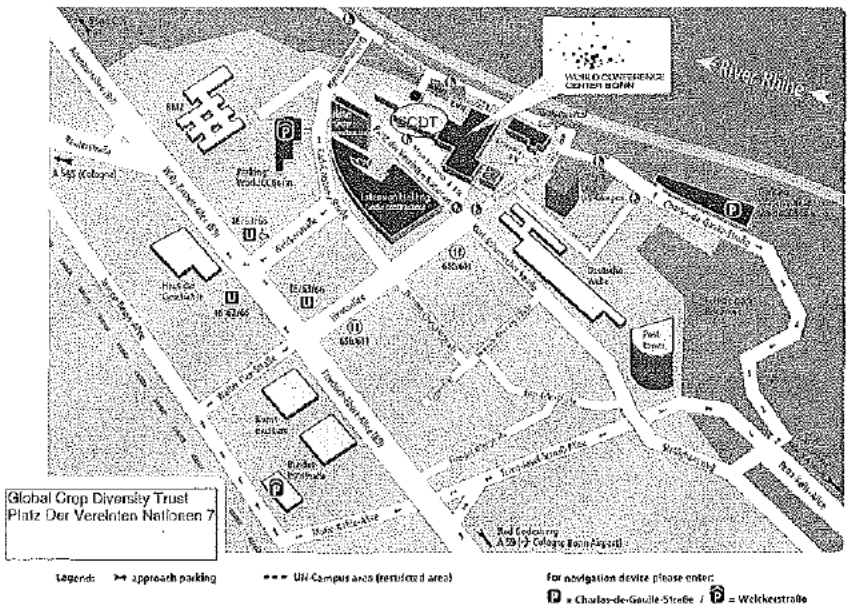
## 2. LOCAL TRANSPORTATION/STERNHOTEL BONN – GCDT OFFICE

**Ticket Purchase:** The name of the type of ticket you should get is 1b and the price of a one-way ticket is 1,80 Euro. Please remember to have coins. All underground stations have a ticket machine inside. Please purchase them beforehand at a machine.

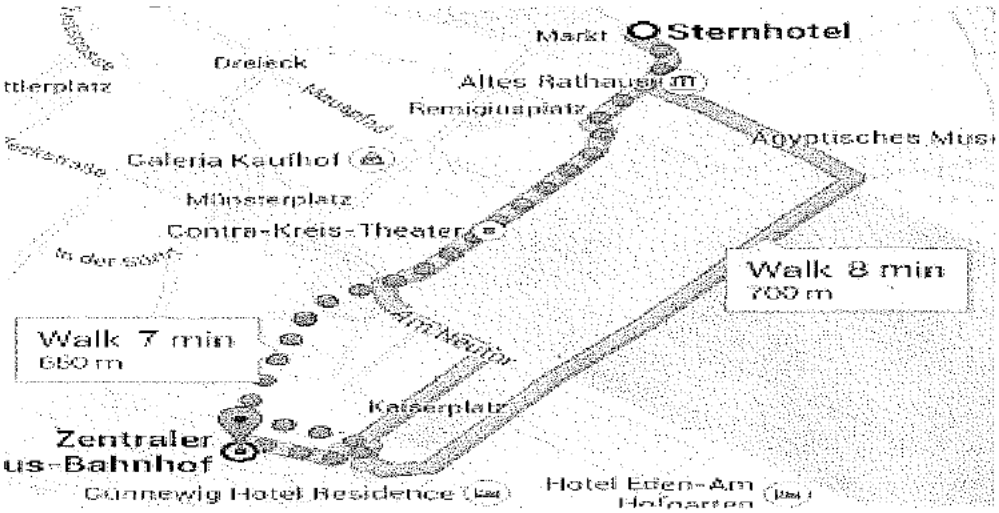
**By Underground:** underground tram station Universitaet / Markt (towards Bonn - Bad Godesberg). Take tram no. 16 or 63 (towards Bad Godesberg), or 66 (towards Bad Honnef,Ramersdorf). You get off at the stop Heussallee/Museumsmeile. Refer to map below for a guide on how to get from your hotel to the station.




Heussallee Station to Platz der Vereinten Nationen 7/ GCDT Office (please see map below)



**By Bus:** Walk towards the Bonn Central Station and look for the bus stop marked "C4". Take either the 610 or the 611 towards Bad Godesberg. Busses come every 10 minutes. You get off at the stop called Deutsche Welle. Please look at the map below for guide on how to get from the hotel to the station. Tickets cost 2.80 EURO per one way and can be purchased from the driver.




## 3. IMPORTANT INFORMATION

### WEATHER
Please visit https://weather.com for current and future information

### MONEY EXCHANGE
If you need to exchange money, please do so upon arrival at the airport. Most Hotels in Bonn do not offer this service. There are banks in Bonn where you will be able to withdraw money with your credit card (fees may apply). You should get in touch with your bank or credit card provider to tell you which banks in Germany offer free cash withdrawal or the one with the lowest fees (For example: If you are with Barclay's, you may withdraw free of charge in Deutsche Banks around Germany and Spain because of the Global Alliance). Credit Cards are **not** accepted everywhere. If you need to exchange money in the city, there is a Western Union Branch in Bonn Central Station.

### PASSPORT & VISAS
It is in your own responsibility for providing a valid passport and visa (if applicable) to enter Germany.

