DRAFT FOR DISCUSSION

Generation Challenge Programme

Partnerships in modern crop breeding for food security

White paper on the GCP service component The Integrated Breeding Platform

Table of Contents

Acronyms and abbreviations used in the IBP papers

AfricaRice	Africa Rice Center
Agropolis–CIRAD	Centre de coopération internationale en recherche agronomique pour le développement, France
AP	IBP Analytical pipeline comprising of data analysis and decision-support tools
ARI	advanced research institute
BGI	Beijing Genomics Institute, China
BIOTEC	National Center for Genetic Engineering and Biotechnology, Thailand
BSS&H	Breeding and Support Services & Helpdesks (of IBP)
CAAS	Chinese Academy of Agricultural Sciences, China
CAPs	Coordinated Agricultural Projects, USA (of USDA)
CB	capacity building
CD	compact disc
CGIAR	CGIAR Consortium of International Agricultural Research Centers
CI	see RI
CIAT	Centro Internacional de Agricultura Tropical (*International Center for Tropical Agriculture*)
CIMMYT	Centro Internacional de Mejoramiento de Maíz y Trigo (*International Maize and Wheat Improvement Center*)
CIRAD	see Agropolis–CIRAD
CIS	crop information system
CoPs	communities of practice
CORAF/WECARD	Conseil ouest et centre africain pour la recherche et le developpement agricoles/ *West and Central African Council for Agricultural Research and Development*
CRPs	CGIAR Research Programmes
CWS	configurable workflow system
FAO	Food and Agriculture Organization of the United Nations
FARA	Forum for Agricultural Research in Africa
GBS	genotyping by sequencing
GCP	Generation Challenge Programme (of the CGIAR)
GDMS	Genotyping Data Management System (of GCP)
GENESYS	platform or “gateway” to plant genetic resources
GenStat	A statistical data analysis software for biological and life sciences, developed and supported by VSN International – a United Kingdom firm.
GIPB	Global Partnership Initiative for Plant Breeding Capacity Building
IARI–ICAR	Indian Agricultural Research Institute, India (of ICAR)
IBP	Integrated Breeding Platform (of GCP)

ICAR	Indian Council of Agricultural Research, India
IChIS	International Chickpeas Information System
ICASS	International Cassava Information System
ICIS	International Crop Information System (of CGIAR)
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
ICS– CAAS	Institute of Crop Sciences, China (of CAAS)
ICT	information and communication technology
IGnIS	International Groundnuts Information System
IITA	International Institute of Tropical Agriculture
IMIS	International Maize Information System
IP	intellectual property
IPhIS	International Phaseolus Information System
IRIS	International Rice Information System (of ICIS)
IRRI	International Rice Research Institute
IS	information system
ISgIS	International Sorghum Information System
IWIS	International Wheat Information System
IVIS	International Vigna Information System
MAB	marker-assisted breeding
MABC	marker-assisted backcrossing
MARS	marker-assisted recurrent selection
MAS	marker-assisted selection
MTAs	marker–trait associations, *also* diagnostic markers
NARS	national agricultural research system
NRCRI	National Root Crops Research Institute, Nigeria
NSF	The United States National Science Foundation
OptiMAS	Breeding decision-support tool for MARS
PABRA	Pan-African Bean Research Alliance
PBI	Plant Breeding Institute, Australia (of University of Sydney)
QTLs	quantitative trait loci
R4D	research for development
RI	Research Initiative (of GCP), *formerly* Challenge Initiative (CI)
SMEs	small and medium-sized private enterprises
SGRP	System-wide Genetic Resources Programme (of CGIAR)
SNP	single-nucleotide polymorphism
UCR	University of California–Riverside, USA

USD	United States dollar (USA currency)
USDA	United States Department of Agriculture, USA
VSNi	VSN International, a crop informatics firm based in the UK that are the proprietors of GenStat
WUE	water-use efficiency

Background and process

A series of white papers are being drafted by the Generation Challenge Programme (GCP) team in collaboration with external experts. The purpose is to communicate the outputs and deliverables from each research and service component during 2004–2014, and to explore options for enabling and ensuring that the potential benefits of these components will be fully used in the future. At this stage, the white papers are just an initial analysis for purposes of internal use. They are expected to develop over time, shaped by progress made during GCP's remaining time and by the evolution of international agricultural research for development (R4D), particularly in terms of the 'moving landscape' of socio-economic, political and environmental issues in which the research portfolios of the CGIAR Consortium of International Agricultural Research Centres (CGIAR) and related CGIAR Research Programmes (CRPs) operate. Each white paper is designed to contribute to GCP's orderly closure in 2014 by considering the following three questions:

1. What research assets will be completed by the end of the GCP's lifetime in December 2014?
2. What research assets can best continue as integral components of the new CGIAR Research Programmes (CRPs) or elsewhere?
3. What research assets may not fit within existing institutions or programmes and may require alternative implementation mechanisms?

This paper focuses on the outputs and options for GCP's Integrated Breeding Platform (IBP). Because the development of this component is a major part of the research portfolio of GCP Phase II, the reader will find, embedded as annexes within this overall paper, five shorter papers. These deal with the Platform's five key elements:

- Annex 1 Breeding information and Communities of Practice
- Annex 2 Data management software
- Annex 3 Analysis and decision support software
- Annex 4 Data management service
- Annex 5 Breeding and support services

An Annex 6 Genomics tools and resources will be incorporated in due course

Outputs have been achieved through (a) collaborative work among three sets of actors: a broad network of partners in regional and country research programmes, the CGIAR and academia; and (b) through capacity enhancement to assist developing-country researchers to tap into new genetic diversity and access modern breeding tools and services. GCP research activities[1] have produced the research products described below.

[1] GCP is supported by the generous funding from an array of donor organisations listed at http://www.generationcp.org/network/funders. See also descriptions of products at http://www.generationcp.org/impact/product-catalogue and of the institutions that generated them at http://www.generationcp.org/research/research-projects.

Introduction and rationale

Growth in scientific knowledge and innovation and advances in information and communication technology (ICT) over the past two decades have jointly provided new tools, avenues and resources to address food security. The global agricultural research agenda must take advantage of this progress to enhance the agricultural knowledge base and provide innovative solutions.

Biological sciences are now extremely 'data-rich'. This quantitative nature of modern biology demands closer collaboration between biologists and informaticians, as well as strong partnerships between researchers on both developing and developed countries, across sectors and disciplines. The concurrent revolutions in genomics, molecular biology and information technology offer unprecedented opportunities to enhance breeding programmes. Consequently, molecular characterisation, accurate phenotyping, analytical tools and overarching information systems must be integrated with breeding workflows combining pedigree, phenotypic, genotypic and adaptation data for better predictions on the performance of different genotypes across various environments.

Integrated breeding hastens genetic gain by combining phenotypic selection with cost-effective, precise and faster molecular breeding methods. However, although molecular breeding approaches have been readily adopted by and proved to be of great benefit to the private sector, they have had limited impact in the public sector and in small private enterprises. The reasons for this vary, but include: lack of personnel, inadequate high-throughput genotyping capacity, unreliable phenotyping practices and protocols, inadequate infrastructure, poor information management systems backed by inadequate analytical tools, and generally insufficient resources. These have slowed development of new cultivars and compromised food security.

The development of the Integrated Breeding Platform[2], led by GCP in collaboration with a broad set of partners from CGIAR Centres, Universities and National Programmes, is intended to help overcome these bottlenecks. This 'one-stop shop' provides information, services and training, and supports scientific communities of practice focusing on integrated plant breeding (ie, traditional breeding methods enhanced by molecular screening and advanced analysis and decision-support).

The IBP's overall objective is to provide developing countries with access to modern breeding technologies, breeding materials and related information in a centralised and practical manner, thus facilitating their adoption of molecular breeding approaches and improving their plant breeding efficiency.

The short-term objective is to establish – based on a client-centred approach – a minimum set of tools, data management infrastructure and services to address the needs and enhance the efficiency of selected breeding projects designated as "user cases". Through these user cases, marker-based approaches would be used to improve seven crops in sub-Saharan Africa and South & Southeast Asia.

---

[2] See website at http://www.integratedbreeding.net

The IBP is expected to be operational as an integrated analytical pipeline by end of 2012, at which time it will be opened to a broader public beyond the user cases, both within and outside the CGIAR system. The platform will enable breeding programmes in the public and private sectors to accelerate variety development for developing countries, using marker technologies ranging from simple gene or transgene introgression to gene pyramiding for complex traits.

To achieve these objectives, GCP allocated about USD 22 million to the IBP initiative over five years (July 2009 to 2014), with financial support primarily from the Bill & Melinda Gates Foundation and also from the European Commission (EC) and the UK Department for International Development (DFID). This effort represents about 15% of GCP's total budget (USD 150 million over 10 years).

The establishment of the Platform in 2009, at the beginning of GCP's Phase II, was timely and highly appropriate, as it coincided with the implementation of several projects on molecular breeding within and outside GCP. The development of the platform is driven and informed by the 14 user cases, which are implemented by a broad set of partners and cover different crops. The details of the user cases are provided in Annex 1 *Breeding information and user communities of practice*. The user cases are defining user requirements and, hence, the design and development prioritisation of the Platform's different elements. The Platform's reciprocal contribution to these breeding projects is to help them bypass bottlenecks that would compromise product delivery, and enhance their overall efficiency by providing appropriate tools and support.

At programme level, the research components aim to demonstrate – through the user cases – that modern and integrated breeding approaches can have a significant positive impact on crop productivity in developing countries. The service component, fundamentally the Integrated Breeding Platform, is conceived as a vehicle for the dissemination of knowledge and technology. It not only permits broad access, but also the proactive distribution of crop genetic stocks and breeding materials; molecular, genomics and informatics technology and information; cost-effective high-throughput laboratory services; and capacity-building programmes.

Activities and products of the Integrated Breeding Platform

The IBP's primary stakeholders are plant scientists, specifically breeders, who lead the selected molecular breeding projects of the 14 pioneer user cases. The Platform particularly aims to enable breeders in developing countries to access and use modern breeding approaches. The social component of the Platform is therefore key to achieving this objective, and to ensuring some sustainability for the mid- to long term.

From a technical point of view, the IBP has three broad components (Figure 1): a web-based portal and helpdesks; an open-source information system (IS) that incorporates an adaptable integrated breeding Configurable Workflow System; and breeding and support services.

The IBP's developmental phase brings together highly regarded public research teams, which are made up of institutes and individuals who work on the challenges of crop information management and

analysis, biometrics and quantitative genetics. The resulting team of bioinformaticians, statisticians and developers aims to design and develop different elements of the Platform, based on needs and priorities, as defined by the user cases. To maximise efficiency, and given that coding is not a strength of GCP's CGIAR partners, several components of the Configurable Workflow System (CWS) have been subcontracted to private-sector service providers for development.

