B Agricultural Research

Hugo Besemer, Chris Addison, Francesca Pelloni, Enrica M. Porcari and Nadia Manning-Thomas

1 Introduction

Agricultural science combines amongst others applied socioeconomic disciplines, applied plant animal physiology and environmental sciences (soil science, hydrology, erosion/geomorphology).

Research workflows, like for other applied sciences, depend on the disciplines and methods that are applied, as well as on the way that the organisation that does the research is embedded in the agricultural sector. This chapter was written from the perspective of the Consultative Group on International Agricultural Research (CGIAR), a global partnership that unites organisations engaged in research for sustainable development with funders, including governments, foundations and international and regional organisations. CGIAR's mission implies working for international development, but many of the processes apply to national agricultural research organisations as well. As it impossible to give a general framework for research workflows in our field, we will present case studies from the CIAGR to illustrate the diversity. Typically these workflows include processes such as:

- problem identification and analysis,
- observations/acquisition of data,
- analysis of results,
- possibly design and testing of a remedy (e.g. pest or erosion control measures),
- validation (e.g. checking that a proposed solution work at farm level),
- publishing and dissemination.

In our field, problem identification includes consultation with organisations of beneficiaries and it may be done in the framework of applications for funding. In the case of the CGIAR, these are usually organisations for international development. Such organisations may have specific requirements for the disclosure of research results and some are in the process of formulating their own policies with regard to storage and accessibility of data sets.

Data may be acquired directly through observations but may also be acquired from other parties. For example, satellite images or digital maps are used for research with a spatial component. These images are often purchased from commercial firms. Observations on farm level require collaboration with farmers and may often be done in collaboration with local institutes or for example extension organisations. The latter do the initial data collection, while the data is further processed at research centres. Both in the case of satellite images and in the case of on-farm surveys, the question arises who owns the data, as data is collected and value is added in a chain.

The validation step is crucial for agricultural research as its results need to be communicated with potential beneficiaries, like rural communities. Some decades ago there was in many countries a clear division of labour between research bodies and extension agencies that were often publicly funded. In recent years, agricultural extension has changed and it is beyond the scope to describe the very diverse way that the agricultural knowledge systems have developed in different countries and different parts of the world. But as this picture is getting less straightforward, the communication of research results beyond the circle of scientific peers is becoming a direct concern for agricultural research organisations. In this context, in 6.5 Knowledge sharing, we will be drawing some lessons from CGIAR's efforts with regard to knowledge sharing. Data sets may be used to communicate with and collaborate with scientific peers, but they may also be knowledge products to communicate the results of research with the potential beneficiaries and the general public. The case studies were selected to illustrate these issues and the diversity of scientific methods in agricultural science. The CGIAR is described as an organisation in more detail in section 2.

From the socioeconomic angle, there are two case studies:

Socioeconomic surveys: which concentrate on the level of individual households or farms. Agriculture is an activity that is often carried out by such smallholders. For these surveys, the data curation/data repository approach appears to be appropriate.

Ongoing agricultural research and development capacity survey: which concentrates on the level of national agricultural knowledge systems. While agriculture is often an activity of smallholders, research and development is often an activity that is not done within the individual enterprises, but in national or regional organisations. Raw data is acquired in a variety of forms, and is stored in a central database.

From the genetic and environmental angles, there are also two case studies:

Multisite agricultural trial database for climate change analysis:

which describes an effort where traditional outputs of agricultural research (field trials) are combined in a model with existing environmental (climate) and geographical data.

Plant genetics resources – the Singer system and further: which describes research activities where the primary output is not a collection of data (in a collection of sets or a database) but a collection of certified and documented seeds. The data collection activities are aimed at making the collections of seeds accessible for experiments where genetically uniform plant material is required and taking stock of the certification and documentation process. Technically this is a central database importing on an ad hoc basis updates from local databases.

The acquisition management, publishing and dissemination of these sets are illustrated in the example case studies chosen here.

Lessons for OpenAIRE

- Models and other integrated knowledge products may be an alternative to repositories to bring together data sets from different sources.
- Research results may be captured in databases rather than static data sets. A data preservation and re-use policy should take databases into account.

1.1 Case study: socioeconomic surveys: International Food Policy Research Institute (IFPRI)

The IFPRI is an international agricultural research centre working on informing national agricultural and food policies to find sustainable solutions for ending hunger and poverty. Much of the Institute's research work relies on data collected through socioeconomic surveys and experiments. This case study describes the steps involved and the issues affecting the acquisition, storage and dissemination of these data (Figure B.1).

The story begins with the collection of raw data in the field. This has changed recently with the adoption of new technologies for the recording of information and new approaches to capture data.

The IFPRI Mobile Experimental Economics Laboratory (IMEEL) was established in 2007 by the Markets, Trade, and Institutions Division (MTID) of the IFPRI. Its primary objective is to collect data through economics experiments in the field to better understand the behaviour of smallholders and the poor in rural areas, especially in Africa, Central America and the Caribbean, Latin America and south-east Asia (Vargas, 2010; Vargas and Viceisza, 2010). These experimental data are usually combined with survey data to understand farmers' decisions on the adoption of new technologies,

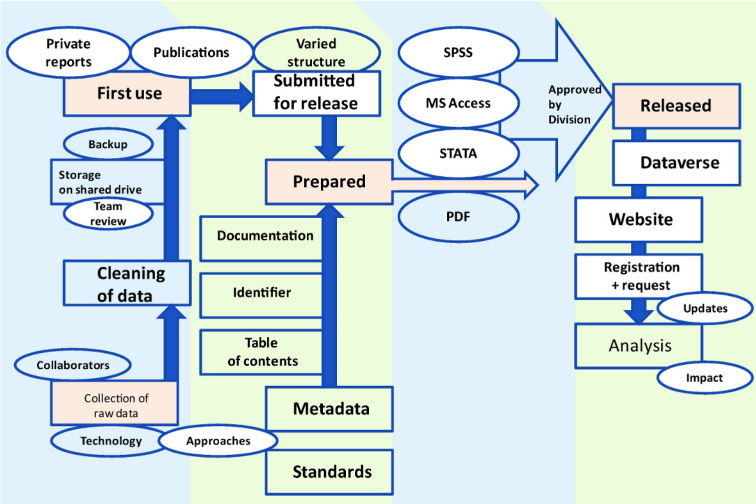

Figure 1: Workflow of data from collection by researcher through dissemination to user to analysis of use

Figure B.1 IFPRI workflow of data from collection by researcher through dissemination to user to analysis of use

participation in marketing activities; contracting arrangements and farmer groups.