### LOCAL INFORMATION
https://www.bonn.de/tourismus_kultur_sport_freizeit/tourist_information_aktuell/tourismus_und_hotels/?lang=en

**Bretting, Peter**

| | |
|---|---|
| **From:** | E. Marden [redacted] |
| **Sent:** | Tuesday, December 01, 2015 9:33 PM |
| **To:** | Susan McCouch; Susan McCouch |
| **Cc:** | Bretting, Peter |
| **Subject:** | Fwd: DivSeek meeting in Bonn - Comment on ASU Report |
| **Attachments:** | DS_SC-2_15_5 (Update on ASU study).pdf |

Dear Susan,

I wanted to touch base before next week's meeting about the Report of the ASU Study (attached to the original distribution and included below). There was very little visibility into the development of this report - and I am a little concerned about what we are going to do with it. I am copying Peter Bretting on this email as he was also a member of the Governance Expert Committee. He and I have not discussed the ASU Report, and I welcome his comments as well.

1. As I understand it, the ASU study was co-funded jointly by Crop Trust and the Treaty to generate an independent report on governance as it relates to data sharing and open source in order to help guide DivSeek on this front. I have gone back through a series of emails on the report - and though my understanding is correct as the original intent of the ASU study, I can see where over the course of the year, this work stream seems to have been characterized more and more broadly. It now seems to cover everything, without disclosing the research or findings. Moreover, Daniele - of the Treaty - is now an author. So, two major issues: (a) we don't know what the report actually says; and (b) it is now arguably a Treaty perspective.

2. The report's authors propose to present the full findings at the January 8 "roundtable." I think this is unwise. As above, we don't know the content of the report and it is arguably a Treaty document. The roundtable in January will only help further DivSeek if it is a tightly orchestrated session presenting paths forward. I had hoped that next week's SC meeting would result in a recommendation on governance that would allow the organization to focus on substance. If we introduce a separate report, we will likely lose ourselves in endless discussion. Further, if the report is presented at the Jan 8 meeting, the Assembly will (rightly) perceive it as a SC-endorsed output.

I am available to discuss if that is helpful.

Best regards,

Emily

Begin forwarded message:

**From:** Peter Wenzl <peter.wenzl@divseek.org>
**Subject: DivSeek meeting in Bonn**
**Date:** December 1, 2015 at 2:44:00 AM PST

Duplicate email trail removed

DS/SC-2/15/5

# Report from the governance research project

This document contains a memorandum developed under the governance research project that Arizona State University (ASU), in cooperation with CIRAD, is conducting to present DivSeek with options for governance, data sharing and membership.

The Committee received a first progress report by the research team at its first meeting.

The Committee is invited to appraise the memorandum and advise on the next steps, including for the preparation of the DivSeek workshop/roundtable in January 2016.

**DivSeek Governance Research Project**
**Memorandum for the Steering Committee**
***Eric Welch, Selim Louafi, Federica Fusi and Daniele Manzella***
**8 December 2015**

## Purpose of the memorandum

Following the progress report to the first meeting of the Steering Committee in May 2015, this memorandum is to update the Committee at its second meeting, on progress made with the governance research project. The memorandum also contains preliminary observations derived from the research, on themes that are on the agenda of the Steering Committee.

This *interim* memorandum does not explicitly address the broader policy context in which the initiative takes place. Constraints and sensitivities deriving from such a context are likely to impact goal setting, data sharing, resources and membership, and the DivSeek community is encouraged to consider them in conjunction with the findings of the research project, for instance when addressing developing country involvement and capacity development. This memorandum focuses on general governance aspects that are applicable across the case studies analyzed in the research project, independently from their policy context.

## Methodology and case study selection

As outlined briefly in the first progress report of May 2015, the research team has organized the work in three main phases:

(1) Literature review and sample frame development. In line with the objectives of the study, nineteen cases in the area of food and agriculture and seven cases in the human health sector were identified for possible inclusion in the study. Appendix 1 contains the criteria used for evaluating those cases.

(2) Initial selection of ten cases most relevant for DivSeek. Initial selection was followed by 45-60 minute interviews of the ten Project Managers and Executive Directors to identify six cases for in-depth data collection and analysis. Appendix 2 contains the initial list of the ten cases and the final group of six cases.

(3) All six Project Managers and Executive Directors were sent invitations asking them to participate in the in-depth analysis of their organizations. The research team has conducted 45-75 minute interviews with staff, partners and clients of the six organizations, but has not completed all interviews.

This memorandum provides a second report on initial findings to date. The full report, to be delivered by the end of January 2016, will elaborate on the findings presented here.

## Preliminary Findings

Since data collection and analysis under the project are still taking place, the following observations should be considered preliminary. Nevertheless, the research team hopes that the findings and observations will be of use to the DivSeek Committee.

*Goal setting*

Based on the case analyses to date, it is clear that governance structures have strong accordance with program goals. Case study interviewees consistently explain that clear

specification of program goals is an essential first step for program success. All interviewees, whether leadership or staff, were able to articulate program goals and explain how their goals differ from others or fill a particular research or service gap. Cases show that program goals are the substance around which governance structures are built as they facilitate accomplishment of the goals.

Program goals vary across the cases. In some cases, program goals are mainly technical. In others, goals are broader and seek to integrate the values and perspectives of heterogeneous groups through some set of agreed-upon rules for access and behavior of members. Cases demonstrate that specification of the goals is often a slow process requiring two or more years. Some of the interviewees have noted that their projects or programs began small, with only a few actors who were able to achieve consensus to set program goals. Consensus on goals is much more difficult when a large more heterogeneous group of actors is involved. However, homogeneity and small initial size can also reduce inclusivity, ignore obvious complexities and work against long-term buy-in.