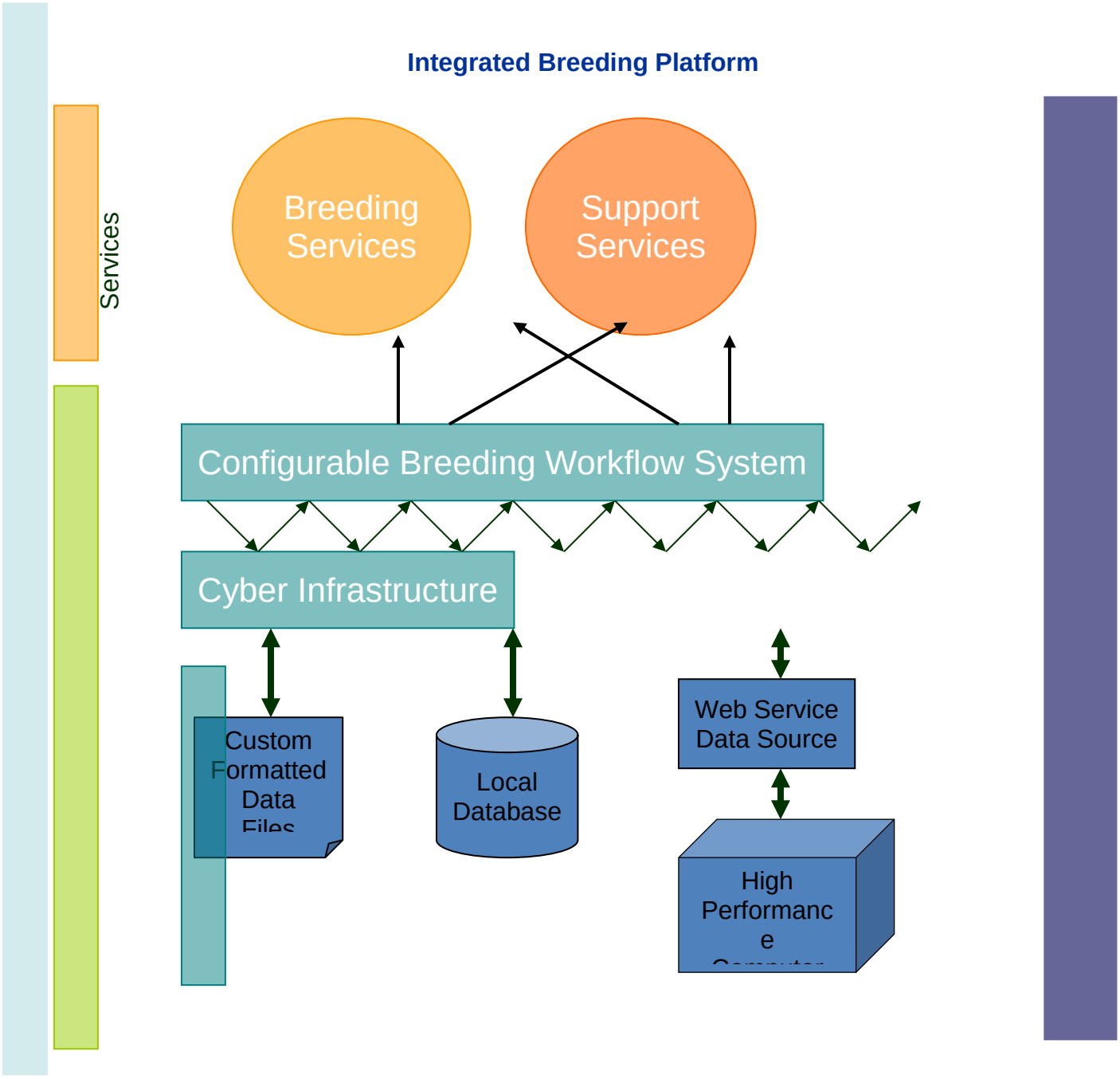


Figure 1. The Integrated Breeding Platform and its three main components

A continuous dialogue between users, developers and service providers ensures a healthy balance

between having a needs-oriented user-driven platform on the one hand, with a reasonable degree of 'technology push' on the other hand to ensure that users keep abreast of technological solutions they may not be aware of but which would facilitate and accelerate breeding work.

By the end of 2012, the IBP is proposed to be accessible to all plant scientists, particularly breeders in both the public and private sectors. By providing forums and other community tools via a user-friendly portal, it will stimulate the development of crop- and discipline-based communities of practice (CoPs). The CoPs will promote the application of molecular breeding techniques and use of facilitative information management technologies, enhance data and germplasm sharing, and generally advance modern breeding capacity by linking CGIAR Centres and advanced research institutes with developing-country breeding programmes and research organisations. They will also facilitate and accelerate a paradigm shift to a more collaborative, outward-looking, technology-enhanced approach to breeding.

The CoPs are expected to become the Platform's primary stakeholders by December 2014. Their needs and expectations will therefore need to be taken seriously into account when defining the IBP's future.

The integrated breeding portal and helpdesk

Inaugurated in October 2011, but still not broadly advertised, the portal is the online gateway through which users access all the tools and services found on the Integrated Breeding Platform. Through the portal, users will select and download tools and user instructions, order materials and procure laboratory services. The portal's helpdesks facilitate its use and ensure access for users who cannot efficiently use the web interface – by providing the information, tools and services they need through email, compact discs and other offline media. The portal's social networking component provides forums, blogs, event calendars and other community tools.

The information system

The value of an Information System resides in both the quality of the individual tools or modules that are part of it, and in the cyber infrastructure or middleware that ensures both cohesion across tools, and efficient communication with databases.

The IBP information system is structured as a Configurable Workflow System (CWS), with access to both local databases and distributed resources, such as central crop databases and molecular databases on GCP partners' sites and public initiatives such as Gramene and GrainGenes. This CWS is the operational representation of the information system to be implemented. It assembles informatics tools into applications configured to match specific integrated breeding workflows (eg, for MAS, MABC or MARS) while also supporting the informatics needs of traditional breeding. The tools are developed in a series of functional modules that comprise the Integrated Breeding Workbench – the background structure that implements the CWS. The Information System of the IBP is therefore a complete system that includes the following modules:

- Project planning module
- Germplasm management module
- Robust relational database

- Data collection and cleaning tools
- Analytical standards
- Analytical and decision support tools
- Query tools
- A cyber infrastructure linking the multiple tools in a cohesive and user-friendly way

These elements can be grouped into two components: information management (see Annex 2 *Data management software*) and information analysis (Annex 3 *Analysis and Decision-support software*).

IBP services

The Services component comprises three modules: the first supports data management (Annex 4 *Data management service*), the second all analytical steps involved in breeding projects (Annex 5 *Breeding and Support Services*,) and the third deals with the intellectual property (IP) and policy.

The *Data management service* includes, among other elements, support and training on data generation, access, curation and documentation. All these elements are critical for any successful experiment but, unfortunately, are too often undervalued and underestimated in terms of resource allocation. Additional effort is also needed to properly document data so that it can be shared and used at a global level. Here again, researchers generally do not accurately anticipate the effort needed. This service therefore aims to educate breeders on and support them in managing their information. The service particular aspires to communicate to breeders that proper data management is a critical component of their work and the IBP – as without good data all other elements of their work and the Platform are greatly diminished in value.

The *Breeding and support services* include Breeding Plan Development. Developing a workplan with a cost/benefit analysis is essential before conducting a multi-cycle molecular breeding project. It also includes a technical element to assist users in accessing a set of online options for high-throughput marker service laboratories in the public and private sectors which operate under clear concessionary contractual terms. These service laboratories are selected on the basis of competitive costs, compliance with defined quality requirements and expeditious delivery of services. The *Design and analysis* element of the support services is core. It provides support on statistics, bioinformatics, quantitative genetics and molecular biology. Through the *Phenotyping sites and screening protocols* element, users can access information on phenotyping sites, protocols and potential collaborators to ensure that selection is carried out under appropriate biotic and abiotic stresses, and that germplasm adaptation is well characterised.

The *IP and policy* module provides support on intellectual property rights and the freedom to operate in the arena of biotechnology and germplasm use.

Measuring success and impact

The Platform is primarily about access to information, tools and services. The best means of measuring success is therefore to look at adoption rates. The opening of the IBP to a wider community of plant

scientists and breeders at the end of 2012 will be a cornerstone in the IBP's development. The portal has facilities for tracing and retrieving information on who has been consulting and downloading what. The first evaluation of demand rates and user composition should take place by the third quarter of 2013. Box 1 summarises the different impact indicators that may be considered for the short term (1–3 years after opening up to a wider public), mid-term (3 years after opening up to 2020) and long term (after 2020). Quantifiable indicators for the long term will need to be defined, according to economic studies to be conducted by 2015.

Of course, the cost that might be requested for the right to access some IBP tools and services will have an impact on demand and rate of use, and therefore on the different impact indicators listed below.

Box 1. Impact indicators

Short-term (2013–2015)

1. Number of requests for the different products available through the service (eg, databases, molecular markers, analytical tools, protocols and training materials): 150 expected in 2013; 300 in 2014
2. Number of support requests received from country programmes and SMEs (eg, planning for breeding projects, informatics, data management, data analysis, genotyping, training and IP): for 2013–2014, 200 per year
3. Number of citations of IBP per article (both reviewed and general) per year: starting 2014, at least 100
4. Number of first-time and returning users for the different high-throughput and breeding databases, tools and services: for 2013–2014, 150 per year
5. Number of datasets published through crop Lead Centre databases: 50 in 2013; 80 in 2014
6. Number of researchers that join established CoPs: 150 per year

Mid-term (2016–2020)

7. Number of operational molecular breeding programmes in both public and private sectors in developing countries that routinely use the IBP: at least 300 in 2015
8. Number and membership dynamics of active crop CoPs: at least 50 members per CoP
9. Number of improved crop varieties released or developed, using IBP tools and services: 50 by 2015; 80 by 2016; more than 100 per year, starting 2017

Long-term (after 2020, impact studies)

10. Acreage under crop varieties produced through molecular breeding
11. Yield increases in a given ecoregion and time period that are attributable to crop varieties from molecular breeding programmes
12. Number of farm households with enhanced livelihoods that are attributable to increased crop productivity because of improved varieties from molecular breeding programmes

Reviewing the dynamics of demand for support services will be another indirect means of measuring success and impact. In theory, demand for the support services should increase over the first few years following the opening up the Platform with increases in the number of new users. It would then stabilise as the tools and breeding services are improved and enhanced based on user feedback, hence lessening the need for support and help as provided by the support services. Additionally, if the CoPs indeed become active forums for effective peer-to-peer support and discussions, questions and issues raised by users may well be addressed directly by their peers in the professional networks.

Creating and supporting active networks of scientists sharing similar professional interests and aspirations is one of the objectives of the IBP. Such networks can be organised by crop, research theme or profession – creating fora for collaborative access to capacity building, to tools and services, phenotyping and other protocols, and collective troubleshooting. The primary drivers of CoP membership are intended to be sharing of expertise and constructive dialogue around molecular breeding, rather than accessing funds. CoP dynamics and the number of members over time will thus also be a highly reliable indicator of the Platform's success.

However, realising this objective of community-building is a significant challenge, given that few success stories have been reported. Most such networks fall apart when establishment financial resources end or are insufficient. Over the last year, GCP has tried to form CoPs, building on the scientific groups involved in GCP's crop Research Initiatives (RIs), but with highly variable success across the RIs. Successful leaders, who are the key, tend to emerge naturally because they have the right attitude, drive and charisma to motivate members to play active roles in a social network. If leaders do not have these characteristics 'in their blood', then their leadership simply does not work, even if they are the lead scientists in their field. GCP is now exploring a bottom-up approach in the hope that leaders and other community enthusiasts will emerge from outside the GCP network once the Platform is opened up to the broader public at the end of 2012. Participation will be stimulated through incentives directly linked to participation in community activities – such as innovative blog posts, frequent participation in Q&A forums, provision of feedback on tools and services.

GCP's intention is that the CoPs should be the IBP's key stakeholders and, ideally, representatives of these communities should in the future contribute actively to IBP's management.

On a relative scale of 0 to 5, where 5 represents the largest impact across all kinds of GCP products, and 0 no impact, all GCP's efforts to develop and deploy the IBP are estimated to have an impact factor of 5. Such a high score indicates that the IBP and related activities have the potential to make a huge impact on plant breeding efficiency in developing countries.

Post-GCP sustainability of the Integrated Breeding Platform

This section deals with IBP's sustainability in general terms, considering the Platform as described in our proposal. However, such an exercise is not empirical because the Platform may evolve in a different direction, depending on the magnitude and nature of the capacity-building, social-networking and even research components that may be included as part of it post-GCP. This section will consider some possible scenarios to more accurately predict the shape and nature of the IBP after December 2014.

The IBP's development is secured until June 2014. After that date, because of the commitment made by the Bill & Melinda Gates Foundation in Phase I, funding is still highly likely to be available to support the Platform's further development and ensuring the continuation of some of its support services.

From a business plan perspective three options may be considered for the post-GCP era: (1) That the Platform will be fully supported by donor contributions, (2) That the Platform will be supported partly

by donors and partly by income generated from cost to access tools and services, and (3) That the Platform will be fully self-sustaining.

Considering IBP's public goods and humanitarian components, and its orientation towards providing scientists in developing countries with access to molecular breeding at reduced costs, Option 3 (fully self-sustaining) appears unrealistic for the short- to medium-term, although not for the long-term (after 2018). Option 1 (fully donor-funded) is also not very desirable or feasible, given the uncertain nature of donor funding and because it would undermine incentive to develop quality products that can win a market niche on their own merit. As a principle, GCP also believes that "free is not good". That products must be paid for puts pressure on the development teams to generate a saleable product that is of sufficiently high quality to ensure that users will not only learn how to use it, but will indeed use it. This leaves Option 2 as the most suitable option for the coming years – partially funded by donors and partly by internally generated funds.

Two models may be considered for generating income: (1) where users/institutions directly pay for tools and services; and (2) where users/institutions pay a membership fee to access the Platform's facilities. The chargeable fees could be defined according to a market-and-demand analysis and also the level of donor contributions received at any given time. The membership option could offer different packages at different fees, and in both cases fees would be adjusted to different kinds of users. However, IBP even in the post-GCP era should still enable scientists in developing countries (excluding emerging economies) to access the Platform for free or at nominal cost.