A number of methods are used for collecting data including a variety of personal digital assistants, cell phones and tablets. Whilst there may be different risks in digital collection, the advantages of software to improve data collection provide increased efficiency in the collection and reduce the need for processing. For example, the software includes controlled responses and range checking, thus reducing errors in collection. The main software used for the surveys includes mQuest, Satellite Forms and Lime Survey. The output in each case is a rectangular data file readable into statistics packages or Microsoft Excel. The choices of handheld devices for data capture is based on their battery life, ease of use and their durability.

The capture of raw data involves a number of collaborators from other organisations who are often given the opportunity to use the raw data for their own studies. The capture of data in digital media in the field means that adequate backup needs to be in place in the field environment.

The data captured is then cleaned by the research team and will then be stored in a shared area for review and validation. Whilst the data is held on the shared drive it is regularly backed up from the Institute's servers.

The data will then be used within the organisation either for the production of a donor report or limited distribution report or for a publication. The software used to analyse the data during this stage is SPSS, Stata, Excel or Access. Any models produced or developed during this stage are held on the researcher's machine or the shared drives. Several of these models will be worked into a knowledge product and shared with the public through the institutions website.

The data is not released until the derived research is published. Once used for a publication, the publications review committee will require the author to submit the supporting data set. This may submitted in several forms: STATA, SPSS, Excel, Access and PDF. It is then tidied, documented and packaged by the Library and Knowledge Sharing Unit in discussion with the researcher. A table of contents will be produced to indicate the various supporting components of the data set which comprises original questionnaires and resultant data sets. Attention will be paid to ensuring anonymity of survey participants, standard formats for files where applicable and the addition of appropriate metadata.

Once approved by the Division, the resultant files and records are then published both on the internet site and in and external repository: Dataverse.[1] Dataverse is a data repository run by Harvard which provides metadata storage, file format conversion, collection management and customisation of display. Users coming through the website are asked to register and record how they will be using the data, so that analysis of use is possible later and users can be informed when there are updates of the data set or similar data sets are available.

Models and tools developed during the analysis may be similarly packaged but are normally provided online as knowledge products through the institute's website. Increasingly tools are being provided online through the site itself and through portals. For example, the welfare simulator embedded on the food security portal allows users to use their own data and run the simulator online.

There has also been a move not just to provide data online for download but to provide access to the data through application programming interfaces and visualisations of the data through interactive maps and graphs.

1.2 Case study: ongoing agricultural research and development capacity survey

Agricultural research capacity can only be developed if it can be measured. This is the premise of the formation of the Agricultural Science and Tech-

[1] http://dvn.iq.harvard.edu/dvn/dv/IFPRI.

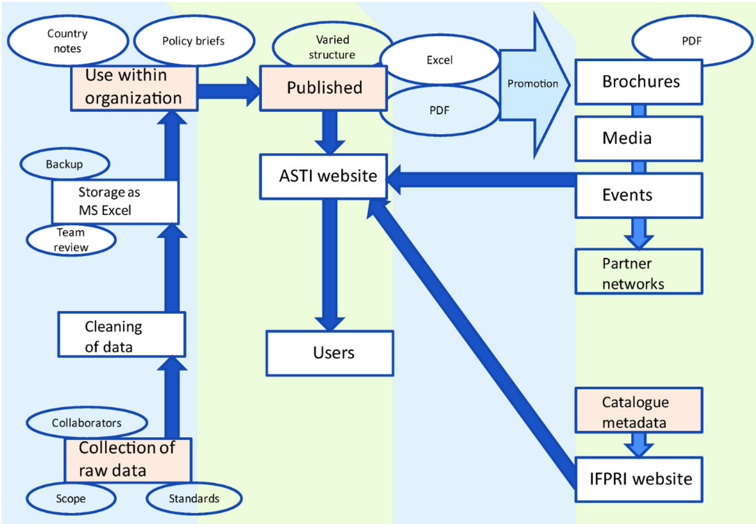

Figure B.2 ASTI data collection, management, dissemination and promotion

nology Indicators (ASTI) project to capture data on the current state of agriculture research in a selection of countries (Figure B.2). The site `http://www.asti.cgiar.org` hosts this data and the country notes and policy briefs which result from their analysis.

Raw data is collected in collaboration with partners using the OECD Frascati manual with some adjustment for the collection of data. This allows the data collected to be compared with other data sets more readily. Standard definitions are used to define scope. The data is collected with collaborators and consultants and the national partners coauthor and copublish the data in the form of country notes. These notes are produced from the raw data set but will share different levels of detail depending on the Intellectual Property Rights agreed.

The form of questionnaire for the collection differs according to the source of information. There are currently three types of questionnaire: one for NGOs and government departments, one for higher education and one for the private sector. These are constantly improved and revised as necessary.

ASTI manages a portfolio of data, from time series data across country, regional and global level covering agricultural research and development investments, institutional arrangements, funding sources, degree qualifications and female participation in agricultural research and development (ASTI, 2011; Norton, 2010).

The ASTI project has recognised the importance of promoting the data sets, realising that although the sets are valuable and there is a ready demand, active steps need to be taken to reach the potential audience. With this in mind, they have an active communication strategy and have produced a number of specific promotional products and held a number of media events and policy seminars.

The three levels of output have been the country briefs and notes; the data sets themselves and the website to allow the user to investigate the data themselves. These outputs have been complimented by promotional activities such as the ASTI blog, brochures, flyers and posters. ASTI seminars and outreach events to reach the variety of stakeholders and media and working through partners' own workshops and capacity-building programmes to raise awareness of the data sets. One of the major challenges has been to communicate with such a diverse range of stakeholders and make ASTI data known.