These observations are significant for DivSeek. For example, alternative strategies that the DivSeek community might pursue, could 1) aim to establish technical standards critical for advancement of cross-crop genomics research; or 2) aim to establish a heterogeneous community – industry, OECD country universities and national agriculture research organizations in developing countries – around technical standards, capacity development and sharing norms. Governance structures for the former will certainly be less complex than the latter. Both strategies require a reflection about which actors to include in the goal setting process and which actors might be invited later.

*Defined resources*

Case analyses to date show that governance structure also depends on the resource(s) that the program/project seeks to develop and make available to the community. The choice of resources is usually dependent on the identification of a demand that arises from the field. All interviewees to date were able to identify specific human, software, data, standardization, negotiation/brokering, material, computational or storage resources that are not offered (or are of low quality, or are poorly distributed) by other programs. They talked about how their organizations developed strategies to fill the identified gap. DivSeek should consider specifying the resources that 1) are not provided elsewhere; 2) are in high demand; and 3) may stimulate convergence across different sub-communities. Subsequently, governance structure should establish who has access to the common resources and under what conditions.

*Sharing policies*

There appear to be two main approaches to data sharing: centrally determined or locally autonomous. Centrally determined rules make products/data/findings 1) completely open such that all products/data/findings are publicly available; 2) partially open such that there is an embargo on products/data/findings of one to two years before making products publicly available; or 3) closed such that only members share among themselves (This last one is a theoretical category. None of the cases are completely closed). Organizations with centralized sharing policies can design monitoring and controlling mechanisms to prevent opportunistic behaviors, such as unbalanced contributions of resources to the common pool. Thus far, cases show that imposing centralized rules for resource sharing does not always encourage sharing behavior; actors may withdraw from organizations where there are sharing requirements.

Locally autonomous systems allow actors to decide whether or not to share products/findings/data publicly. Locally autonomous systems often include flexibility that leaves actors free to decide with whom and under what conditions they will share. Some of these projects create and manage formal and informal services such as match-making, brokerage activities, communication or technical support that build trust among members, with the ultimate aim of encouraging sharing.

*Membership, including by private sector*

The interplay between resources and sharing policies influences membership. Different types of resources and governance structures provide alternate incentives for participation. The influence of such an interplay on the distribution of membership is relevant to multiple categories, e.g. activity and branch of knowledge (molecular biologists, breeders, genebank managers, bio-informatics), geographical (developing and OECD country), sector (public and private). In keeping with the scope of this memorandum, we consider private sector involvement to illustrate the argument. According to interviewees, industry requires clear incentives to participate in any type of collaborative arrangement. Research programs that provide exclusive access to precompetitive knowledge or results can induce private sector involvement. Research that pools scientific and technical human resources to produce products and processes that are precompetitive, publically available and of high value can attract private sector involvement. Pooling of data, materials and information that have precompetitive value may also be an inducement. In none of our cases to date, interviewees have affirmed that private sector companies voluntarily contribute internally developed data or information into a pool.

Membership defines the boundaries of the initiative. Although several cases include a heterogeneous set of actors, interviewees also admit that not all needs can be accommodated within the same initiative. Whether or not programs include private sector actors provides an example of what we mean. The cases indicate that private sector involvement proceeds down one of three tracks. In one, the program works with large private sector companies that pool substantial financial resources from industry for the development of new publically available products that are useful to all members but too expensive for one company to produce independently. In another, programs seek out small and medium sized firms, not large firms. SMEs often provide better partners for data sharing initiatives as they are often more resource limited and more willing to pool resources. Finally, some programs make no effort to appeal to integrate the private sector. The exclusion of the private sector is not always related to property rights considerations – some interviewees have noted that universities often drive harder bargains than companies. Rather, exclusion of the private sector is often related to choices the program makes about the membership, resources and access.

*Leadership structure and attributes*

Case studies have not demonstrated that a single CEO form (or strong leader model) is consistently appropriate. Leadership structure depends on the goals, membership and resources embedded in the program. In most of the cases studied, leaders are members of a small team of highly respected individuals who have demonstrated competencies in the field and/or represent the key membership constituencies. Yet, leadership qualities also depend on the political context within which the program operates. Highly politicized contexts require a leader who is above all highly respected by the key constituencies. Leaders and leadership teams are only effective when they have sufficient financial resources that can be applied to accomplish goals and demonstrate the value of common and shared resources from initial stage.

The food and agriculture domain is populated by diverse array of genetic resources related initiatives. It can be characterized by having 'moderate to high' levels of political and legal uncertainty. The domain also includes a heterogeneous set of actors. This complexity is likely to impact goal setting, data sharing, resources and membership, and calls for a leadership model that is multifaceted and incorporates the range expertise that instills confidence among all key constituencies.

*Management capacity*

Management emerges as an important aspect across all analyzed cases. All initiatives have invested considerably in a reliable and efficient management structure/team/system. Management generally includes scientific support, community building and outreach activities, and technical expertise. Interviewees have highlighted how management is fundamental to create value and build trust towards the capacity of the initiative to achieve its goals. Actors will not consider engaging in any project, if they do not recognize its technical, financial and operational capacity. Management activities are necessary to ensure a high quality product that respond and quickly adapt to the needs of its members and users. Management also guarantees the quality and reliability of internal processes that address sensitive issues such as security and intellectual property.