Under its current structure and functionality, the basic maintenance cost of the Platform is estimated at about USD 2 million per year. The level of adoption and the nature of users will define the income that could potentially be generated internally by the Platform. If that income is high, charge-backs on maintenance and support services may indeed bear the entire cost of developing and maintaining the Platform.

The magnitude of buy-in from small and medium-sized private enterprises (SMEs) will determine the level of income that the Platform can generate. The potential is large, as hundreds of SMEs are involved in the seed business, especially in Asia, Central and South America and to a less extent Africa. These potential IBP users are seeking reliable data management systems and analytical pipelines to start breeding activities over and above their current seed multiplication business or to scale up their breeding activities. Of course, the SMEs will adopt and invest in the IBP only if the system guarantees sustainability and if the support services are well established and of good quality.

In December 2012, the portal will be opened up to all kinds of users. Right from the outset, tools will be accessible under licensing agreements but, for the first year, no fee will be charged as the entire system will still be under development and implementation, and feedback from all users will be requested to improve the Platform. In December 2013, a first survey will be conducted to ascertain demand and evaluate the quality of the tools and services. Results of that survey will be used to decide if the Platform could generate income in the following year, starting December 2013 or in 2014 at GCP's sunset.

Post-GCP placement of the Integrated Breeding Platform

Five components of the IBP have been identified that relate to knowledge sharing and users, databases, data management services, tools and support services. For each component, an analysis was conducted as objectively as possible on what may happen after December 2014. This analysis is contained in the Annexes 1 to 5 attached to this document. Indeed, by the end of the initial five-year project, the different IBP components can be separated and positioned in different existing programmes or initiatives. The analyses are presented in the individual papers (Annexes 1 to 5) and demonstrate that the option of dismantling the IBP and placing its various components separately or on a crop basis has not been excluded.

Dismantling the IBP will clearly and dramatically reduce its overall value and potential impact. What makes the Platform unique, adding considerable value to each of its individual components, is to have them together, accessible in an integrated and configurable way through a single web portal. The next section, therefore, considers the IBP's future only as a single and integrated product.

What will be finished by December 2014

The initial five-year phase of the project will end in July 2014. By that time, the Platform is expected to be fully functioning, as projected in the initial proposal. The first three years (from July 2009 to 2012) focused on developing the different tools. At the annual meeting in June 2011, all stand-alone tools were presented and, in the following year, the first version of the integrated Configurable Workflow System was introduced to users for testing. The remaining two years of the project will see the implementation of the different tools, taking into account feedback from users within and outside GCP, and the development of the different support services. By December 2012, the CWS will be fully operational and the IBP will be available to scientists and breeders in the public and private sectors. By December 2014, the fully operational IBP will be two years old and a thorough evaluation of the Platform's adoption shall be conducted at that time. Put simply: by GCP's sunset, the project will be fully finished, as projected in the proposal. However, it will, of course, have only just begun its life as a public platform.

Extending activities to CRPs, Centres or other institutions

The current nature of the crop CRPs suggests that research, capacity building and networking will develop as integrated parts of those programmes. Thus, if the IBP remains under the CGIAR umbrella, it will operate in a relative austere format, essentially focusing on data management, analysis and decision-support functions. The IBP would serve the needs of the CRPs and partners and will be a demand-driven support platform that provides best practice and access to proper data management, breeding tools and services. Although the IBP would also be a natural home for knowledge sharing and access to eLearning materials and protocols, the core of CB activities would, in fact, be embedded in the CRPs and related research activities.

It is important to underline upfront that the CGIAR centres, independent of the evolution of the shape and content of the IBP, would need to play an important role in the hosting and maintenance of the various databases for their mandate crops (both phenotypic and genetic) associated with the IBP.

Without that critical component, the IBP CWS will just hook into a bunch of independent local databases at each user organization, with no sharing of data or access to up-to-date global information. This would not be ideal, as data access and exchange at a global level is one of the key objectives of the IBP.

Preliminary discussions with representatives of the CGIAR Consortium Office suggest that a clear interest exists to integrate the IBP as a cross-cutting platform to support the CRPs along the lines described above. Support for developing some research elements may not be forthcoming if they directly compete with those of the CRPs. As for for the social networking component the CGIAR Consortium Office has expressed strong commitment to allocating additional resources to bring new partners on board and develop solid networking opportunities and systems within the CRPs.

Under the scenario of an IBP within the CGIAR system, the integration of the genetic resources and stocks components should also be considered. The Global Crop Diversity Trust is leading the development of a platform (GENESYS) to manage the genetic resources of the different CGIAR genebanks, providing access to accessions and related passport information. A natural and desirable move would be to link the two platforms, whereby the IBP provides access to databases and to related analytical tools for managing and analysing phenotypic and genotypic data collected on those resources, thus promoting access to genetic stocks. Such efforts would facilitate the discovery of new alleles for breeding and the development of new genetic stocks for pre-breeding activities.

The IBP could also be embedded in an institution outside the CGIAR. However, no obvious option comes to mind that would facilitate a proactive collaboration. The web-based elements of the Platform are currently hosted by the iPlant Collaborative, a community of researchers, educators, and students working to enrich all plant sciences through the development of cyber-infrastructure – the physical computing resources, virtual machine resources, collaborative environment, and interoperable analysis software and data services that are essential components of modern biology. However, a post-GCP alliance with iPlant will represent only a partial solution. The iPlant cyber infrastructure is suitable for hosting the tools and databases of the IBP, and indeed this is already part of the plan – users will be able run the CWS online, taking advantage of iPlant's considerable computing power to run heavy analysis. But the iPlant environment is not designed to provide maintenance and support services for the tools hosted on its cyber-infrastructure, nor to support the CB and networking components of consolidated IBP.

Many options are probably open for the IBP to be hosted by a legal institute, and some have already been identified (see next section). However, considering that the IBP supports an ongoing large initiative in breeding, none of the other current efforts in the public sector will have the magnitude nor the diversity of research areas, in both people and crops, needed to take full advantage of the IBP's potential. In that context, the CGIAR is in an unrivalled position.

Another option is for the IBP to be converted into a commercial entity, as a start-up enterprise or linked with an existing commercial analytical pipeline. This second option would ensure that, if the demand from users is sufficient, the tools developed during the first 5 years of the IBP project will be maintained, improved and supported in a professional way by a commercial enterprise that is already

well established in the arena of plant science and analysis. The firm VSN International (VSNi) is a possible option for a couple of reasons. They have successfully commercialised GenStat, a statistical data analysis software for biological and life sciences. GenStat is marketed worldwide on a model that resonates with the strategic objectives of GCP and IBP – specifically, promoting and facilitating access to modern breeding technologies by developing countries. A basic version of the software, GenStat Discovery, is available free of charge to breeding programmes in developing countries, with requisite support services. GenStat Discovery does not include the most advanced versions of the different tools, but allows users to run basic analyses. VSNi is already involved in the development of elements of the Analytical Pipeline of the IBP Configurable Workflow System. The main drawback of this option is that the IBP and related products will not be in the public domain anymore as global public goods. The commercial priorities pursued by VSNi, or any similar partner, in the future might also not always align with the needs and demands of the IBP stakeholders.

Embedding the IBP in a new entity after December 2014

The IBP's core activities are, without doubt, geared to provide a data management and analysis system, and tools and services that will support breeders in their routine activities. However, those core functions can be extended to convert the IBP into a livelier and more comprehensive platform that would better serve a broad and diverse community of users, focusing on scientists in developing countries. It could include research elements that would provide exposure to modern breeding through small grants (e.g. 5 to 10 thousand USD, as is currently done through the GCP Genotyping Support Service concept). Small grants might also be allocated to support fingerprinting as new advanced lines get developed by the various users. This is common practice in the private sector, to characterise new lines and identify their potential as parental lines in new crosses. Similar exercises should also be conducted automatically as good practice by CG Centre breeding programmes, but that might not be possible, due to resource challenges, for lines developed by breeders in developing countries, reducing the potential impact of new varieties in those countries. A research component could also include specific and targeted CB activities that would involve research activities to enable scientists from developing countries to participate more in the arena of modern breeding. Support in implementing local infrastructure could also be considered.

Under such a scenario, the level of CB can also be magnified beyond a strict support in the use of the Platform's tools and services. It could include courses, support material and forum discussions on cross-cutting themes such as theoretical aspects of molecular breeding, exposure to new genomics tools, phenotyping screening protocols, IP, and policy issues. Of course, adding a research component and expanding on CB, compared with what would be offered under the first scenario above, would provide favourable conditions for developing and expanding a proactive and diverse network of partners, including CoPs.

Although, at a first glance, some of these activities may be seen to compete with what can be achieved in the different CRPs, this would not be the case because the entry point will be different. That is, CRP activities are organised around research objectives, while IBP activities would be organised in a crosscutting way, following a thematic approach, thus enabling developing-country scientists to be exposed to, access and use modern-breeding approaches.

The creation of a new entity for incorporating the IBP appears to be the most suitable option, considering the IBP's form and functionality as described in this section, and the flexibility and freedom it needs to adjust and respond to the demands of a broad range of users and stakeholders, while focusing on scientists in developing countries.

If, at the end of the day, this second scenario is selected, the management and governance of such an entity can then be defined. GCP's experience suggests that an external scientific monitoring and advisory team would need to be established. Considering that the fundamental objective is to serve a broad set of users, the establishment of a consortium to oversee IBP's evolution appears critical to the success of that scenario. Consortium members would include representatives from developing countries, CGIAR Centres, universities and SMEs.

The choice between an independent legal institution and a host agent arrangement would need to be determined in due course. In preliminary discussions, several institutions have already expressed interest in potentially hosting the IBP, provided that the non-legal status is retained. Such institutions include the Global Crop Diversity Trust, the Syngenta Foundation for Sustainable Agriculture and regional organisations such as CORAF/WECARD and FARA. Considering the investment and direct interest of the Bill & Melinda Gates Foundation in the IBP, they could possibly host the Platform.

Discussions advanced furthest with Chinese representatives during the annual meeting held in Beijing, June 2012. CAAS is establishing, in collaboration with the Beijing Genomics Institute (BGI) and some CGIAR Centres, a new Biological Breeding Innovation Research Institute in Shenzhen. As part of this effort, and after discussions with Wang Ren, the vice-president of CAAS and president of Shenzhen Biological Breeding Innovation Research Institute, CAAS is willing to consider the possibility of physically hosting the IBP on its new campus. By early 2013, the campus is expected to host more than 2,000 scientists; provide high-throughput sequencing and genotyping facilities (together with biological experts to facilitate data analysis and interpretation); and modern CB facilities.

Conclusion

Embedding the IBP into an existing, large, breeding initiative such as the CGIAR (Scenario One), has some pros and cons. It implies, among other things, that the Platform will focus on a support and service role with limited flexibility to conduct, on its own initiative, research, CB or networking activities because these activities will be led by and embedded in the initiative's own research projects. Of course, the division of responsibilities does not have to be black and white, and a middle ground can surely be found, especially for the CB and networking elements. Overall, however, not much freedom can be expected for IBP management to extend beyond core support and service functions.

One advantage of Scenario One is that the IBP will have a ready-made and large community of users with clear demand stemming from numerous ongoing breeding projects. The transition after December 2014 would therefore be made smoother by the existence of a core group of potential members and users. In addition, most of the CGIAR Centres now leading the different CRPs participated in the IBP's development in one way or another, thereby creating some sense of

ownership. They are also hosting the central databases for their particular mandate crops. Last, but not least, the IBP builds on the best practices of the different Centres at different levels (data management, tools and CB, among others). A clear added value for using the IBP should therefore exist for the Centres and partners, as the Platform was, in a sense, designed for them in the first place.

Given the above discussion, the key questions are now: (1) does the IBP need to be embedded in the CGIAR system to best serve the CRPs? and (2) will the CRPs suitably address the needs and expectations of GCP stakeholders in developing countries in terms of their needs for CB and networking? Depending on the answers to these questions, Scenario Two could be considered.

A new entity, with a governance to be defined, will provide flexibility to the IBP stakeholders to build and extend the Platform's core support and service components. Such extension may include research activities that introduce scientists in developing countries to modern breeding approaches and encourage them to adopt those approaches. Capacity-building activities can be organised according to theme and directly related to the facilities and services offered by the Platform, thereby building and nurturing the development of active research networks and CoPs.

Such a scenario will need to build on strong collaboration with the CGIAR, as several of the Platform's components, including the crop databases, rely on some of the CGIAR Centres' assets. Therefore, the advantage of interacting, from day one, with an active community of users from the crop CRPs, as described under Scenario One, should not be lost. But this second scenario will allow the IBP to have some independence in organising activities related to modern breeding.