1.3 Case study: multisite agricultural trial database for climate change analysis

The online database developed at agtrials.org is the development platform for the CGIAR research programme on Climate Change and Food Security (CCAFS) Global Trial Sites Initiative. It shows the result of discussions between plant breeders running the agricultural trials and the geographers from a spatial data background (Figure B.3).

Agtrials.org is a development organised through the community working within the CCAFS and emphasises a pragmatic approach to the collection of metadata and data which reflects the realities of the diverse research environments involved. A series of trials were identified which could be easily incorporated into the database with emphasis on what was possible within existing time and resource constraints. The application development focussed on providing a data repository application where users could easily load historical trial metadata and information on current trials within the CCAFS programme. It needed to provide both private and public access. It built on experience on previous systems which were purely location based and incorporates the requirements of the plant breeders.

Data is provided in a variety of formats and development of the application is continuing to accommodate the design of the database and metadatabase, which can cope with the different types of user. Researchers also provide, where available, information on weather conditions during the trial and soil characteristics. There was no off-the-shelf solution to this requirement and the project develops with the contracts between the programme and researchers

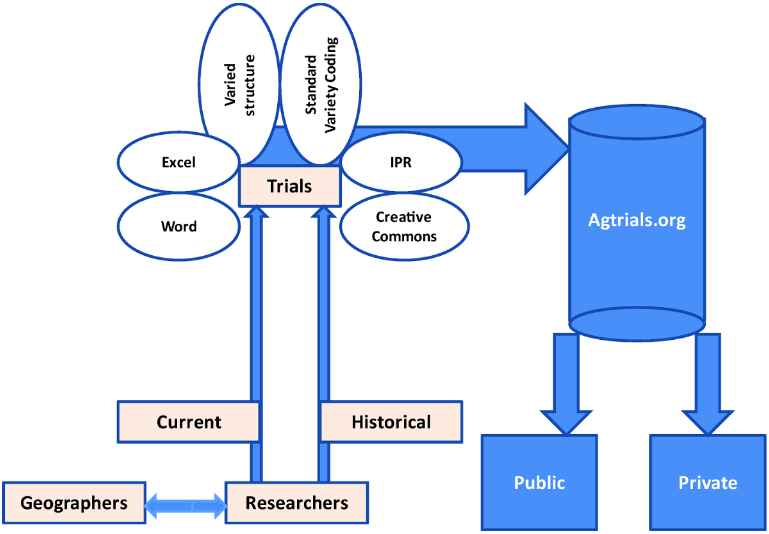

Figure B.3 Trials data storage and sharing using agtrials.org

developing hand in hand with this reference system. Most data is in Microsoft Excel file formats or Microsoft Word with no standardisation on format, but all sets follow a standard nomenclature for varieties. The technical format of the trial results has been kept open at this stage to encourage registration of a variety of sets.

Whilst each data set will have a statement on intellectual property rights, the same rights are not used across the site. Guidance is provided on the use of Creative Commons Licences but as many different organisations are involved there is no blanket statement. A series of user guidelines are available to explain the use of data from the site. In addition to current and historical trials, there is now an option to add "simulated" trials from crop simulation models. More models are planned to be developed within the group.

The interesting part of this system is not the database alone but the process by which the community is developing a data reference point for the CCAFS programme, with the dual approach of developing the research relationships with the programme and consolidating the reference index for the trials involved. The subsequent phases of the project will include models that can identify analogue environments so that the result of one set of trials can provide information on the performance of a variety in a similar environment elsewhere.

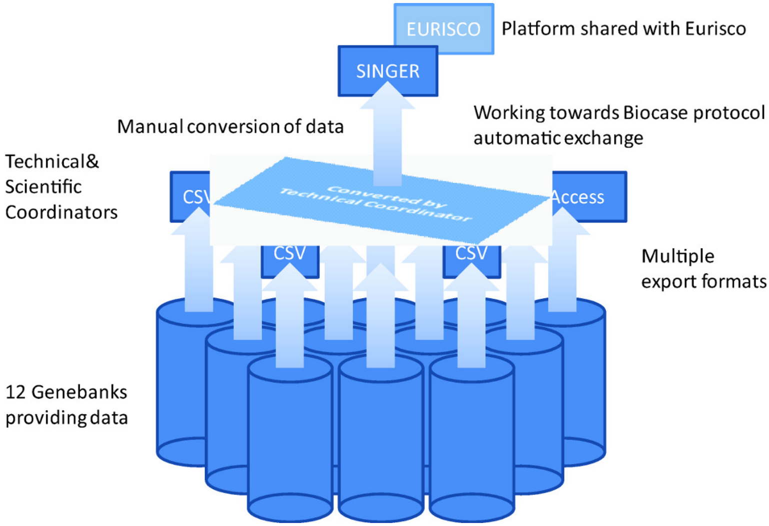

Figure B.4 The Singer system providing access to the CGIAR gene banks

1.4 Case study: plant genetics resources: the Singer system and further

There are a variety of data-oriented systems maintained by CGIAR centres and partners in the field of plant genetic resources. We will now concentrate on the oldest and most mature system. Singer gives access to the collections of the gene banks of different centres (Figure B.4). Of more recent data is the Crop Genebank Knowledge Base where instructional materials are collected on best practices for gene banks. For the "Generation Challenge Program", a facility is under development to collect data sets from this programme on molecular plant breeding and there is a portal to collect field reports on different crops.

The basic currency that is dealt with in gene banks and the Singer system are accessions, certified plant propagation, such as bags of seeds. Certain fields in agronomic research cannot do without them and that is probably the reason why this is one of the oldest and most mature cross-centre data-oriented operations within the CGIAR. The system is managed through an informal working group of technical and scientific representatives from the 12 participating gene banks. There is a scientific coordinator at Bioversity International as well as a technical manager who performs all database-oriented operations.