**Next steps**

For the project report, the research team will create a template for the presentation of the in-depth case studies, reflecting metrics of case study evaluation, and synthesize research findings in both descriptive and analytical modalities.

The DivSeek January workshop will be an important moment to present the findings of the research project and facilitate the discussion of governance options for DivSeek. The research team stands ready to contribute to the workshop based on the guidance by the Steering Committee and in cooperation with the Joint Facilitation Unit.

*Appendix 1*

**First evaluation of case studies: criteria**

## 1. RESOURCES CHARACTERISTICS

**Criteria:** Programs in which scientists and other actors exchange and use both data and genetic materials, under different legal constraints.

**Rationality:** While both data and material sharing might be restricted by legal constraints (Chokshi, Parker, & Kwiatkowski, 2006; Contreras, 2014) , material sharing is more exposed to legal burdens at the national and international level (Eric Welch & Selim Louafi, n.d.).

**Metrics:**

| Flows | |
|---|---|
| Data | Does the project provide access to data? |
| Material | Does the project provide access to material? |

## 2. ACCESS TERMS OF CONTRIBUTION TERMS OF USE

**Criteria:** Programs in which individuals and organizations freely access and use data and materials, but are also encouraged to contribute back to the common pool by sharing their data and materials. Programs with a good and clear data use and contribution policy have been preferred.

**Rationality:**

- **Access policies:** Different legal contexts might affect the contracts that actors can stipulate among each other and might limited the opportunities for data and material exchange. For instance, the existence of IP rights might limited the use and diffusion of data and might prevent actors from freely sharing their resources (Chokshi et al., 2006, 2006; Contreras, 2014; Eckersley et al., 2003; Eric Welch & Selim Louafi, n.d.; Kosseim et al., 2014)
- **Presence of data policies:** As legal issues might affect the use and contribution of data and material, data-sharing initiatives have to develop their own policies to facilitate the exchanges among involved actors (Elta Smith, n.d.; Knoppers, 2009; Kosseim et al., 2014).
- **Data production:** Whether resources are collectively produced, sharing and governance are easier than when resources are individually collected and produced, as researchers might seek for more recognition and reward (Chokshi et al., 2006)

**Metrics:**

| Access policies | |
|---|---|
| Open | The access is open and free, even if users may be required to register or agree with the use/contribution policy |
| Specified access | Access is reversed to specific groups (i.e. members) |
| **Presence of data policies** | |

| | |
|---|---|
| Users | Rules that regulate the use of data and material |
| Contributors | Rules that regulate users' contribution |
| **Data production** | |
| Internal | Data are produced by the partners / members of the projects |
| External contribution | Data can be uploaded by anyone who respects the contribution policy |
| Aggregator of public data | Data are collected through publicly available datasets |

## 3. GOALS

**Criteria:** Programs that aim at integrating multiple goals such as fundamental research, scientific and technical innovation and sharing or whose goals have evolved over time to integrate multiple interests.

**Rationality:** Differences in goals might negatively affect individuals and organizations motivation for sharing and collaboration, while overlapping goals might foster interactions among involved actors and increase the motivation to participate in the initiative (Foster & Sharp, 2007; Powell, White, Koput, & Owen - Smith, 2005; Strandburg, Frischmann, & Cui, 2014).

**Metrics:**

| **Goals typology** | |
|---|---|
| Database | The project offers access to data and material produced by members, other institutions or upload by contributors. The project does not have any research goal on its own. |
| Research project | The project aims at pursuing its own agenda or supporting research initiatives |
| Platform | The project provides an open platform for the collection, management and sharing of data and material. |

## 4. INTERDISCIPLINARITY

**Criteria:** Programs that include a diversity of stakeholders, including from public, non-profit and private sector, and from different regions in the world present one of the many dimensions of heterogeneity. Heterogeneity across actors is likely to increase the numbers of barriers that data sharing projects should have to overcome (Dedeurwaerdere, 2006).

**Rationality:** Whereas stakeholders belong to the same institutional background or the same professional / scientific field, shared codes and paradigms facilitate knowledge exchange and collaboration towards common goals, (Dove, Faraj, Kolker, & Özdemir, 2012; Möllering, 2005; Tasselli, Kilduff, & Menges, 2015; Tsai & Ghoshal, 1998). Different codes and paradigms might stress differences might create conflicts and tensions. Diversity might reduce the level of trust embedded in the relationships among actors which is positive related to the exchange of knowledge (Knoppers, 2009; Kosseim et al., 2014; Tsai & Ghoshal, 1998).

**Metrics:**

| **Partners** | |
|---|---|
| Public | Public institutions are part of the project |

| | |
|---|---|
| Private | Private institutions are part of the project |
| Non profit | Non-profit institutions are part of the project |
| Research | Universities and research centers are part of the project |
| Main partners | Funders |
| **Geographical characteristics** | |
| Focus of the project | Geographic area to which the project aims at providing benefits |
| Stakeholders nationality | Main stakeholders nationality |
| **Developing countries & Breeders** | |
| Developing countries | The project collaborates or involves actors from developing countries |
| Breeders | The project collaborates or involves breeders in training and education activities |

### 5. INFORMATICS

**Criteria:** Programs that offer innovative tools for the analysis, storage and management of data will be preferred, assuming equal conditions in the other criteria.

**Rationality:** Genomics research is becoming increasingly data-driven. Scientists need instruments for the analysis, storage and management of data (Bouffard et al., 2010; Schroeder, Gonzalez-Perez, & Lopez-Bigas, 2013). The possibility to access to technical tools is one of the reason that might incentive or reduce the propensity to share.