The model of an independent entity creates some autonomy of action for a platform to be led by stakeholders within and outside the CGIAR, with strong participation and leadership from scientists from developing countries and including mentorship from experts in CGIAR Centres and universities. This model, of course, echoes GCP's current way of operation and, in fact, builds on the GCP partnership. Creating a new consortium with a mix of partners from different horizons, including the private sector, may be an attractive option, as it has already demonstrated its potential. Such a scenario also received highly supportive feedback from key country partners such as CAAS in China, ICAR in India and BIOTEC in Thailand.

IBP's future therefore remains an open question. Two realistic scenarios have been laid out in this paper, with the possibility of considering other alternatives to these scenarios. The two questions above must be answered to identify the best way forward. We hope to have them addressed by our stakeholders and partners during a workshop to be held in early 2013. The composition of the IBP user community after its broad opening in December 2012, comparing users and potential users within the CGIAR system with those outside the system, may also be a significant element to take into account when considering the best model to adopt.

Annex 1 Breeding Information and Communities of Practice

Introduction and rationale

A fundamental function of GCP's Integrated Breeding Platform (IBP) is to provide useful, timely and relevant breeding information to target client groups. The goal is to accelerate adoption and facilitate application of molecular breeding technologies by programmes in developing countries in particular. In serving this function, the Platform, through the IBP Web Portal, provides access to information pertinent to molecular breeding. Such information relates not only to IBP's core activities and the different tools and services offered through the Platform, but also to announcements (eg, training courses, meetings, protocols and e-learning materials) and news (eg, germplasm release, papers and meeting proceedings) from plant science initiatives both within and outside the GCP collaboration networks.

A significant amount of this information is collated from GCP-supported research work, and from collaborators. It is also generated from and through crop communities of practice (CoPs) and the user cases of the IBP project. The crop CoPs are groups of crop researchers willing to share experiences and information on modern breeding methods, best field practices and improved varieties, and also to provide peer-to-peer support through the Platform's community tools. The user cases, in contrast, comprise 14 marker-assisted breeding projects for eight crops in 32 developing countries in Africa and Asia. These pioneer users of the Platform collaborate in designing and testing IBP's services and informatics tools. In GCP Phase I (2009–2012), IBP's development is driven by the needs of the user cases.

These communities and user cases are intended as the vanguard for both the dissemination of this information, and its further development and accumulation. The CoPs are of particular significance because they are expected to be core stakeholders of the IBP in the post-GCP era.

Activities and products

At the time of preparing this paper, the IBP is providing access to the following information:

- Nine crop databases that are maintained and curated by the respective Crop Lead Centres for beans, cassava, chickpeas, cowpeas, groundnuts, maize, rice, sorghum and wheat. In addition to a downloadable database for each of these crops, the Platform provides an online query interface for each. More details are provided in Annex 4 *Data Management Service*.
- Information on various informatics tools for data management, and analysis and decision support, details of which are covered in Annex 2 *Data Management Methodology and Tools* and Annex 3 *Analysis and Decision Support Methodology and Software*

- *GCP Crop Ontology*, an online tool that was developed as part of the GCP crop ontology project. This tool allows community participation and curation, and is therefore constantly expanding with new and updated information
- Downloadable *Trait dictionaries* for the nine crops define nomenclature and measurement protocols, emphasising the 50 most commonly measured breeding traits. These traits have also been incorporated into the Integrated Breeding FieldBook

- Information on genetic resources, through links to the CGIAR's System-wide Genetic Resources Programme (SGRP) and GENESYS of the Global Crop Diversity Trust
- Information on genomic resources, particularly marker information, captured in the Molecular Marker Toolkit and the Genotyping Data Management System (GDMS) – both of GCP – and by outside repositories.
- Information on Breeding Services and service providers and on Support Services. Details are covered in Annex 5 *Breeding and Support Services*
- Information on capacity-building resources and opportunities, both internal and third party. Details are covered in the Paper on *Capacity-building and training courses*
- A growing database of relevant professional publications
- Information generated by the intra-community interactions on blogs and forums on the IBP Portal, including insights and experiences shared by community members
- Relevant news and events, for example, meetings, training courses, germplasm and variety releases, recent discoveries, funding and fellowship opportunities, learning and other resources

The CoPs are intended to promote the application of molecular breeding techniques and use of facilitative information management technologies, enhance data and germplasm sharing and generally advance modern breeding capacity by linking CGIAR Centres and advanced research institutes with developing-country breeding programmes and research organisations. The CoPs should also facilitate and accelerate a paradigm shift to a more collaborative approach to breeding that is outward-looking and technology-enhanced.

The researchers, working in the IBP-use cases, form the CoPs' nuclei. They are designated as user testers for IBP tools and services, providing feedback from iterative development and testing cycles. Thus, they ensure that the Platform's tools and services are demand driven, addressing the breeder's practical needs. Over time, the CoPs are expected to take up leadership, driving the Platform's further development.

Testers drawn from the 14 cases of IBP use listed in below are variously involved in supplying, developing, testing and using these information outputs. Six CoPs have also been formally convened: *Bean* based on the Pan-African Bean Research Alliance (PABRA); *Cassava*, focusing on African cassava breeders; *Chickpea* with members from several African countries and India; *Cowpea & Soybean*, with members from Africa and USA; *Rice*, with members from the Mekong region countries; and *Sorghum*, with members from several African countries and Australia.

Integrated Breeding Platform User Cases

	User case[a]	Region and countries	Project title
1	**Beans** (Lead Institute: CIAT)	*Africa:* Ethiopia, Kenya, Malawi, Tanzania	Improving tropical legume productivity for drought-prone environments in sub-Saharan Africa
2	**Chickpeas** (Lead Institute: ICRISAT)	*Africa:* Ethiopia, Kenya *Asia:* India	Improving tropical legume productivity for drought-prone environments in sub-Saharan Africa
3	**Cowpeas** (Lead Institute: UCR)	*Africa:* Burkina Faso, Mozambique, Senegal	Improving tropical legume productivity for drought-prone environments in sub-Saharan Africa
4	**Maize** (Lead Institute: CIMMYT)	*Africa:* Angola, Ethiopia, Kenya, Malawi , Mozambique, Tanzania, Uganda, Zambia , Zimbabwe	Drought-tolerant maize for Africa
5	**Maize** (Lead Institute: CIMMYT)	*Asia:* China, India, Thailand	Drought-tolerant maize for Asia
6	**Rice** (Lead Institute: IRRI)	*Sub-Saharan Africa* *South Asia*	Stress-tolerant rice for poor farmers in Africa and South Asia
7	**Rice** (Lead Institutes: CAAS and IRRI)	*Africa* *Asia*	Green Super Rice for poor farmers of Africa and Asia
8	**Sorghum** (Lead Institute: CIRAD)	*Africa:* Ethiopia, Mali	Yield improvement of sorghum in Africa through marker-assisted recurrent selection
9	**Wheat** (Lead Institute: PBI)	*Asia:* India	Molecular-marker technologies for faster wheat breeding in India
10	**Wheat** (Lead Institutes: USDA and CIMMYT)	*Asia:* China *Africa:* Ethiopia, Kenya	Durable rust resistance in wheat
11	**Rice** (Lead Institute: AfricaRice)	*Africa:* Burkina Faso, Mali, Nigeria,	Drought-tolerant rice for West Africa
13	**Wheat** (Lead Institute: ICS–CAAS)	*Asia:* China	Breeding and selection strategies to combine and validate QTLs for WUE and heat tolerance
12	**Wheat** (Lead Institute: IARI–ICAR**)**	*Asia:* India	Breeding and selection strategies to combine and validate QTLs for WUE and heat tolerance
14	**Cassava** (Lead Institute: NRCRI, Nigeria)	*Africa*	MARS projects to breed for disease tolerance and drought tolerance in cassava in West Africa

Full names of institutes and other entities are listed in *Acronyms and abbreviations used in the IBP papers*

Members of these communities, some of whom are not directly involved in the user cases, also serve as testers and thus further widen the networks for information dissemination and peer-to-peer support. The dynamism of the various CoPs, however, remains disparate. GCP is now employing a more bottom-up approach to attract more active CoP members from outside the immediate GCP circle. This approach will be fully implemented when the Platform is opened up to a wider audience in December 2012 (see the main body of this paper).

CoPs particularly target developing-country breeding programmes where their impact is potentially large in the area of knowledge-sharing, information dissemination and capacity building generally. The communities are designed to promote interaction between scientists working on a common crop, help build alliances and facilitate coordination and synergy of research initiatives by members. The CoPs are expected to map existing knowledge and identify gaps for proactive interventions and promote innovation by creating new knowledge and developing new capabilities. They should facilitate the sharing of both information and experience for problem solving, and promote the building of a shared repertoire of resources, tools and methodologies. In doing so, they are expected to standardise professional practices by leveraging best practices.

Ultimately, given the relatively small numbers of crop scientists in developing countries, the CoPs are expected to leverage economies of scale to secure resources and opportunities for members and engage in advocacy to engender a positive environment – in policy-making, statutory and regulatory terms – that promotes food security and improved livelihoods for agriculture-dependent households.

Post-GCP sustainability and projected impact

The GCP breeding information products are intended to be freely and readily accessible to all potential users as global public goods, with minimal intellectual property encumbrances as far as possible. Where products are subject to copyright restrictions, such as publications, pertinent information is provided to enable clients to conveniently choose what they need and access it as appropriate. As a unique feature, the IBP has made considerable progress in collating, at one node, breeding information that is often distributed across different sources. The clear added value of having this information thus consolidated promises high impact on the user breeding community.

The IBP CoPs have been slow to start, but are anticipated to have considerable impact once established and active. The pioneer user cases will have completed their role when the fully functional platform is opened to the general public. However, as the informatics tools will need to develop improved and newer versions in the future, and additional services will be identified and made available through the Platform, maintaining links with breeding programmes will be necessary to ensure that the new or upgraded tools and services meet real needs. The evolution, size and composition of the CoPs will be a very good impact indicator for the Platform. Indeed, if there is no significant buy-in by the CoPs, the IBP is certain to remain strictly as a service platform with core tools and services but little collateral capacity-building and networking activities (see scenario 1 in the main body of this paper).

Post-GCP placement of Breeding Information and Communities of Practice

Breeding information management is a dynamic process that requires constant effort. If by December 2014, as envisaged, a considerable amount of relevant breeding information will be available through the Platform, then the management of this information will necessarily extend beyond GCP's lifetime. Of course, the magnitude of that activity will depend enormously on the shape and nature the Platform takes up after December 2014. Under scenario one – the austere version of the IBP – activity would be considerably reduced. Under scenario two – where core functions are extended to include strong capacity-building and networking components to make the IBP a livelier and more comprehensive platform – activity will be commensurately enhanced. The same rationale applies to the future of the CoPs beyond 2014 because, as indicated above, CoP dynamics and IBP's future form are interdependent.

The post-GCP future of both breeding information and CoPs can also be projected in a more independent way, such as by crop. Under that option, the CRPs, which maintain information repositories pertinent to their particular mandate crops, will be a perfect home for the breeding information. Today, they mainly target in-house players and immediate collaborators, but the respective crop CRPs aim for a more general dissemination of information to interested parties, include developing-country breeders.

The same applies to CoPs – the CRPs may also have communities built around their mandate crops. Good examples include the bean research networks in Africa maintained by CIAT and the rice networks of IRRI and AfricaRice. However, they usually do not provide community-interaction tools that are purpose-built, customised and online. The communities are again generally restricted to the immediate collaborators of their research projects.

Again, however, the CRPs are supported to overcome these limitations and move towards a more integrated and inclusive partnership. However, some open and very pertinent questions remain: how long will this objective take to achieve? How far will the dissemination of information and building of effective partnerships go? Would they address the expectations of the current IBP stakeholders?

Breeding information and CoPs could also be positioned in other platforms such as the Global Partnership Initiative for Plant Breeding Capacity Building (GIPB). This multi-partner platform was convened by FAO to improve institutional capacity for effective crop variety development and their distribution through seed systems (http://km.fao.org/gipb/). But that platform, as well as others developed by various universities, is not embedded in strong research components. This makes the use of accumulated breeding information more challenging and the continuity of the CoPs in a sustainable way more doubtful.

Other homes worth considering are existing crop breeding networks, such as the Coordinated Agricultural Projects (CAPs) supported by the United States Department of Agriculture (USDA). The CAPs aim to strengthen research in some selected crops (eg, legumes and small grains), education, and extension communities by focusing on the genetics and genomics aspects of nutrition in these

important food crops. Although the CAPs have significant links with developing countries in terms of partners and products, they nevertheless remain US-centric.