Data from the different gene banks are sent in by the different gene banks in a wide variety of technical formats, like MS/Access databases, CSV files, etc. Conversion to the central database is performed on case-by-case bases by the technical coordinator at Bioversity. Around 2005, attempts were made to develop a more standardised updating method using the WSDL/UDDI web services technology that was developing at that time. The Biocase[2] protocol was developed in conjunction with Singer, and was deployed at six centres. However, there were two bottlenecks: the performance (speed) of the system that implemented the protocol and the relative difficulty to produce the flat files that were required by the system from the various database implementations with which the participating gene banks are managed. The Biocase protocol had been implemented successfully elsewhere, for the Global Biodiversity Information Facility (GBIF). In the Singer network, it is recognised that data-transfer methods and an upload facility should be developed in the years to come. But the Singer case shows that if there is a perceived need for exchange a cooperative system can be kept alive without a sophisticated technical backbone. Until now, the most important investments have been done in the intellectual foundations of the system.

The minimal data elements to describe an accession have been laid down as the FAO/IPGRI Multicrop-passport Descriptors[3] that were agreed in 1997 and updated in 2001 … (Alercia, Diulgheroff, & Metz, 2001). For specific needs, there are extensions like the guidelines for developers of crop descriptor lists.[4] The data model of the Singer database is documented on the Singer website, thus giving further guidance for data harmonisation.

With regard to data quality, there is an ongoing capacity development effort aiming at improving the management of gene banks. Instruction materials can be found at the Crop Genebank Knowledge Base.[5]

There are internal agreements on how the data that is entered in the system is used. The purpose of the Singer system is the discovery of accessions and it should lead to a transaction whereby a scientist requests seeds or other plant propagation materials for further research. These transactions, the documentation to come with the material and the obligation to share results with the originators of the material are governed by the "Standard Material Transfer

[2] http://www.biocase.org/products/protocols/index.shtml.

[3] http://www.bioversityinternational.org/nc/publications/publication/issue/faoipgri_multi_crop_passport_descriptors.html.

[4] http://www.bioversityinternational.org/index.php?id=19&user_bioversitypublications_pi1[showUid]=3070.

[5] http://cropgenebank.sgrp.cgiar.org.

Agreement (SMTA)".[6] Singer data is shared with the European network of gene banks, Eurisco, that is using the same database facilities at Bioversity.[7]

The most important lesson from this case is the need to invest in the intellectual infrastructure of a network for data exchange.

2 Current status of research infrastructure workflows and research life cycle

2.1 Introduction to the research infrastructure

The CGIAR is a global partnership that unites organisations engaged in research for sustainable development with the funders of this work. The funders include developing and industrialised country governments, foundations and international and regional organisations. CGIAR research is dedicated to reducing poverty and hunger, improving human health and nutrition and enhancing ecosystem resilience. It is carried out by 15 members of the Consortium of International Agricultural Research Centers in close collaboration with hundreds of partner organisations, including national and regional research institutes, civil organisations, academic institutions and the private sector.

A research organisation like the CGIAR that studies agriculture from all different angles should be compared to a university as a whole rather than to individual research groups and institutes. For the CGIAR, an extra complication is that the organisation operates in centres on different continents. The centres are partly organised on a disciplinary basis (e.g. rice, tropical crops, genetics) but increasingly on a more multidisciplinary basis (e.g. arid environments, agroforestry) and along new interdisciplinary programmatic axes, the CGIAR research programmes.

The Consortium of International Agricultural Research Centers[8] was established in April 2010, as part of a major reform of the CGIAR, this year celebrating its 40th year. The Consortium was formed to ensure closer alignment with the needs of partners and beneficiaries and to lead, coordinate and support the 15 member centres that make up the Consortium, some of which have been carrying out agricultural research with resource-poor farmers and their communities, for over 50 years. The Consortium supports and facilitates system-level approaches and interactions and has responsibility for

[6] http://singer.cgiar.org/index.jsp?page=smta.
[7] http://eurisco.ecpgr.org/static/about_eurisco.html.
[8] http://consortium.cgiar.org.

formulating strategy[9] and for developing multiyear and multicentre research programmes that implement on that strategy.[10] The Consortium employs over 9000 staff operating in over 150 locations (Figure B.5).

2.2 Scientists, centres and system-wide programmes

Whilst individual scientists are employed by one of the CGIAR centres, they are increasingly outposted on other campuses of other centres or partner organisations. They therefore use the research infrastructure of their host, their employer and the programme for which they work. A few of these examples were drawn upon in the case studies already presented. The first example showed the situation in a centre, the next the situation for a CGIAR research programme and the others for existing system-wide programmes. The solutions vary for different work, but groups sharing data-platform requirements across the CGIAR centres (e.g. the Consortium for Spatial Information; see 4.5.2) are increasingly using shared data platforms and carrying out joint developments.

2.3 Knowledge sharing

The approach to the role of knowledge sharing in the CGIAR has changed significantly during the last 5 years. The system has developed a dedicated group to encourage knowledge sharing across the system and many of the centres now have job titles including knowledge sharing.

Research organisations like the CGIAR cannot be satisfied just knowing they have produced high-quality science. It is essential that the outputs of their research are communicated and put to use, in the village, on the ground, in the lab or across the negotiating table.

Therefore, the Triple-A Framework[11] was developed by the CGIAR ICT-KM programme looking at the availability, accessibility and applicability of research outputs and knowledge from the CGIAR. According to the framework, the three As are:

- **availability:** "can I find it?" – the need to assemble and store outputs so they will be permanently accessible and to describe them in systems so others know, and can find, what has been produced.

[9] See the Strategy and Results Framework (SRF): http://consortium.cgiar.org/our-strategic-research-framework.

[10] See the CGIAR Research Programs (CRPs): http://consortium.cgiar.org/our-strategic-research-framework/cgiar-research-programs-crps.

[11] http://ictkm.cgiar.org/document_library/program_docs/ICT-KM%20AAA_complete.pdf.

Figure B.5 Research structure of the CGIAR

- **accessibility:** "can I put my hands on it?" – the need to make outputs as easy to find and share and as open as possible, in the sense that others are free to use, re-use and redistribute them.
- **applicability:** "can I make use of it?" – the need to ensure that research and innovation processes are open to different sources of knowledge and outputs that are easy to adapt, transform, apply and re-use.