**Metrics:**

| **Platform characteristics** | |
|---|---|
| Access to germplasm data | The platform provides access to public datasets |
| Tools for data management | The platform provides tools to manage data |
| Web based / downloadable | The platform is available only online / The platform may be downloaded |
| Upload or no of own data | Users can upload and manage their own data |
| Private / public options | Users are allowed to choose with whom and to what extent share their data |
| Storage facilities | The platform provides storage facilities for users |
| Visualization tools | The platform provides visualization tools |
| Open source code | The code of the platform is open source |

***Appendix 2***

**Final selection of case studies**

| *Project name* | *Private actor engagement* | *Heterogeneity of actors involved* | *Developing country involvement* | *Involvement of members in decision making* | *Data sharing policies* | *IT infrastructure complexity* |
|---|---|---|---|---|---|---|
| **Integrated Breeding Platform** | Medium | Medium | Yes | Low | Decentralized | High |
| **iPlant** | Low | Medium/High | No | Low | Decentralized | High |
| **Open Science Grid** | None | Low | No | Medium | Decentralized | High |
| **Structural Genomics Consortium** | Very high | Medium | No | High | Centralized | Low |
| **Global Alliance for Genetics and Health** | High | High | Yes | High | Centralized and decentralized | Low |
| **IRIC** | Low | | Yes | Medium | | Medium |
| Seeds of discovery | Low | Low/Medium | Yes | Low | Centralized | Medium |
| Cacao Genome Database | High | Low | Yes | Low/High | Centralized | Medium |
| Cassavabase NextGen Cassava | Low | Low | Yes | Low | Centralized | Medium / High |
| SciCrunch | Low | Low | No | Low | Decentralized | Medium |

*Note: the table represents the evaluation of the case studies* ***before*** *completing all the interviews. Thus, errors in the evaluation of the criteria might be due to the low availability of data (one interview and collection of information through the website).*

**Bretting, Peter**

| | |
|---|---|
| **From:** | Bretting, Peter |
| **Sent:** | Wednesday, December 09, 2015 7:37 AM |
| **To:** | E. Marden |
| **Cc:** | Susan McCouch |
| **Subject:** | Re: DivSeek meeting in Bonn - Comment on ASU Report |

Hi Emily--greetings from Chicago! Hope that all is well in Bonn.

I had a chance to review the DS/SC-2/15/2 document last night. The contribution by the GCDT articulates well the original intent for DivSeek, and its aspirations. The explanation on p. 4 about responding to the request from the Governing Body to investigate the implications of the technology for the ITPGRFA is on the mark, from my perspective. DivSeek doesn't have the expertise nor mandate for that. It's really up to the Governing Body to conduct such assessments and draw conclusions.

Thanks,

Peter

Peter Bretting
National Program Leader
USDA/ARS Office of National Programs
George Washington Carver Center
4-2212, Mailstop 5139
Beltsville, MD 20705-5139
301-504-5541
Cell [redacted]
peter.bretting@ars.usda.gov

On Dec 3, 2015, at 10:42 AM, E. Marden [redacted] wrote:

> Thanks Peter. As always, your comments are thoughtful and constructive.
>
> Perhaps next week we can acknowledge the interim report, but note that it does not disclose findings. We can ask that the full report be shared with the full SC before a decision about the release and/or presentation.
>
> I agree that it may be very useful on the various other aspects of DivSeek's goals.
>
> Cheers,
>
> Emily
>
> On Dec 3, 2015, at 5:42 AM, Bretting, Peter <Peter.Bretting@ARS.USDA.GOV> wrote:

**Hi Emily and Susan—apologies for the delayed reply. I was fully engaged in a series of job candidate interviews conducted here yesterday, with more today.**

**I don't know much more than y'all do about this research project commissioned by the GCDT and the Treaty earlier this year. Here are some observations/queries.**

1) **At this point I'm not terribly troubled that Daniele Manzella is a co-author of the draft study. Whether or not other members of the JFU (especially the GCDT, but also the Global Plant Forum and the CGIAR) have been consulted during the project, have reviewed early draft reports, and/or provided input and approval are more germane. Perhaps the other JFU members could be queried about their participation.**
2) **The project report should be reviewed and discussed by the Steering Committee (especially its governance subcommittee) before it is released to all the partners. The memorandum provided to the SC now isn't sufficient for such a review. For the final report to be discussed at the January workshop, it should be made available to the SC well-in-advance of the January workshop. If it's not, then presentation of the ASU project report to the Divseek partners should be deferred to subsequent meetings. Recall the notes from the May SC meeting:**

> *23. In conjunction with the decision to convene a governance expert group, the Committee was informed about an on-going research project by Arizona State University (ASU) on institutional and organizational factors for enabling data access, exchange and use, which the Global Crop Diversity Trust and the Secretariat of the International Treaty were co-funding. Mr. Manzella, of the Joint Facilitation Unit and the International Treaty, informed the Committee of the preliminary research activities conducted by the ASU research team for the project, and distributed a progress report. The Committee invited Ms. Marden to coordinate with the ASU research team to obtain early access to the results of the study for consideration as part of the work of the governance expert group.*