Another alternative, assuming that the IBP will survive beyond December 2014 in an integrated and comprehensive way, is that an entity inherits the Platform's accumulated information and community-building and interaction tools as part of its core functions. Thus, it would both add to the stock of information and also broaden the audience that the information reaches. As core functions of the entity, these activities would benefit from the most efficacious technologies and optimum allocation of resources.

Splitting IBP's core activities by crop may split the risk of failure, as each will be managed independently. Those activities would also be close to the pre-existing and active teams of breeders as exist in the CRPs and the CAPs who are part of the target audience. The fundamental question to be addressed is whether this is a significant enough advantage to counterbalance the added value of having breeding information and CoPs in an integrated effort cutting across crops, enabling a broad set of diverse users to simultaneously access tools and services to conduct their breeding activities. A new entity would also be particularly well placed to reach and integrate breeders and researchers from developing-country programmes who are not directly involved with the CGIAR CRPs, thus facilitating and accelerating their transition to molecular breeding approaches.

Either way, the need for pertinent breeding information is perpetual. It is particularly critical for developing-country breeding programmes as they progressively adopt molecular breeding technologies. It will therefore be necessary to maintain the systems and structures for identifying, generating, collecting and disseminating this information. The possible future of these functions is discussed on an overall basis in the Paper on the IBP.

Annex 2 Data Management Software

Introduction and rationale

Data management is a critical element of the Integrated Breeding Platform. A key technical constraint to the efficient management of crop information at institutional and global levels is lack of standardization and consistency. In order to develop, maintain and disseminate best practices in crop informatics for each crop, communication, training and knowledge management needs to be enhanced. A central platform designed to support these activities and technologies would be the most efficient way to achieve this.

As a general principle GCP tries to build on existing expertise and infrastructure. A major objective of the IBP is to standardize and integrate data management across partners working on the same crops. However, this standardization is not intended to create alternative systems. It should result in an analytical pipeline that will be used routinely by scientists and partners, starting with field design, field books and electronic data collection developed and used by CGIAR Centres, and ending up with breeding information that will be stored in Lead Centre databases. It is also very important that relevant, public data generated through the platform by CG and outside users be of good quality and accessible to all users through appropriate information management systems. To achieve these objectives, GCP would like to establish a formal network of Crop Lead Centres. This concept will soon be discussed at the CGIAR Consortium level.

Our vision is that best practices for breeding data management should be the responsibility of CGIAR Centres with the mandate for particular crops. This will facilitate the continuity and extension of these best practices, create ownership of the platform and ensure its sustainability beyond the life of the GCP. This initiative will not therefore create different ways of running integrated breeding activities in parallel to those promoted by the lead centres but will adopt these best practices, standardize them across partners and crops where possible, and promote them to outside users.

The public sector, particularly in developing countries, does not at this time have an alternative comprehensive suite of user-friendly tools that facilitate the collection and management of breeding data across a wide range of crops and institutions, while linking that data seamlessly with analysis and breeding decision tools for easy use. Another of IBP's key objectives is to ensure that such organisations and projects working on crop improvement for development, and which cannot afford to buy or develop quality informatics systems, should still have access to high quality and efficient methodology and tools to support their efforts.

To deploy and support community-defined standards and best practices, GCP has installed a quality data-management workflow and analytical pipeline as part of the Integrated Breeding Platform.

Activities and products

As indicated above, the IBP data management system is built on the structures used by CGIAR Centres. It links these structures through analytical and decision support tools that are being developed as part of the Platform. The IBP also stores phenotypic, genotypic and pedigree information collected by

breeders, and includes all the historical crop-related data that have been generated over time within various institutions.

The IBP project works with partners to develop common standards, and design and deploy systems. This is done partly through the use of the *Crop Ontology Project*, which ensures that consistent nomenclature is being used across breeding programmes and institutions, and provides the tools for facilitating this.

Figure 1 depicts the basic structure of data management in international agricultural research.

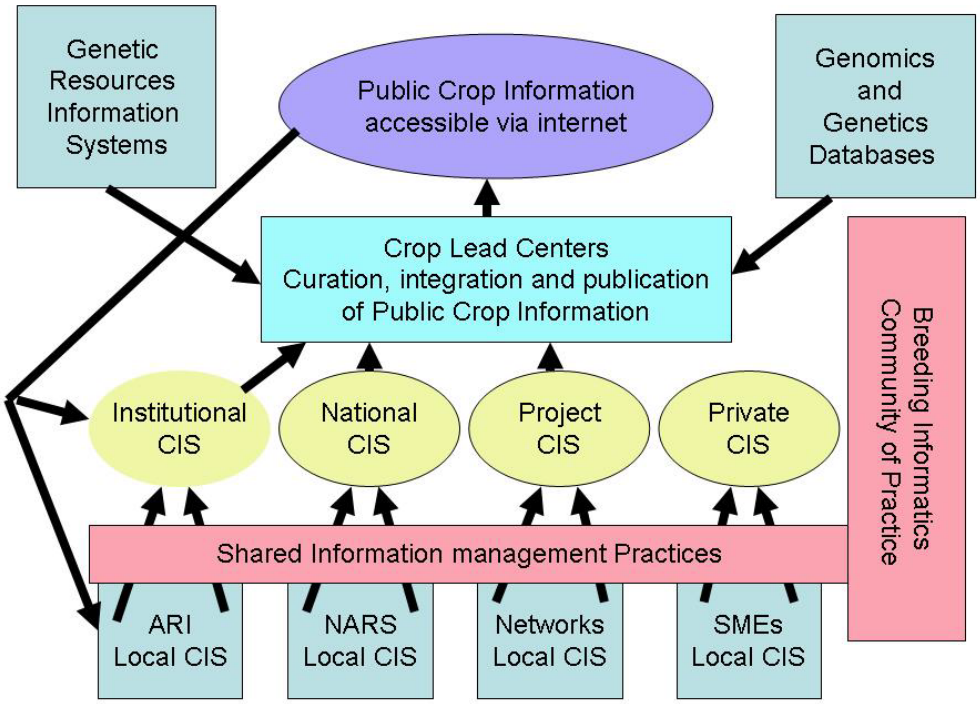


Figure 1 Information flows in plant breeding in international agricultural research

At the base of the structure, breeding projects are implemented by breeding programmes in developed and developing countries, networks of these breeding programmes and small and medium sized enterprises in the private sector. Each breeding project requires a local crop information system (local CIS), most of which are, at present, unsophisticated. A major activity of the IBP is to upgrade these systems, which are at the level where breeders conduct most of their immediate work, to take full advantage of new breeding technologies.

The user-friendly Integrated Breeding Configurable Workflow System enables users to access data management methodology and software at different stages of the breeding process, within a customisable decision roadmap, enabling efficient and convenient utilisation of standardised data structures and protocols. Figure 2 illustrates the overall structure of the CWS.

On the top row, the breeding process has been separated into a set of different activities – from project planning through data collection and storage, to data analysis, and finally breeding decision-

making. The lower row lists applications that have been developed by the IBP to meet the needs of each of those activities. These applications have been integrated into the single CWS platform in a way that data can flow seamlessly from one application to the next. The first 4 sections of the CWS deal with the logistics, database and data management tools, while the last 2 sections jointly constitute the Analytical Pipeline, discussed in Annex 3 *Data analysis and decision support methodology and software*.

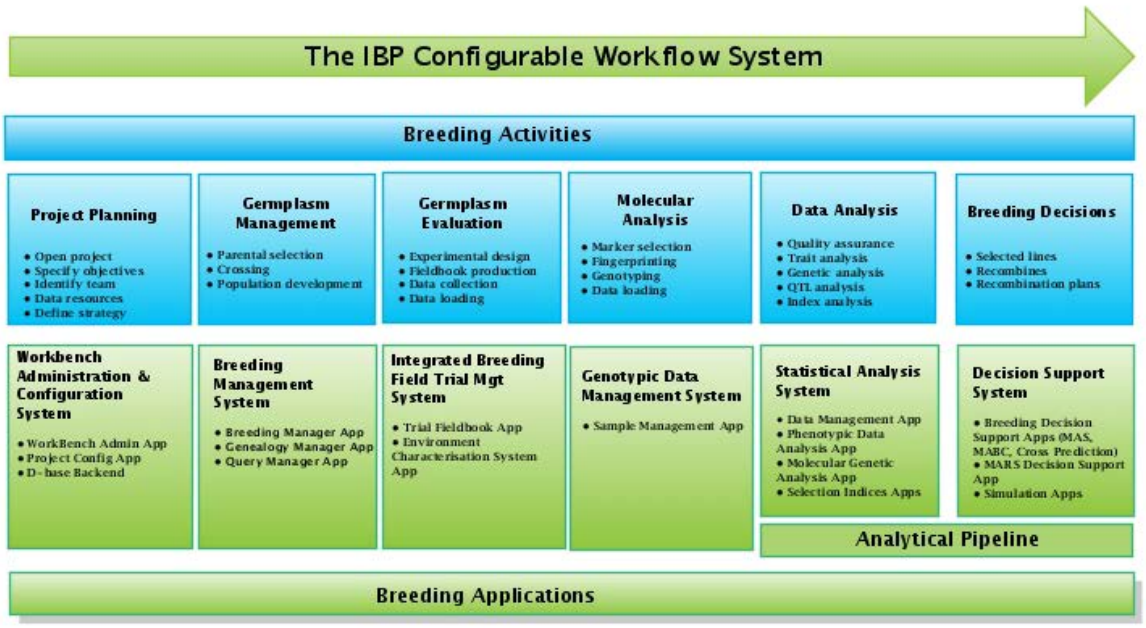


Figure 2 - Configurable Workflow System

The integrated breeding logistics and data management methodologies and software comprise a suite of applications in four categories, each with several tools as listed in Data management tools of the IB CWS:

- *Project planning*: The first category provides tools for project planning, including definition of objectives, project sites, project teams, strategy, etc.
- *Breeding/Pedigree Information Management:* The second category provides tools to facilitate sample tracking for lists of germplasm to be evaluated in nurseries, trials or laboratory assays. They manage the recording of pedigrees, chronology of breeding processes and naming of germplasm passing through a breeding project. They enable nursery composition and management of breeders' seed inventory.
- *Trial/Field Data management:* The third category provides applications which facilitate the production of electronic field books for germplasm screening, characterisation and evaluation. Electronic fieldbooks enable randomisation of experimental design, and improve data capture and management, reducing overall possible human error. They also facilitate the production of field maps and labels, including bar-coding, to ensure traceability of genotypes and samples from the field to the laboratory and back to the field – a critical element of the selection process.
- *Genotypic Data Management:* The fourth category facilitates the capture of genotypic information into local databases, information derived from the requisite laboratory work.

Data Management Tools of the IB CWS

Category	Applications	Tools
Project Planning & Logistics	IB WorkBench Administration Application	• Installation Tool • Users Management Tool • Projects Management Tool
	IB Project Configuration Application	• Locations Tool • Persons Tool • Institutions Tool • Breeding Methods Tool • Naming Conventions Tool • Storage Conventions Tool • Trait Dictionary Tool • Trait Templates Tool • Gene Catalogue Tool
	Database back-end for the IB WorkBench	• IB WorkBench Database
Breeding Management	Develop Breeding Manager Application	• Seed Inventory Tool • Germplasm Lists Tool • Crossing Block Tool • Nursery List Tool • IB Nursery Book
	Develop Geneaology Manager Application	• Pedigree Import • Pedigree Editor
	Develop Query Manager Application	• Pedigree Viewer Tool • Study Browser Tool Genotype Viewer Tool
Field Trial Management	Develop a Trial FieldBook Application	• Seed Inventory Tool • Germplasm Lists Tool • Trial Entries Tool • Data Import Tool • IB FieldBook
	Develop an Environment Characterisation System Application	• Site Characterisation Tool • Soil Data Tool • Climate Data Tool • Socioeconomic Data Tool
Genotypic Data Management	Sample Management Tool	• Gene Catalogue Tool • Sample Manager Tool • Genotyping Data Manager Tool • Fingerprinting Tool • Data Import Tool

These tools together comprise Crop Information Systems for breeding programmes, most of which are conducted by a group of partners from similar or different types of institutions. For enhanced value and effectiveness, the CISs also need to integrate with global public information for each crop.

GCP also supports a project that is building a bioinformatics pipeline to analyse next-generation sequencing data to harness the power of genotyping by sequencing (GBS) for crop improvement. Results are made available for marker-assisted breeding or genome-wide selection.

Users of both the workflow and pipeline require support and training. This support is provided through the *Data Management Service* and *Capacity-building Service*, described in separate white papers.