The framework is aimed at managers, researchers and information professionals to help them better understand the current AAA status of their research knowledge, how to identify, choose and develop pathways to improved accessibility for their outputs and eventually to improve chances that they will be put to use.

The first part of the framework is a benchmarking exercise which seeks to evaluate and measure the availability and accessibility of research outputs at a given time. This then helps CGIAR centres and programmes and their scientists decide on the level of AAA they want for their research outputs and also the pathways with which to turn these outputs into international public goods.

The Triple A approach has been developed and promoted to encourage sharing of international public goods produced by research. There has been more adoption of action-oriented research, more knowledge sharing during projects, changes in the peer-review process and more interim results are made available. There are new requirements from external stakeholders, such as journalists requesting access to data. There are new ways of working with fellow researchers outside the organisation, such as platforms like Basecamp and wider use of social media (Figure B.6).

3 Current status of Open Access in agriculture

As there is no clear indicator of how to measure the state of Open Access in a domain, we have approached it in two ways: (i) assessing how Open Access journals are indexed in major indexes, i.e. Web of Science (ISI) and Scopus; and (ii) overviewing Open Access document repositories within the CGIAR.

3.1 Coverage of agricultural Open Access journals in scientific journal metrics indexes

Our main question address the success of the Golden route to Open Access for agriculture, compared with two other subject domains.

[12] http://ictkm.cgiar.org/document_library/program_docs/ICT-KM%20AAA_complete.pdf.

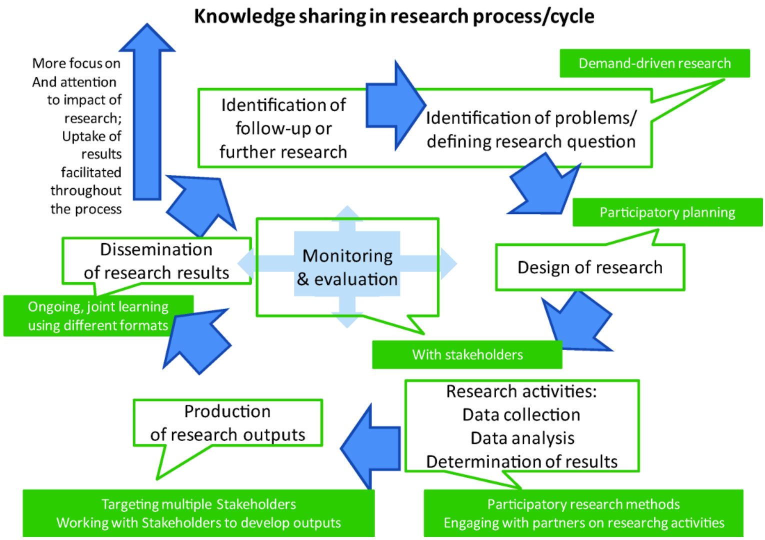

Figure B.6 Knowledge sharing in research processes and cycles (from Nadia Manning-Thomas from background note[12] for the Consortium of CGIAR Centers)

The most comprehensive list of Open Access journals is the Directory of Open Access Journals (DOAJ). We have matched (using title, ISSN and e-ISSN) the DOAJ journal list with the list of the most important lists for scientific journal metrics, i.e. Scopus, that calculates Scimago Journal Rankings – SJR and SNIP values) and the Journal Citation Report (ISI) that calculates the journal impact factor (IF).

For a cross comparison, we kept the DOAJ subject classification for each journal. The coverage of Open Access journals in different subfields of agriculture is given in Table B.1 and Figure B.7. In short, an Open Access journal in agriculture has a chance of 38% to be included in Scopus and 27% to be included in JCR (ISI).

In Table B.2 we compare these figures with the field of biology. The table shows that Open Access publishing in the field of biology is more successful than in the field of agriculture.

Finally we did the same for the field of medicine and health sciences (Table B.3). These results indicate that Open Access journals are less successful in medicine than in agriculture and in biology.

The question remains what the percentage of Open Access journals is of the total of journals in Scopus and JCR. We can make this comparison partially

Table B.1 Publications in Open Access journals: agriculture

DOAJ category	Scopus	JCR (ISI)	DOAJ
Agriculture (general)	33	24	104
Animal sciences	30	24	74
Aquaculture and fisheries	6	3	9
Biotechnology	16	10	25
Forestry	7	7	28
Nutrition and food sciences	4	5	26
Plant Sciences	19	10	33
Total	115	83	299
Percentage of DOAJ	38	27	100

Table B.2 Publications in Open Access journals: biology and agriculture

DOAJ category	Scopus	WOS (ISI)	DOAJ
Biochemistry	15	12	36
Biology	63	53	65
Botany	23	16	63
Cytology	6	2	7
Ecology	11	5	35
Genetics	20	17	37
Microbiology	14	11	38
Total	152	116	281
Percentage of DOAJ	54	41	100

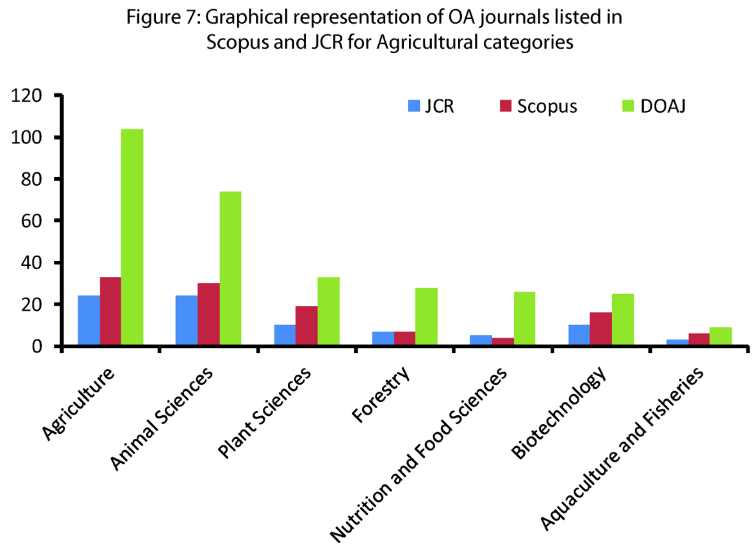

Figure B.7 Graphical representation of Open Access journals listed in Scopus and JCR for agricultural categories

for Scopus, and not at all for JCR(ISI): in both cases, the subject categorisation is different. For JCR a meaningful comparison is not possible; Scopus lists the Agricultural and Biological sciences as one category. In that category, 13.4% of the journals are Open Access, against 5.9% of the journals in the medical field.