3) **Based on the memorandum, the ASU research project is generating a taxonomy of IT and germplasm projects that might be relevant to DivSeek. The taxonomic traits employed include private-sector engagement, heterogeneity of actors involved, developing country involvement, involvement of members in decision making, data sharing policies, and IT infrastructure complexity. From that analysis, the study will examine goal setting, defined resources, sharing policies, membership, leadership structure and attributes, and management capacity. So the ASU research project is much broader than the SC governance sub-committee's work, which has focused on reforming elements of the current governance structure so that DivSeek operates more effectively. The ASU research project**

**and the governance sub-committee effort overlap in the area of leadership/management. But the SC governance subcommittee is more sharply focused on devising an optimal structure for operating the DivSeek organization. Thus the two efforts are somewhat complementary; we'll know for sure once the ASU research project final report is available.**

4) **If anything, all this uncertainty and concern would support reforming the DivSeek organizational structure from the JFU model to an executive operational model.**

**Anyway, hope that this helps, and hope that Susan is enjoying Switzerland!**

**Peter**

**Peter Bretting**
**USDA/ARS Office of National Programs**
**Room 4-2212, Mailstop 5139**
**5601 Sunnyside Avenue**
**Beltsville, MD 20705-5139**
**Phone 1.301.504.5541**
**Fax 1.301.504.6191**
**Mobile Phone** [redacted]
**E-mail** peter.bretting@ars.usda.gov
**Web site:** http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301

**From:** [redacted] [mailto [redacted] **On Behalf Of** Susan McCouch
**Sent:** Wednesday, December 02, 2015 10:32 AM
**To:** E. Marden
**Cc:** Bretting, Peter; Susan McCouch
**Subject:** Re: DivSeek meeting in Bonn - Comment on ASU Report

Dear Emily,

I just arrived in Zurich and am reading the docs sent by Daniele about the ASU governance study now. I would like to chat with you by Skype sometime in the next 2 hours, if at all possible. You are 9 hours behind us, so this is cutting it very short.

It's 4:23pm here now, and I will have to leave for dinner at 6:30pm. If we can't squeeze in a Skype conversation before I leave for dinner, perhaps we could talk when I return? I am not sure when that will be, but I expect I'll be back by 9:30 or 10:00pm.

I'm in meetings all day tomorrow, so the other alternative would be to talk sometime on Fri afternoon, or over the weekend.

Please let me know what works for you. It might be best to catch up once you've heard back from Peter B, who we hope will shed some light on this.

Thanks,
Susan

On Tue, Dec 1, 2015 at 9:32 PM, E. Marden [redacted] wrote:

Dear Susan,

I wanted to touch base before next week's meeting about the Report of the ASU Study (attached to the original distribution and included below). There was very little visibility into the development of this report - and I am a little concerned about what we are going to do with it. I am copying Peter Bretting on this email as he was also a member of the Governance Expert Committee. He and I have not discussed the ASU Report, and I welcome his comments as well.

1. As I understand it, the ASU study was co-funded jointly by Crop Trust and the Treaty to generate an independent report on governance as it relates to data sharing and open source in order to help guide DivSeek on this front. I have gone back through a series of emails on the report - and though my understanding is correct as the original intent of the ASU study, I can see where over the course of the year, this work stream seems to have been characterized more and more broadly. It now seems to cover everything, without disclosing the research or findings. Moreover, Daniele - of the Treaty - is now an author. So, two major issues: (a) we don't know what the report actually says; and (b) it is now arguably a Treaty perspective.

2. The report's authors propose to present the full findings at the January 8 "roundtable." I think this is unwise. As above, we don't know the content of the report and it is arguably a Treaty document. The roundtable in January will only help further DivSeek if it is a tightly orchestrated session presenting paths forward. I had hoped that next week's SC meeting would result in a recommendation on governance that would allow the organization to focus on substance. If we introduce a separate report, we will likely lose ourselves in endless discussion. Further, if the report is presented at the Jan 8 meeting, the Assembly will (rightly) perceive it as a SC-endorsed output.

I am available to discuss if that is helpful.

Best regards,

Emily

Begin forwarded message:

**From:** Peter Wenzl <peter.wenzl@divseek.org>
**Subject: DivSeek meeting in Bonn**
**Date:** December 1, 2015 at 2:44:00 AM PST
**To:** Andreas Graner <graner@ipk-gatersleben.de>, David Marshall <David.Marshall@hutton.ac.uk>, "Elizabeth Arnaud (Bioversity-France)" <e.arnaud@cgiar.org>, Emily Marden [redacted] Peter Bretting <peter.bretting@ars.usda.gov>, "Rajeev Varshney (ICRISAT-IN)" <R.K.Varshney@cgiar.org>, "Ruaraidh Sackville Hamilton (IRRI)" <r.hamilton@irri.org>, Sarah Ayling <sarah.ayling@tgac.ac.uk>, Susan McCouch <srm4@cornell.edu>
**Cc:** Daniele Manzella <daniele.manzella@divseek.org>, Ruth Bastow <ruth.bastow@divseek.org>, "Wayne Powell (CGIAR Consortium)" <w.powell@cgiar.org>, Dagny Poser <dagny.poser@croptrust.org>

Dear DivSeek Steering Committee members,

Please find attached the Agenda and other Documents for the **DivSeek Steering Committee meeting on December 8th** next week. The meeting will start at **9:00am** at the premises of the Crop Trust (Platz der Vereinten Nationen 7, 53113 Bonn).

In a follow-up message later this week, we will send you two additional Documents (No. 2, 3) with notes and updates from Susan and the JFU organizations.