Sustainability and projected impact

A comprehensive suite of user-friendly tools for the collection and management of breeding data across a wide range of crops and institutions, and linking that data seamlessly with analytical and breeding decision tools for easy use, is simply not available in the public sector, especially in developing countries. Hence, the IBP data management system that supports the configurable workflow system meets an important need to help accelerate the rate of development of improved varieties for developing country farmers. It is at the early stages of deployment, but it is already being well received by the users. The full and continuous deployment of the system, along with appropriate training and maintenance support, will occur after the completion of the GCP mandate in 2014.

Data integration and exchange across teams provides opportunities for wider impact and is one of the major objectives of the IBP. The CRPs and CG centers will undoubtedly need to play an important role in the hosting and maintenance of the various databases (both phenotypic and genetic) associated with the IBP. Crop Lead Centres will need to ensure the continuous maintenance and appropriate curation of the data being stored in the central systems. If this is not realised, the result will be a bunch of independent local databases at each user organization, with no sharing of data or access to up-to-date global crop information. This is **not** what the IBP is aiming for, and it is difficult to see how the IBP data management system would be sustainable if this does not happen, and needs to be a key element in the discussions about how the IBP will be handled post-GCP.

Where possible, the future sustainability of products has been enhanced by ensuring that all components and the source code are publicly available and usable by interested organisations and communities. Nevertheless, the best option for sustainability beyond GCP is adoption by users: breeding programmes, companies, organisations or communities with a mandate to support such activities across a wide range of staple crops. But as with the other components, the Platform will still need to generate some income to support the data management methodology and software. Income could be generated through a charge back system, licensing fees or membership subscription, depending on the products. In keeping with the IBP's goal that scientists in developing countries (not including emerging economies) should access the Platform for free or at nominal cost, membership charges/fees would be adjusted to different kinds of users and would be defined according to a market-and-demand analysis and the level of donor contributions received at any given time. The key to this resource mobilization strategy is to find the right graduated payment mechanism to fully or partially support continued maintenance and development

The impact of the deployment of common standards and tools in the GCP and user case projects will be substantial in terms of accelerated and targeted crop improvement within those projects. But the potential impact beyond the GCP, dependent mainly on adoption rates and scope, is enormous as data management practices, tools and services are generally very poor in the public sector beyond the single user level. GCP is investing heavily in both support (see Annex 4 *Data management service* and Annex 5 *Breeding and Support Services*) and capacity building (see the white paper on *Capacity-*

building services and training materials) to ensure adoption as widely as possible. With numerous breeding projects using these tools, together with the data analysis and decision support tools of the CWS (see Annex 3 *Data analysis and decision support methodology and software*), even a small impact on data management at a corporate level will have a massive collective impact on crop improvement.

Post-GCP placement of Data Management Software

As noted above, the immediate objectives of GCP will be met by the end of the programme, but for the full potential of sustained high quality data and information management on crop improvement for development to be realized, post-GCP continuation will be essential. The most basic model would be to ensure that the methodologies, tools and source code are all in the public domain, allowing interested parties to adopt whatever components they require, while adding improvements to the 'open source' inventory.

The greatest impact, and the most efficient maintenance and development of these products, would however be achieved by keeping all the IBP components together – crop information, breeding logistics, data management, data analysis and decision support. To do this would require the identification of an organization or community with the mandate and resources to take on or commission the continued maintenance, support and development of all the tools in tandem. Considering the key role of the CGIAR Centers in the maintenance and curation of the centralized databases, a core part of the system, it is difficult to perceive another institution or network taking on the data management component of the IBP in a smooth and efficient way.

Another alternative, assuming that the IBP will survive beyond December 2014 in a consolidated and comprehensive way, is to create a new entity that would inherit the accumulated information and community-building and interaction tools of the Platform as part of its core functions, and both add to the stock of information and also broaden the audience that the information reaches. As core functions of the entity, these activities would benefit from the most efficacious technologies and optimum allocation of resources.

The post-GCP future of the data management methodology and software can also be projected by crop. Under that option the CRPs/Centres, which maintain information repositories pertinent to their particular mandate crops, will be a suitable home for the data management methodologies and tools. Today, they mainly target in-house players and immediate collaborators – but the respective crop CRPs could aim for a more general dissemination of information and tools to interested parties, including to developing country breeders.

Alternatively, some users (e.g., seed companies, large breeding programmes) may wish to pick certain elements of the CWS, for example the data analysis and decision support tools, which may be more easily integrated with and into their internal systems. Another option is to license the methodologies and tools to a company or consortium who will undertake the maintenance and development, deriving sufficient revenue from well-funded paying commercial users while maintaining free access for needy users.

Whichever option is determined to be most viable and desirable, the logistics and data management functionalities meet an ongoing and growing need. The maintenance, development and dissemination of the CWS of which they are an integral part should therefore be continued even after GCP, and conducted in an integrated fashion (supported by an appropriate data management service and capacity building interventions).

Annex 3 Analysis and Decision Support Software

Introduction and rationale

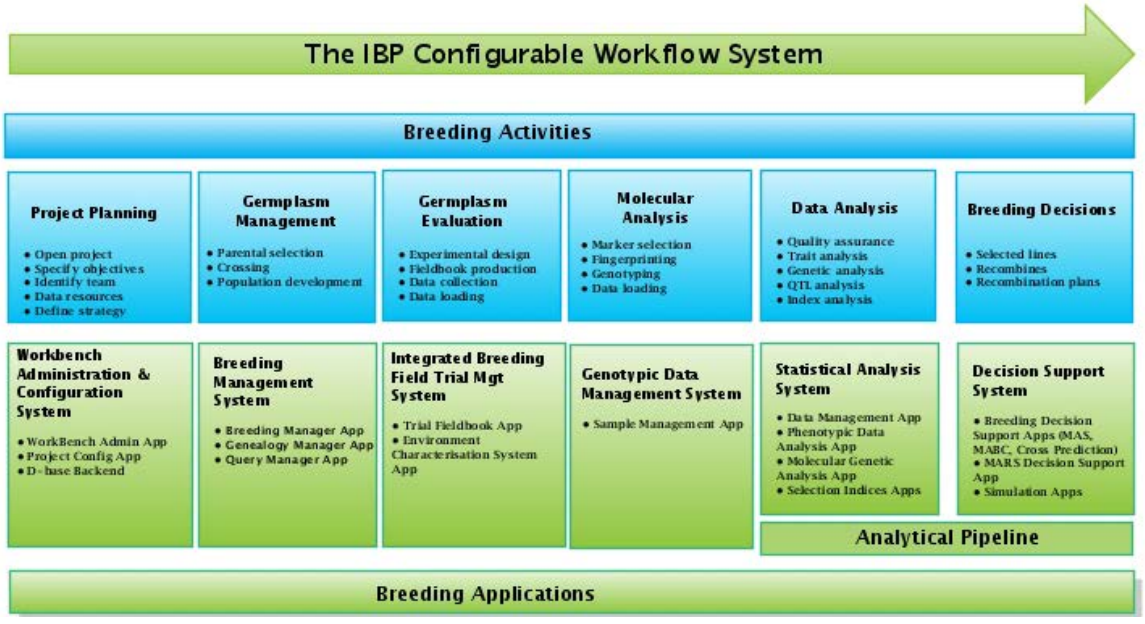
For a number of years now, plant breeders in the private sector have taken advantage of the latest molecular biology and informatics technologies to implement molecular markers in their programmes. Today, in many crops, breeding using molecular markers is considered routine, making good use of all available genetic information.

In contrast, in the public sector most activities involving markers have focused on discovery research with little routine implementation. Issues around data integration, tool accessibility and human resource capacity, are key roadblocks to public sector breeders adopting practices that are considered routine to their private sector colleagues. This is especially true in developing countries, where breeders are lagging behind in the adoption of molecular breeding techniques.

To overcome these hurdles and facilitate the wide adoption of molecular techniques in plant breeding in developing countries and render the plant breeding process more productive and more efficient, the Integrated Breeding Platform is developing a suite of data analysis, decision support and simulation tools linked to a comprehensive data management system through a user-friendly Configurable Workflow System (CWS). The CWS will provide easy access to fully integrated resources facilitating plant breeding projects, including the rapid discovery and implementation of molecular markers in the day to day activities of a breeder – with particular utility for developing country breeding programmes, in both national research institutes and small- and medium-size enterprises.

Activities and products

The Integrated Breeding Configurable Workflow System, incorporating the Analytical Pipeline, enables users to access statistical analysis, decision support and simulation tools at different stages of the breeding process, within a customisable decision roadmap, enabling efficient and convenient integration of marker information into breeding decisions. Figure 1 – IBP Configurable Workflow System Illustrates – the overall structure of the CWS.



In addition to traditional analysis applications, the CWS provides decision-support tools based on graphical visualisation of genotypes and phenotypes, and on quantitative genetic and eco-physiological modelling and simulation. On the top row, the breeding process has been separated into a set of different activities, from project planning through data collection and storage, to data analysis, and finally breeding decision-making. The lower row lists applications that have been developed by the IBP to address the informatics needs of each of those activities. All those applications have been integrated into the single CWS platform in a way that data can flow seamlessly from one application to the next. The user therefore needs to use only a single application to meet all of his or her needs. The first 4 boxes of the CWS deal with the database and data management tools, discussed in Annex 2 *Data Management Software*, while this paper deals with the last 2 boxes jointly designated as the Analytical Pipeline - the core of the data analysis and decision support tools comprising of two main applications:

- *Statistical Analysis System*: This element provides access to statistical methodologies and applications for appropriate and timely analysis of phenotype and genotype data. It also provides tools for characterising target populations of environments, weighting the influence of different environment types, analysing GxE, and epistatic interactions, mapping molecular markers and detecting marker-trait associations in breeding populations and constructing selection indices from trait and marker data.
- *Breeding Decision Support System*: This element provide analysis models and applications for conducting marker-assisted selection, after the analysis of marker trait associations in different types of populations. In addition to classical bi-parental populations, it will address complex breeding populations involving a broader diversity, such as Multi-Parent Advanced Generation Inter-cross (MAGIC) populations and populations derived from several connected bi-parental populations, as well as large-scale introgression line breeding populations involving different donors.

The Analytical Pipeline provides all the tools, listed at Tools of the Integrated Breeding Analytical Pipeline, necessary to conduct an analysis of phenotypic and genotypic data generated as part of a breeding or evaluation experiment, covering all needs from traditional breeding to advanced molecular breeding applications. For instance, the phenotypic data analysis by itself will allow the traditional and molecular breeders to evaluate their progenies using the most sophisticated statistical methods available, and to make selections on which lines to advance to the next phase of development. For marker-assisted breeding applications, the combined analysis of phenotypic data and genotypic data will allow the development of new QTL or other trait linkage information that can be used with the decision support tools. This data analysis section of the Analytical Pipeline provides state of the art statistical routines developed in two different environments (R and Genstat) to ensure the widest uptake by potential users. These are customized and integrated to reflect the philosophy of the CWS.

Marker implementation methods can be varied but the tools required need to help the breeder make a quick informed decision on what to take forward to the next generation. What plants need to be crossed; what plants can be kept and which ones can be discarded? The decision support tools provided by the CWS will help the breeder make these decisions. These include a tool (OptiMAS) to facilitate the development of new progenies combining favorable alleles from both parents in the marker-assisted recurrent selection (MARS) scheme.

Tools of the Integrated Breeding Analytical Pipeline

Category	Applications	Tools
Statistical Analysis	Data Management Application	• Experimental Design Tool • Quality Assurance Tool • Data Manipulation Tool
	Phenotypic Data Analysis Application	• SSA Tool • METs Tool • Variance Component Analysis Tool • GxE Tool
	Molecular Genetic Analysis Application	• Genetic Map Construction Tool • Haplotyping Tool • QTL Analysis Tool • QTL x Env Tool • Genetic Diversity Tool • Design Target Genotype Tool
	Selection Indices Application	• Phenotypic Selection Indices Tool • MARS Selection Indices Tool • GWS Selection Indices Tool
Breeding Decision Support	Breeding Decision Support Applications	• Molecular Breeding Design Tool • Marker Assisted Selection Tool • Marker Assisted Back-crossing Tool • Marker Assisted Recurrent Selection Tool
Simulation & Modeling	Simulation & Modelling Applications	• QuLine QuHybrid QuMARS

The integration of the analytical pipeline and OptiMAS the in CWS, along with the data management and data collection tools, will offer all tools and methodologies necessary to detect and implement QTLs in the context of a MARS scheme. In simpler applications, the breeding decision tools also provide a means to implement markers in a marker-assisted selection (MAS) and marker-assisted backcrossing (MABC) approaches which are the most commonly used approaches in marker-assisted breeding (MAB). This includes tools to simulate a breeding scheme in order evaluate different approaches for marker implementation. As other MAB approaches, such as genome-wide selection (GWS), become main-stream these will be included in the CWS pipeline and appropriate breeding decision-support tools developed.