The overall conclusion of this exercise is that agriculture is not behind the other research fields studied. One caveat is we have only looked at the total number of journals listed, not the relative importance of the journals (as expressed through IF, SJR and SNIP values).

3.2 Open Access repositories in the CGIAR

CGIAR's activities to make its publications available and accessible has resulted in publication databases/institutional repositories at all centres (Table B.4).

These repositories have been analysed in an article by Arivananthan, Ballantyne and Porcari, (2010). The content of the repositories from six centres was assessed. It appeared that there is a huge variation with regard to the availability of full text for publications (from 19% to 100% of the document descriptions). Especially articles from peer-reviewed journals were missing.

Table B.3 Publications in Open Access journals: medicine and health sciences

DOAJ category	Scopus	WOS (ISI)	DOAJ
Medicine (general)	167	69	378
Allergy and immunology	12	3	19
Anatomy	3	1	10
Anaesthesiology	5	0	11
Cardiovascular	22	11	59
Dentistry	14	3	67
Dermatology	9	1	20
Gastroenterology	9	3	26
Gynaecology and obstetrics	9	1	32
Internal medicine	104	36	237
Neurology	34	15	80
Nursing	6	2	31
Oncology	31	9	66
Ophthalmology	7	1	26
Otorhinolaryngology	3	0	17
Pathology	15	2	32
Pediatrics	14	3	44
Pharmacy and materia medica	29	7	65
Physiology	13	9	25
Psychiatry	16	7	40
Public health	50	17	47
Sports medicine	3	1	17
Surgery	24	9	69
Therapeutics	29	12	64
Urology	9	0	15
Total	637	222	1497
Percentage of DOAJ	42	14	100

Around 40% of peer-reviewed journal articles and 54% contributions to peer-reviewed books could be found in Google Scholar. However, it should be remarked that, since the study, many centres have collaborated in the Google Books publishers programme and this may likely have improved the coverage of books and book chapters.

To improve the accessibility of CGIAR's publications, there are more opportunities such as collecting pre-prints of articles in peer-reviewed journals (Green route to Open Access). It is suggested that the coverage in search engines can be improved, for example by uploading sitemap files. In the Opendoar[13] registry of Open Access repositories, seven CGIAR centres can

[13] http://www.opendoar.org.

be found, while 11 centres participate in FAO's AGRIS system.[14] Through AGRIS, these publications are also indexed in Google Scholar, but material from centres that do not participate in AGRIS can also be found back there. It is beyond the scope of this study to check systematically the coverage of CGIAR publications in these external systems, but it would be interesting to see how it develops in view of these efforts.

Table B.4 Open Access repositories in the CGIAR

Institute/programme	Start year of systematic collection of full texts	Earliest publication
Bioversity[15]	2004	1977
CGIAR secretariat[16]	"latest titles"	
Center for International Forestry Research (CIFOR)[17]	2001	1993
International Center for Agricultural Research in the Dry Areas (ICARDA)[18]	2005	1977
International Center for Tropical Agriculture (CIAT)[19]	2001	
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)[20]	2001	
International Food Policy Research Institute (IFPRI)[21]	2000	
International Institute of Tropical Agriculture (IITA)[22]	2005	1990
International Livestock Research Institute (ILRI)[23]	2008	1977
International Maize and Wheat Improvement Center (CIMMYT)[24]		
International Potato Center (CIP)[25]		
International Rice Research Institute (IRRI)[26]	2007	1999

[14] http://agris.fao.org.
[15] http://www.bioversityinternational.org/publications/search.html.
[16] http://www.cgiar.org/publications/secretariat.html.
[17] http://www.cifor.cgiar.org/online-library/browse.html.
[18] http://icarda.catalog.cgiar.org/textbase/search.htm.
[19] http://ciat.catalog.cgiar.org/ciat_bibliography.html.
[20] http://dspace.icrisat.ac.in.
[21] http://www.ifpri.org/pubs/pubs_menu.asp.
[22] http://biblio.iita.org/index.php?page=pubyear&kind=year&type=iita.
[23] http://mahider.ilri.org.
[24] http://www.cimmyt.org/en/services/library/recent-publications.
[25] http://cip.catalog.cgiar.org/cat-cip.asp.
[26] http://ricelib.irri.cgiar.org:81/screens/opacmenu.html.

International Water Management Institute (IWMI)[27]	1984	
World Agroforestry Centre (ICRAF)[28]	2004	1978

In 2006, CGIAR launched the CGIAR Virtual Library providing access to research on agriculture, hunger, poverty and the environment. This is a shared, integrated service that allows users to tap into leading agricultural information databases, including the online libraries of all 15 CGIAR centres.

3.3 Open Access mandates

Advocacy is an important component of the CGIAR's Open Access policy. In 2004, the idea that research outputs should be treated as global public goods was introduced. Four years later, the Triple-A framework of availability, accessibility and applicability was introduced to encourage specific activities to communicate CGIAR knowledge to potential beneficiaries.

To create more awareness of Open Access to publications and related issues, a workshop was held at Bioversity in 2010. In the same year, a discussion about deposit mandates arose. Two centres (CIAT and ICRISAT) made their respective mandates public. Both statements require that scientists deposit their version of an article as soon as it is accepted by a journal. Neither statement includes a clause prohibiting publication in non-Open Access journals, but the CIAT statement requires scientists to consult the intellectual property rights officer before they transfer their copyrights.

In 2010, there was also a discussion about whether there should be a deposit mandate covering the entire CGIAR. This discussion was instigated by a letter sent by a number of science writers to CGIAR management.