As a reminder, here's a quick overview of the logistics:

- **Accommodation:** Accommodation has been booked at the Hotel Stern, Markt 8, 53111 Bonn, Germany. Accommodation and breakfast have been pre-paid by the Crop Trust. All extras are to be paid by the guest upon check out.
- **Transfer:** please see attached sheet for directions from the airport to the hotel, and from the hotel to the Crop Trust offices.
- **Per diem:** we'll pay you a per diem to cover costs for meals not provided by us. It should also cover incidentals such as transport to/from the airport.
- **Dinner on 7 Dec:** you're invited for a joint dinner at 19:00 on the evening before the meeting at the Ruland Restaurant, Bischhofsplatz 1, Bonn. This is a 2-min walk from the hotel.

Please do not hesitate to contact us in case you've any questions in regard to the agenda, documents or the logistics of the meeting.

Looking forward to meeting you next week!

On behalf of the Joint Facilitation Unit,

Peter
--
Peter Wenzl
DivSeek Liaison
Global Crop Diversity Trust
Platz der Vereinten Nationen 7
53113 Bonn, Germany
Office: +49 228 85427 126
Mobile: [redacted]
www.croptrust.org

*Securing our Food, Forever*

--
Susan McCouch
Professor, Plant Breeding & Genetics
Cornell University
162 Emerson Hall
Ithaca, NY 14853-1901
Phone: +1 607-255-0420
Fax: +1 607-255-6683
Email: srm4@cornell.edu or mccouch@cornell.edu
Alternate Email: [redacted]

**Bretting, Peter**

**From:** E. Marden <[redacted]
**Sent:** Thursday, December 03, 2015 11:42 AM
**To:** Bretting, Peter
**Cc:** Susan McCouch
**Subject:** Re: DivSeek meeting in Bonn - Comment on ASU Report

Thanks Peter. As always, your comments are thoughtful and constructive.

Perhaps next week we can acknowledge the interim report, but note that it does not disclose findings. We can ask that the full report be shared with the full SC before a decision about the release and/or presentation.

I agree that it may be very useful on the various other aspects of DivSeek's goals.

Cheers,

Emily

On Dec 3, 2015, at 5:42 AM, Bretting, Peter <Peter.Bretting@ARS.USDA.GOV> wrote:

> **Hi Emily and Susan—apologies for the delayed reply. I was fully engaged in a series of job candidate interviews conducted here yesterday, with more today.**
>
> **I don't know much more than y'all do about this research project commissioned by the GCDT and the Treaty earlier this year. Here are some observations/queries.**
>
> 1) **At this point I'm not terribly troubled that Daniele Manzella is a co-author of the draft study. Whether or not other members of the JFU (especially the GCDT, but also the Global Plant Forum and the CGIAR) have been consulted during the project, have reviewed early draft reports, and/or provided input and approval are more germane. Perhaps the other JFU members could be queried about their participation.**
> 2) **The project report should be reviewed and discussed by the Steering Committee (especially its governance subcommittee) before it is released to all the partners. The memorandum provided to the SC now isn't sufficient for such a review. For the final report to be discussed at the January workshop, it should be made available to the SC well-in-advance of the January workshop. If it's not, then presentation of the ASU project report to the Divseek partners should be deferred to subsequent meetings. Recall the notes from the May SC meeting:**
>
> ***23. In conjunction with the decision to convene a governance expert group, the Committee was informed about an on-going research project by Arizona State University (ASU) on institutional and organizational factors for enabling data access, exchange and use, which the Global Crop Diversity Trust and the Secretariat of the International Treaty were co-funding. Mr. Manzella, of the Joint Facilitation Unit and the International Treaty, informed the Committee of the preliminary research activities***



duplicate email trail removed

**Bretting, Peter**

| | |
|---|---|
| **From:** | Bretting, Peter |
| **Sent:** | Thursday, December 03, 2015 8:43 AM |
| **To:** | 'Susan McCouch'; E. Marden |
| **Subject:** | RE: DivSeek meeting in Bonn - Comment on ASU Report |

**Hi Emily and Susan—apologies for the delayed reply. I was fully engaged in a series of job candidate interviews conducted here yesterday, with more today.**

**I don't know much more than y'all do about this research project commissioned by the GCDT and the Treaty earlier this year. Here are some observations/queries.**

1) **At this point I'm not terribly troubled that Daniele Manzella is a co-author of the draft study. Whether or not other members of the JFU (especially the GCDT, but also the Global Plant Forum and the CGIAR) have been consulted during the project, have reviewed early draft reports, and/or provided input and approval are more germane. Perhaps the other JFU members could be queried about their participation.**
2) **The project report should be reviewed and discussed by the Steering Committee (especially its governance subcommittee) before it is released to all the partners. The memorandum provided to the SC now isn't sufficient for such a review. For the final report to be discussed at the January workshop, it should be made available to the SC well-in-advance of the January workshop. If it's not, then presentation of the ASU project report to the Divseek partners should be deferred to subsequent meetings. Recall the notes from the May SC meeting:**

*23. In conjunction with the decision to convene a governance expert group, the Committee was informed about an on-going research project by Arizona State University (ASU) on institutional and organizational factors for enabling data access, exchange and use, which the Global Crop Diversity Trust and the Secretariat of the International Treaty were co-funding. Mr. Manzella, of the Joint Facilitation Unit and the International Treaty, informed the Committee of the preliminary research activities conducted by the ASU research team for the project, and distributed a progress report. The Committee invited Ms. Marden to coordinate with the ASU research team to obtain early access to the results of the study for consideration as part of the work of the governance expert group.*