As more marker-trait associations (MTAs) or “diagnostic” markers become available for any crop and for any trait (qualitative and quantitative) these will be stored in the relevant genotypic databases so that they can be readily selected by the breeder for screening of their germplasm (e.g. for MAS or MABC). These genotypic databases are an important part of the data management system, discussed in Annex 2 *Data Management Software*. The tools and interfaces required for this will be provided in the CWS. By being seamlessly linked to all the other tools in the CWS, a breeder can rapidly verify the decisions being made (i.e. that the plants chosen can be selfed or crossed) and make any adjustments.

Sustainability and projected impact

It is important to emphasise that the data analysis and decision support tools are an intrinsic part of the CWS. As standalone tools, they would still have some value and impact, but would not fit the philosophy of the Integrated Breeding Platform and CWS – which is to fully integrate data and tools to rapidly and easily implement breeding projects. By providing a single integrated tool where a breeder can conduct every activity linked to their programme and readily access all the data they need, the impact is much greater than simply providing a set of standalone tools. Integration and automated connection with databases limits human errors and also provides opportunities for wider impact, particularly in the area of data mining. Such an integrated suite of applications for data analysis and breeding decision-support, coupled with the data management facilities put in place by the IBP, is unique in the public sector. This is especially so in the developing world, where breeders have very limited access to informatics tools and little support to learn and use them efficiently.

In the short- to medium-term (to 2014) the CWS will be able to serve the needs of the breeding community under the custodianship of GCP. A resource mobilization strategy will be required post-GCP to ensure support and maintenance of the Analytical Pipeline as it is more widely implemented in developing-country and other breeding communities and to ensure new, improved and advanced versions of the tools are made available or adapted for changing needs and additional crops. This strategy would include ongoing training for new users of the tools, and support for changing technologies and methodologies for implementation of molecular breeding that incorporate new ways of detecting MTAs and implementing markers. Such a strategy would of necessity entail both cost sharing by beneficiaries and donor support for some elements. The key is to find the right graduated payment mechanism.

Income could be generated through a charge-back system, licensing fee or membership depending on the products. Fees/memberships would be adjusted to different kinds of users and will be defined according to a market-and-demand analysis and also the level of donor contributions received at any given time. One of IBP's goals is to ensure scientists in developing countries (not including emerging economies) access the Platform for free or at nominal cost.

Enabling the breeding community to take individual elements of the CWS, such as the analysis and decision-support tools of the Analytical Pipeline, may ensure these are continually improved and may increase adoption. However, allowing different developers to make changes to individual elements may result in an Analytical Pipeline that is less integrated with the rest of the CWS and thus less functional, hence moving away from the original incentive for its development.

A more interesting and desirable proposition is to have a community or organization take the complete CWS and be responsible for its maintenance and continuation. This could be under the IBP itself as an independent cross-cutting platform (under or outside the CGIAR umbrella) or possibly commercial entity possessing particular competencies to market the platform and provide technical support. This latter approach, in offering a product for sale may provide the revenues required for maintenance & support and further development of the Analytical Pipeline and the rest of the CWS.

Alternatively some users (e.g., commercial seed companies, large well-funded breeding programmes) may wish to pick-and-choose certain elements of the CWS, such as the Analytical Pipeline, which may be easily integratable into their internal systems; the IBP custodian organisation could then license the tools and technologies to them and use the revenue generated to ensure the continuation of the CWS.

Greatest impact and ultimate sustainability will be realized through adoption, and to maximise this, GCP has been actively engaged in extensive capacity building, and this will be reinforced with a comprehensive awareness-creation and communication effort once the portal is broadly opened to the public in December 2012. The impact of the analytical pipeline in developing countries will be particularly enhanced with the availability of efficient support services, suitable documentation and the possibility for users to join professional networks to share and exchange on common research interests.

Post-GCP placement of the Data Analysis and Decision-support Software

By the end of 2014, a finished CWS containing a complete data analysis and decision-support functionality will be provided to breeders to increase the speed and efficiency of their breeding programmes, particularly in the detection and use of Marker Traits Association (MTAs).

Although it might not be most efficient option, considering the tremendous added value in having all elements of the IBP accessible in a centralized and configurable application, the Analytical Pipeline (the set of data analysis and decision support tools) is probably the component of the IBP that can most easily be positioned elsewhere within an existing infrastructure. A home for the Analytical Pipeline can be found at other platforms such as the Global Partnership Initiative for Plant Breeding Capacity Building (GIPB), a multi-partner platform convened by FAO with the aim of improving institutional capacity for effective crop variety development and their distribution through seed systems.

These products can also be embedded in websites of crop breeding universities (e.g. Cornell in the USA, Hohenheim in Germany or Wageningen in the Netherlands) or Iowa or Illinois Breeding Centers. All these options would ensure access to the products referred to in this paper, but the impact might be limited since they would not be embedded in large and diversified (crop, users) applied breeding programmes with benefits that would come with that, given that maintenance, support services and upgrades will have to be undertaken by the host Institute.

The CWS and Analytical Pipeline can also be hosted by a CGIAR crop Lead Centres or one of the CRPs, or by a cross cutting platform at the CGIAR Consortium level. Such a hosting arrangement would be quite favourable as described in the overall IBP paper. However, whether or not such an arrangement would address the expectations of all the current IBP stakeholders and primary future clients, remains an open and very pertinent question.

As part of the deployment strategy, the CWS will also be accessible and usable online on the iPlant Collaborative cyber-infrastructure by 2013. This will enable users to run the kind of resource-intense heavy analysis made possible by emerging ultra-high throughput genotyping technologies, by taking advantage of their considerable computing power. iPlant Collaborative is a community of researchers,

educators, and students working to enrich all plant sciences through the development of cyber-infrastructure – the physical computing resources, virtual machine resources, collaborative environment and interoperable analysis software and data services that are essential components of modern biology. The iPlant cyber-infrastructure hosts analytical tools developed by other initiatives such as the IBP and will ensure access to the tools of the Platform in a 'relatively' sustainable way, given that it is supported by the National Science Foundation and United States Department of Agriculture. However, iPlant Collaborative is neither designed nor intended to provide maintenance and support services to the tools hosted on its cyber-infrastructure, nor to implement the vital capacity- building and social networking components of the IBP

Last but not least, the CWS, or individual tools such as OptiMAS, can be embedded into existing commercial analytical pipelines such as GenStat or AGROBASE Generation II, so long as provision is made for access by developing country breeders who may not be able to pay the full commercial price for such software. This option will ensure that the IBP products will be handled in a professional and sustainable way, based on proven commercial principles. Of course the main drawback of this option is that the IBP and related products will not be in the public domain anymore. The commercial priorities pursued in the future by the commercial entities owning or managing the tools might also not always align with the needs and demands of the IBP's stakeholders.

Today, there is no single product available where a breeder can effortlessly conduct their breeding activities (analyzing trait data and making informed decisions) in a fully integrated way. The analytical and decision-support tools provided by the IBP in the framework of the CWS offer this. A breeder need not go anywhere else. However, to remain relevant and up-to-date in a rapidly changing field and to meet the changing needs of the client, a strategy to ensure the continual development (including training and support) needs to be developed.

Annex 4 Data Management Service

Introduction and rationale

A quality information system is an integral component of molecular breeding. However, its usability and impact depend on its successful deployment and wide adoption by users. Thus, having a data management service that will facilitate this is crucial. Breeders, users and data managers need training and support on installing the system, customising it and using it for their breeding projects. The beneficiaries also need guidance in curating their data, and access to standards and strategies for quality control and sharing of information.

The GCP Data management service provides assistance in installing and configuring the Integrated Breeding Platform (IBP) information system for use by specific breeding programmes, as well as for global crop improvement databases. The service also assists in curating data into the platform information system for specific breeding projects and in establishing and maintaining global crop improvement databases that integrate data, especially on pedigrees and genotypes, from publicly available sources. The establishment of standard trait dictionaries will facilitate information capture to ensure quality control and further sharing of information.

The objectives of the Data management service are as follows:

1. Train and provide support to users and data managers to install, configure and maintain the IBP system.
2. Train and support breeders and data managers to capture, load and curate existing data relevant to a specific breeding programme or to global projects.
3. Train breeders to use relevant tools of the IBP information system (eg, genealogy management, data management, inventorying, and genotyping data management).
4. Develop training materials to guide users in installing and using tools.
5. Coordinate the maintenance of global crop databases and standards, such as *Trait Dictionaries*, which are important for quality control and sharing of information.

In collaboration with CGIAR Centres, the service coordinates the establishment and maintenance of global crop improvement databases that act as repositories for breeding data published by individual breeding projects.

Activities and products

The following narrative briefly describes the five areas of activities undertaken and products created to accomplish these objectives.

Training in the installation and configuration of the IBP information system
Several training workshops were and will be conducted in relation to this objective. Even as the IBP information system evolves, users are guided in installing the available applications and related

databases. Legacy tools from the International Crop Information System (ICIS) were installed in users' computers during the 2010 training sessions. Feedback received in that period led to the reduction of the installation process to only a few steps, with a facilitative 'wizard'. In 2011, the IB FieldBook was introduced through five workshops where users installed the beta version in their laptops. Subsequent feedback from users became informed further development, with IB FieldBook version 1, released in 2012. First time users directly received this version, while those with the older versions received updates.

Training in the use of tools

Over the past two years, 14 training workshops have been organised to train target groups in the use of various data management tools available from the Platform, such workshops often taking advantage of pre-planned meetings of projects of GCP's crop Research Initiatives (RI). Participants at these workshops provided invaluable feedback that informed continued development. With the first version of IB FieldBook released in February 2012, eight training sessions had been conducted on its use by the end of July 2012, focusing on managing germplasm information and generating an electronic trial fieldbook. The programmes, participants and resources used at these training sessions are ccessible on the IBP Project Wiki. As the tools have evolved and improved, based on feedback from users, they have become more user friendly in addition to offering enhanced facilities and functionalities.

Training and support in curating and capturing data relevant to breeding programmes

This is done through training events, workshops, short stays at GCP headquarters and on-site visits by GCP personnel and consultants. The training events organised around the use of tools also oriented users on best practice and protocols for data capture and curation, utilising the trainees' real-life breeding activities and data. For example, the germplasm database for the GCP Sorghum Research Initiative, that contains information on germplasm materials being used in the project, was uploaded into the the central International Sorghum Information System (ISgIS) database during training sessions. This activity also resulted in the initial *Trait Dictionary for Sorghum*. Training for the Wheat RI scientists also resulted in the establishment of the initial *Trait Dictionary for Wheat* this project.

GCP hosted several users and data managers, working with them to define protocols for curating their germplasm and uploading existing data from their projects into the appropriate crop information system, amongst other cogent activities. These data managers are expected to train and support breeders in their respective institutions, transferring the knowledge they have gained for the benefit of their entire projects.

Developing training materials

Manuals and tutorials were written to guide users during training. Most training materials were developed and customised specifically for the target crop of the course participants, covering Sorghum, Wheat and Cassava Research Initiatives. However, the rice materials were used generically in training on legacy tools, while the cowpeas materials were used generically for training on the latest version of the IB FieldBook because appropriate data were available from those crops. The IB FieldBook also has its own purpose-built tutorial that accompanies every installation. Training materials are accessible from both the IBP Project Wiki and the Integrated Breeding Portal.

***Coordinating the maintenance of global crop databases and standards such as the* Trait Dictionaries**

Central databases and information systems for GCP's nine priority crops were established in collaboration with the CGIAR Centres with the mandate for those crops. These are:

- ICASS – International Cassava Information System with established CIAT and IITA
- IChIS – International Chickpeas Information System established with ICRISAT
- IGnIS – International Groundnuts Information System established with ICRISAT
- IMIS – International Maize Information System established with CIMMYT
- IPhIS – International *Phaseolus* Information System with CIAT (for beans)
- IRIS – International Rice Information System established with IRRI and AfricaRice
- ISgIS – International Sorghum Information System established with ICRISAT
- IWIS – International Wheat Information System established with CIMMYT
- IVIS – International Vigna Information System with established IITA (for cowpeas)

The *Trait Dictionaries*, which form the basis of the trait templates used with the IB FieldBook were established by GCP partners and further documented by the above-mentioned CGIAR Centres. The crop databases and trait dictionaries can be downloaded from the IBP Portal Crop Information pages.