The CGIAR has, however, recently undergone a reform process resulting in the establishment of a new legal entity, the Consortium of Agricultural Centers, supported by a Consortium Office. The Consortium was established to lead, coordinate and support centre research and cross-centre activities through the new CGIAR research programmes. While the centres may be developing individual Open Access policies, it is recognised that a system-wide strategy and supporting mechanisms would improve and speed up the Open Access process. The development of such a strategy falls under the Consortium's mandate and is included in its agenda as part of the Strategy Results Framework.[29]

[27] http://iwmi.catalog.cgiar.org/qryscr/catalogbs.htm.
[28] http://www.worldagroforestry.org/our_products/publications.
[29] http://consortium.cgiar.org/wp-content/uploads/2011/08/CGIAR-SRF-Feb_20_2011.pdf.

4 Open Access to data: overview of CGIAR data sets

4.1 Introduction

We do not have major publishers' systems to examine the position of Open Access for data across the sector. We can, however, look across the CGIAR as a data publisher at the various services provided. There is not an accepted measure for the "openness" of data sets like the "Romeo-Sherpa colours" that are commonly used to indicate how "open" a publication is. To test such a measure we have attempted to classify the data sets that are made available according to the 5 star scheme developed by Tim Berners Lee to assess the degree to which data is made available openly online:[30]

⋆ Make your data available on the web (any format)
⋆ ⋆ Make data available as structured data (e.g. Excel instead of image scan of a table)
⋆ ⋆ ⋆ Use a non-proprietary format (e.g. csv instead of Excel)
⋆ ⋆ ⋆ ⋆ Use URLs to identify things, so that people can point at your data
⋆ ⋆ ⋆ ⋆ ⋆ Link your data to other people's data to provide context

It is not straightforward to classify a system with one star as the data within that system may be varied, the services below have been grouped to reflect the majority of their data content. Examples range from data only available by email request which falls outside the star system through to linked data sets which are fully marked up to comply with linked data requirements.[31] An assessment based on current descriptions (Figure B.8) shows the breakdown of services offering data in the various star categories.

In our view, the classification exercise using the 5 star system goes some way to indicating the degree of opening access to data, as discussed below. Many systems in the study, however, appear to be outside the system or in more than one category. The 5 star system cannot be used alone as a measure of openness as it is designed to apply to the web and software retrieval rather than to the end user directly.

4.2 Outside the star system

No star Data available only on request (any format)

The system listed below shows the data available from centres only by mail, fax or email request.

[30] http://lab.linkeddata.deri.ie/2010/star-scheme-by-example.
[31] http://data.ifpri.org/rdf/ghi.

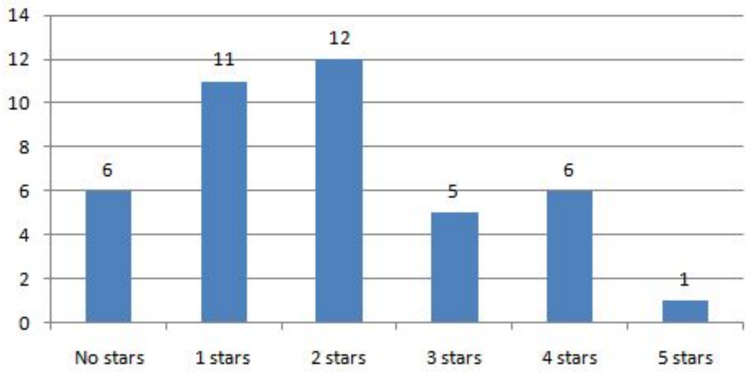

Figure B.8 Approximate number of services offering data in each star category on basis of descriptions

4.2.1 Center for International Forestry Research (CIFOR)[32]

Bogor Barat, Indonesia.

Data policy: access to the data below is possible only via mail, fax or email requests.

- **DOMAIN:**[33] the DOMAIN software uses Geographical Information System layers of environmental factors, such as climate, soil and land use, to construct an environmental habitat envelope or domain on the basis of points for the known distribution points of a species. The application then generates a map showing similarities across areas within the target region.
- **Criteria and Indicators:**[34] the Criteria and Indicators Toolbox Series comprises nine tools that were developed during the CIFOR project on Testing Criteria and Indicators for Sustainable Forest Management. The tools are aimed to help users develop and assess criteria and indicators of sustainable and equitable forest management.
- **FLORES:**[35] the Forest Land Oriented Resource Envisioning System is a model to help explore the consequences at landscape scale of policies and other initiatives intended to influence land use in tropical development. It seeks to provide an accessible platform to foster interdisciplinary collaboration between researchers and resource managers and to facilitate empirical tests of hypotheses and other propositions.

[32] http://www.cifor.cgiar.org
[33] http://www.cifor.cgiar.org/online-library/research-tools/domain.html.
[34] http://www.cifor.cgiar.org/acm/pub/toolbox.html.
[35] http://www.cifor.cgiar.org/online-library/research-tools/flores.html.

- **TROPIS:**[36] the Tree Growth and Permanent Plot Information System promotes more effective use of existing data and knowledge about tree growth in both planted and natural forests throughout the world.
- **VegClass:**[37] is a rapid, cost-effective method of surveying and classifying vegetation in forested landscape mosaics, developed using a minimum combination of variables including vegetation structure, plant species and plant functional types.

4.3 Data available online in publications

⋆ Make your data available on the web (any format)

All centres make their publications available online (see 3.2 Open Access repositories in the CGIAR). A number of data sets are represented within the documents without necessarily being available as separate data files. These collections date back to the 1970s; more recent reports will, of course, have data sets available separately as data files. Some examples are given below:

4.3.1 Africa Rice Center (West Africa Rice Development Association)[38]

Cotonou, Benin.

- **Technical Reports:**[39] which include some data on rice statistics and the genetic evaluation of rice in Africa.

4.3.2 Bioversity International (formally known as the IPGRI)[40]

Rome, Italy. English; documents also available in Chinese, French, Italian, Spanish, Portuguese, and Russian.

- **Bioversity International Publications:**[41] a web-based institutional repository providing access to publications that have been published or sponsored by Bioversity.