3) **Based on the memorandum, the ASU research project is generating a taxonomy of IT and germplasm projects that might be relevant to DivSeek. The taxonomic traits employed include private-sector engagement, heterogeneity of actors involved, developing country involvement, involvement of members in decision making, data sharing policies, and IT infrastructure complexity. From that analysis, the study will examine goal setting, defined resources, sharing policies, membership, leadership structure and attributes, and management capacity. So the ASU research project is much broader than the SC governance sub-committee's work, which has focused on reforming elements of the current governance structure so that DivSeek operates more effectively. The ASU research project and the governance sub-committee effort overlap in the area of leadership/management. But the SC governance subcommittee is more sharply focused on devising an optimal structure for operating the DivSeek organization. Thus the two efforts are somewhat complementary; we'll know for sure once the ASU research project final report is available.**

4) **If anything, all this uncertainty and concern would support reforming the DivSeek organizational structure from the JFU model to an executive operational model.**

**Anyway, hope that this helps, and hope that Susan is enjoying Switzerland!**

**Peter**

**Peter Bretting**
**USDA/ARS Office of National Programs**
**Room 4-2212, Mailstop 5139**
**5601 Sunnyside Avenue**
**Beltsville, MD 20705-5139**
**Phone 1.301.504.5541**
**Fax 1.301.504.6191**
**Mobile Phone** [redacted]
**E-mail peter.bretting@ars.usda.gov**
**Web site: http://www.ars.usda.gov/research/programs/programs.htm?NP_CODE=301**

**From:** [redacted] [mailto[redacted] **On Behalf Of** Susan McCouch
**Sent:** Wednesday, December 02, 2015 10:32 AM
**To:** E. Marden
**Cc:** Bretting, Peter; Susan McCouch
**Subject:** Re: DivSeek meeting in Bonn - Comment on ASU Report

Dear Emily,

I just arrived in Zurich and am reading the docs sent by Daniele about the ASU governance study now. I would like to chat with you by Skype sometime in the next 2 hours, if at all possible. You are 9 hours behind us, so this is cutting it very short.

It's 4:23pm here now, and I will have to leave for dinner at 6:30pm. If we can't squeeze in a Skype conversation before I leave for dinner, perhaps we could talk when I return? I am not sure when that will be, but I expect I'll be back by 9:30 or 10:00pm.

I'm in meetings all day tomorrow, so the other alternative would be to talk sometime on Fri afternoon, or over the weekend.

Please let me know what works for you. It might be best to catch up once you've heard back from Peter B, who we hope will shed some light on this.

Thanks,
Susan

On Tue, Dec 1, 2015 at 9:32 PM, E. Marden <[redacted] wrote:

Dear Susan,

I wanted to touch base before next week's meeting about the Report of the ASU Study (attached to the original distribution and included below). There was very little visibility into the development of this report - and I am a little concerned about what we are going to do with it. I am copying Peter Bretting on this email as he was



duplicate email trail removed

**Bretting, Peter**

**From:** [redacted] on behalf of Susan McCouch <srm4@cornell.edu>
**Sent:** Wednesday, December 02, 2015 10:32 AM
**To:** E. Marden
**Cc:** Bretting, Peter; Susan McCouch
**Subject:** Re: DivSeek meeting in Bonn - Comment on ASU Report

Dear Emily,

I just arrived in Zurich and am reading the docs sent by Daniele about the ASU governance study now. I would like to chat with you by Skype sometime in the next 2 hours, if at all possible. You are 9 hours behind us, so this is cutting it very short.

It's 4:23pm here now, and I will have to leave for dinner at 6:30pm. If we can't squeeze in a Skype conversation before I leave for dinner, perhaps we could talk when I return? I am not sure when that will be, but I expect I'll be back by 9:30 or 10:00pm.

I'm in meetings all day tomorrow, so the other alternative would be to talk sometime on Fri afternoon, or over the weekend.

Please let me know what works for you. It might be best to catch up once you've heard back from Peter B, who we hope will shed some light on this.

Thanks,
Susan

On Tue, Dec 1, 2015 at 9:32 PM, E. Marden [redacted] wrote:

> Dear Susan,
>
> I wanted to touch base before next week's meeting about the Report of the ASU Study (attached to the original distribution and included below). There was very little visibility into the development of this report - and I am a little concerned about what we are going to do with it. I am copying Peter Bretting on this email as he was also a member of the Governance Expert Committee. He and I have not discussed the ASU Report, and I welcome his comments as well.
>
> 1. As I understand it, the ASU study was co-funded jointly by Crop Trust and the Treaty to generate an independent report on governance as it relates to data sharing and open source in order to help guide DivSeek on this front. I have gone back through a series of emails on the report - and though my understanding is correct as the original intent of the ASU study, I can see where over the course of the year, this work stream seems to have been characterized more and more broadly. It now seems to cover everything, without disclosing the research or findings. Moreover, Daniele - of the Treaty - is now an author. So, two major issues: (a) we don't know what the report actually says; and (b) it is now arguably a Treaty perspective.
>
> 2. The report's authors propose to present the full findings at the January 8 "roundtable." I think this is unwise. As above, we don't know the content of the report and it is arguably a Treaty document. The roundtable in January will only help further DivSeek if it is a tightly orchestrated session presenting paths forward. I had hoped that next week's SC meeting would result in a recommendation on governance that



duplicate email trail removed