Post-GCP sustainability and projected impact

Information technology systems and structures require continuous technical support for both existing and new users; otherwise, adoption and use will diminish over time. Moreover, as needs change the tools must be improved and users given updates incorporating the improvements. Hence, a major concern that users often raise during training is the continuity of technical support for users of IBP informatics tools when GCP ends in 2014.

There are several ways to achieve this: one is through the various initiatives of CGIAR Centres that support their respective country partners. As their data managers become more informed about the informatics system, they will gain the knowledge and expertise in using and maintaining it – but they would have only very limited capacity to further develop the systems. However, users who are not partners of any CGIAR Centres may not access that service. Another option is through software companies that are GCP partners in the development of the tools. However, to deploy and teach data management tools for molecular breeding, effective knowledge of breeding and users' needs is a must – something which most software companies do not posses. Yet another way is for the country partners themselves to develop their own in-house expertise. However, even as partners develop their own expertise in the use of the tools, this may not necessarily guarantee support for new users who want to adopt those tools especially where the new users are from outside the primary country partners.

As with the other services, the Data Management Service will require financial support to continue to exist and thrive – either through donor funding or through fees and/or membership to be paid by users. Considering the absolute imperative to have properly managed quality data for successful breeding programmes, there is no doubt that the Data Management Service will have a significant impact on crop improvement. The full advantage and benefits of effective data management, and hence the largest impact, will be best realised at a high level of aggregation and collaboration across teams and institutions.

Post-GCP placement of Data Management Service

When GCP closes in December 2014, the primary goals for this service will certainly have been accomplished, as many training sessions, orientation events and visits will have been carried out. However, sustainable support after GCP ends is a major worry expressed by users. If users are not guaranteed that support, they will hesitate to adopt the protocols and tools. Additionally, potential new users will not adopt a system for which training and technical support is no longer available. To sustain adoption, deployment and use of the IBP informatics tools beyond 2014, it will be essential to perpetuate the activities of the Data Management Service.

The placement of the Data Management Service is more challenging compared to more concrete products such as analytical tools. This is because the placement ultimately depends on the future of the other components of the Platform, mainly: the Data management software, Analysis and decision-support software, and the Breeding and Support services – all of which are direct or indirect beneficiaries of this service. The placement of the Data Management Services cannot therefore be considered in isolation, and must be discussed within a larger context.

Considering a disaggregation of the IBP components by crop, the most logical approach would be for the Data Management Service to be positioned in the Institutions hosting the crop databases and related data management tools. If the CGIAR Centres take the lead on that one, the respective crop CRPs will be the right home for those services. If, on the other hand, the IBP remains as a support platform beyond December 2014 in an integrated and comprehensive way, such an entity shall also be particularly well placed to provide data management support to breeders and researchers from developing country programmes, both those directly involved with the Centres/CRPs as well as those who are not – facilitating and accelerating their transition to molecular breeding approaches.

Either way, whether disaggregated by crop or embedded in a single entity, the need for pertinent data management support is a perpetual one, and is particularly critical for developing country breeding programmes as they progressively adopt molecular breeding technologies. It will hence be necessary to maintain the systems and structures for the identification, generation, collection and dissemination of breeding data and relevant information. The future of those services is discussed at a global level in the main IBP white paper.

Annex 5 Breeding and Support Services

Introduction and rationale

GCP Breeding and Support Services comprise personnel, systems and structures by which users can access, through the Integrated Breeding Platform portal, various requisite services. They can also access technical and logistical support and advice in the use of those services.

Ten years ago, limited access to genotyping facilities was considered as a major bottleneck for marker applications in developing countries. If, today, few question the need for local basic laboratories, most agree that large-scale genotyping activities are best outsourced to cost-effective service laboratories, irrespective of location. Reliable phenotypic data are a must for good genetic studies. Most developing countries lack suitable field infrastructure for conducting good trials and collecting accurate phenotypic data. Therefore, improving access to homogeneous field areas, applying suitable screening protocols, paying attention to good soil preparation and homogeneous sowing are critical.

Several other hurdles limit successful, public-sector, molecular-breeding programmes: some relate to information access; others pertain to data collection, management and storage such as availability of reliable sampling and data-tracking systems; while yet others relate to access to modern analytical methodologies and tools for accurate decision-making.

The GCP Breeding and Support Services primarily aims to promote and enhance the application of modern breeding – using both well-tested and emerging technologies – in developing countries. It strives to bridge gaps in human resources and infrastructure and provide technical backup for the efficient implementation of molecular breeding approaches in those countries. At an operational level, GCP targets specific beneficiaries and, through the BSS&H, seeks to deliver a supported and guided access to marker technologies, trial management and phenotyping screening protocols. The beneficiaries can then efficiently manage and analyse the genotypic and phenotypic data generated from and by their projects.

Breeding Services can be accessed directly from a few selected service providers listed on the IBP portal or, in the case of less experienced users, by communicating with designated IBP resource persons. Support Services provide ongoing technical support to users in information management and data analysis, and in the use of these technologies. Support Services also offers backup and training in a broad set of complementary disciplines to provide support in the use of Breeding Services (and other facilities) and generally ensure that adoption of molecular breeding by users is smooth and sustainable.

Activities and products

The activities and products of the GCP Breeding and Support Services are as follows:

1. *Integrated Breeding Platform Portal* is the online gateway through which users select and download tools and instructions, order materials and procure laboratory services. The portal's helpdesks facilitate its use and ensure access for users who cannot efficiently use the web interface, by providing the elements they need via email and offline media such as CDs. The portal

also provides convenient access to crop information, informatics tools, breeding services, learning resources and tools for community building and interaction.

2. *Genetic Resources Service* facilitates access to suitable germplasm and related information from different partners. As well as providing access to breeding germplasm, this service is designed to link with GENESYS. This latter platform for genetic resources and stocks management was developed by the Global Crop Diversity Trust in close collaboration with CGIAR genebanks. The goals are to ensure quality control, and maintain and distribute materials suitable for pre-breeding activities.
3. *Marker Service* facilitates client-funded access, on concessionary terms, to different high-throughput marker service laboratories in the public and private sectors. GCP personnel provide administrative support and follow-up. This service has been active for several years now. Numerous service requests have already been fulfilled for GCP partners in different countries working on a variety of GCP mandate crops. Under this service, as many as 2,000 SNP markers have been converted for use in KBioscience's KASPar System for 10 crops. Thus, a cost-effective and reliable genotyping platform is now available at https://www.integratedbreeding.net/snp-marker-conversion.
4. *Trait and metabolite service* provides a set of options that facilitate client-funded access to laboratories specialising in the evaluation and analysis of specific traits (eg, quality), pathology screening or metabolite quantification. Analyses of certain secondary traits and metabolites that act as indicators of plant stress tolerance can potentially provide valuable information for use in breeding.
5. *Breeding plan development service* advises and supports clients in the development of viable and sustainable breeding plans. Support focuses on assessing the efficacy of given experiments through, for example, cost-benefit analyses before conducting multi-cycle molecular breeding projects, which may range from transfer of a single region (eg, transgene) to complex selection for the simultaneous transfer of dozens of regions. As part of this service, informatics applications for breeding planning are being developed for MAS, MARS and MABC. When the service is fully implemented, a resource person will be available, through the Platform, to scientists who need support. This person will also endeavour to bridge gaps between service providers and breeders.
6. *Information management & data curation service* provides assistance in installing and parametrising the IBP information system for use in specific breeding projects. It also assists with capturing and curating data, and entering them into a standardised and integrated information system. This step is critical for research data quality control and the mutually beneficial sharing of research information. The service is operated by several full-time staff and consultants, as described in the white paper on the *Data management service*.
7. *Design and analysis service* provides support on statistics, bioinformatics, quantitative genetics and molecular biology. It includes training in data generation, handling, processing and interpretation, and in experimental design from field planting to MAS and MAB schemes. It also provides assistance with 'translating' the molecular context to the breeding context, and ensures that the novel methodology developed for designing and analysing breeding trials is rapidly available to users. Numerous training and support materials have been developed under this service, and several training courses administered.
8. *Phenotyping sites and screening protocols service* facilitates access to information on phenotyping sites, protocols and potential collaborators. The service thus ensures that selection is carried out

under appropriate biotic and abiotic stresses, and that the adaptation of germplasm is well characterised. Many of the sites publicised by this service have received support from GCP to enhance both infrastructure and expertise, covering, for example, best practices, weather stations, irrigation systems, rain-out shelters and greenhouses.

9. *IP and policy advisory service* provides support on intellectual property (IP) rights and freedom to operate in the arena of biotechnology and germplasm use. The service is currently provided through a virtual IP Helpdesk, with issues referred to experts serving on the GCP Intellectual Property Advisory Committee.

The most important and unique feature of the GCP Breeding and Support Services is its integration as a 'one-stop shop' and the availability of dedicated personnel, thus enabling simultaneous and coordinated support from a single node. The services are also integrated with and based on customised capacity-building interventions, purpose-built informatics tools and collaborating service providers whose professional services are aligned with the clients' specific needs. Hence, both the help provided and services rendered are directed towards supporting the practical needs of clients.

Sustainability and projected impact

1. *Post-GCP sustainability*: GCP's Breeding and Support Services are valued by collaborators, especially those in developing countries. The services and related helpdesks address not only long-term needs that will grow with the rise in awareness and adoption of modern breeding, but also the concurrent and recurrent need for resources to keep them going. Breeding Services are self-sustaining, as users pay directly for the services they use. However, funds will be needed to maintain and continue developing the web portal through which these services are accessed, and for personnel to provide support and guidance through the Support Services. As an overall principle of the IBP philosophy, once the IBP is fully operational, targeted beneficiaries will be expected to meet a significant portion of their costs, especially with the growing awareness of the value of molecular breeding in accelerating genetic gain. Payments would probably be in the form of 'per service' fees or subscription. Possible business plans for the IBP are described in the main paper on the Platform.

2. *Impact*: The GCP Breeding and Support Services are certain to accelerate adoption of molecular breeding and enhance impact over a shorter time by speeding genetic gain and contributing to the release of improved varieties that are more productive and resilient. The Breeding Services, and, in particular, access to high-throughput genotyping services, represent a major breakthrough in the application of marker technology in developing countries. This is particularly true for less-studied crops, such as legumes, which, a few years ago, lacked even the most basic genomic resources to run simple genetic analyses.

   Through IBP, users can access as many as 2,000 KASPar SNP markers for ten crops. This will be extended to other crops, based on demand. Enabling simple and effective access to genotyping facilities to run molecular breeding makes the Breeding Services the most likely IBP component to have major impact in coming years. Although more difficult to quantify, the impact of Support

Services and related helpdesks is also anticipated to be highly significant. Impact would not be directly on breeding efficiency in the short-term, but rather in the number of breeders from developing countries with the capacity to use modern breeding as an option in their toolbox.

Post-GCP placement of Breeding and Support Services

The ease and speed of generalised adoption of molecular breeding methodologies will be greatly influenced by the support systems available for those making the transition. These systems must therefore continue (and possibly be enhanced) after GCP to prevent gradual erosion of gains made in the Programme's lifetime, thus leading to the still-birth of potential future gains. Phasing out the GCP Breeding and Support Services is therefore not a desirable option.

In principle, all breeding services can be easily transferred to any existing large initiative dealing with breeding. The services can each stand alone and could be managed in a centralised way by the CGIAR Consortium, or they could be hitched to any existing platform dealing with breeding. They would fit in well with the crop-specific nature of the CRPs, making them ideal to provide guidance and support to developing country breeding programmes, many of which will already be CRP collaborators. The services can also be directly linked to specific research projects – a natural fit within the CRPs/Centres.

The placement of the Support Services and related helpdesks is more challenging as they depend on the future of other components of the Platform, namely: the Data management methodology and tools, Analysis and decision support methodology and software and of course the Breeding Services, all of which are beneficiaries of Support Services. The placement of Support Services and Helpdesks in another entity cannot therefore be considered in isolation, and must be discussed within a larger context.

Considering the nature of the Breeding Services and Support Services presented in this paper, to create a new entity to accommodate them may not make sense, because both can be easily integrated into a large breeding initiative. The CGIAR Consortium is a strong candidate as host. However, to dismantle and disaggregate these services would dramatically reduce their potential impact, compared with keeping them integrated in a comprehensive initiative such as the IBP. The services and helpdesks are, by their nature, best rendered in a centralised and coordinated fashion and in close alignment with a capacity-building function, to maintain economies of scale.

Breeding and Support Services are an ongoing need that should continue even after GCP. The integrated approach that GCP has adopted in creating both the Breeding Services and Support Services offers enhanced value. The future of those services and helpdesk is discussed at a global level in the main white paper.