4.3.3 Center for International Forestry Research (CIFOR)[42]

Bogor Barat, Indonesia

[36] http://www.cifor.cgiar.org/online-library/research-tools/tropis.html.
[37] http://www.cifor.cgiar.org/online-library/research-tools/vegclass.html.
[38] http://www.warda.cgiar.org.
[39] http://www.warda.cgiar.org/warda/techreport.asp.
[40] http://www.bioversityinternational.org.
[41] http://www.bioversityinternational.org/publications.
[42] http://www.cifor.cgiar.org.

- **Publication repository**:[43] catalogue of CIFOR publications searchable by author, title, publication year, language and type of publication.

4.3.4 International Potato Center (CIP)[44]

Lima, Peru.

- **Publications:**[45] catalogue of CIP Publications (books, manuals, reports, working papers and training materials distributed for sale by CIP) on potato, sweet potato and Andean roots and tubers. Database searchable by author, title, publication year, keyword and language.

4.3.5 World Agroforestry Centre[46]

Nairobi, Kenya.

- **Publications:**[47] a wide range of the World Agroforestry Centre publications are searchable and available online.

4.3.6 International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)[48]

Andhra Pradesh, India.

- **Owned services:** AGROPEDIA,[49] ICRISAT Open Access repository,[50] ICRISAT institutional repository.[51]

4.3.7 International Livestock Research Institute (ILRI)[52]

Nairobi, Kenya.

- **Mahider:**[53] institutional repository of ILRI. Mahider is a complete index of research outputs produced by. Where available, the repository gives access to the full content of the output. The repository is built using Dspace; most of the outputs listed are Open Access.

[43] http://www.cifor.org/online-library/browse.html.
[44] http://www.cipotato.org.
[45] http://cip.catalog.cgiar.org/catalogs_menu.asp.
[46] http://www.worldagroforestrycentre.org.
[47] http://www.worldagroforestry.org/our_products/publications.
[48] http://www.icrisat.org.
[49] http://ring.ciard.net/agropedia.
[50] http://ring.ciard.net/icrisat-open-access-repository.
[51] http://ring.ciard.net/icrisat-institutional-repository.
[52] http://www.ilri.org.
[53] http://mahider.ilri.org.

4.3.8 International Rice Research Institute (IRRI)[54]

Los Banos, The Philippines.

- **Publications repository:**[55] owned services: DSpace at IRRI.[56]

4.3.9 WorldFish Center[57]

Penang, Malaysia. English and native languages.

- **Publications repository:**[58] covers WorldFish publications as well as works in other publications by WorldFish scientists and researchers

4.3.10 International Water Management Institute (IWMI)[59]

Colombo, Sri Lanka.

- **Publications repository:**[60] provides access to scientific documents published or jointly sponsored by IWMI. All IWMI publications are global public goods and are available for free from their online database.

4.4 Structured data sets

⋆ ⋆ Make data available as structured data (e.g. Excel instead of image scan of a table)

⋆ ⋆ ⋆ Use a non-proprietary format (e.g. csv instead of Excel)

Several of these services provide data of both types and so have been grouped together.
In many cases, the data is predominantly in proprietary forms as Excel, and statistical programs such as Stata and SPSS are used for the subsequent analysis.

4.4.1 International Center for Tropical Agriculture (CIAT)[61]

Cali, Columbia. English and Spanish.

Data policy: The plant genetic resources conserved by CIAT are a component of the world "designate collection" of the UN Food and Agriculture

54 http://irri.org.
55 http://dspace.irri.org:8080/dspace.
56 http://ring.ciard.net/dspace-irri.
57 http://www.worldfishcenter.org/wfcms/HQ/Default.aspx.
58 http://www.worldfishcenter.org/wfcms/HQ/article.aspx?ID=118.
59 http://www.iwmi.cgiar.org.
60 http://www.iwmi.cgiar.org/Publications/index.aspx.
61 http://www.ciat.cgiar.org.

Organization (FAO). Under a 1994 agreement with FAO, CIAT makes its germplasm available free of charge to farmers, farmer associations, breeders, agronomists, extension agencies, universities and Bioversity institutes with a clearly articulated need.

- **Database on plant genetic resources**
- **Product catalogue**
- **Online methods and query tools**

4.4.2 International Maize and Wheat Improvement Center (CIMMYT)[62]

Mexico City, Mexico. English and Spanish.

- **CIMMYT's Natural Resources Group (NRG) and Maize Program:** produced the *Africa Maize Research Atlas*, *Asia Maize Research Atlas*, and the *Latin American Maize Research Atlas*[63] a stand-alone, CD-ROM software that combines powerful and flexible GIS tools with preloaded data on climate, soils, elevation, population, land use and maize mega-environments for sub-Saharan Africa, southern Asia and Central and South America.
- **CIMMYT Socioeconomics Program (SEP):**[64] provides core data on agricultural prices and production through Open Access databases. Data are gathered from important global sources (World Bank, USDA, FAO, etc.) as well as from CIMMYT metadata projects. Information generated by CIMMYT includes descriptions of SEP projects and those from CIMMYT's former Economics Program and the Impacts, Targeting and Assessment Unit.

4.4.3 International Potato Center (CIP)[65]

Lima, Peru.

CIP databases:[66] are available online, which includes the following:

- **SINGER:**[67] (genetic resource collections) CGIAR genetic resources databases, including information on CIP's collection of potato, sweet potato and Andean root and tuber crops.

[62] http://www.cimmyt.org.
[63] http://www.cimmyt.org/en/services/geographic-information-systems/resources/maize-research-atlas.
[64] http://apps.cimmyt.org/agricdb/default.aspx.
[65] http://www.cipotato.org.
[66] http://cipotato.org/resources/databases.
[67] http://singer.cgiar.org.

- **World Potato Atlas**[68] and **World Sweetpotato Atlas:**[69] information about world potato production with emphasis on developing countries.
- **Inter-genebank Potato Database (IPD):**[70] a global database of potato germplasm available in the member gene banks.
- **World Potato Species Atlas:**[71] distribution maps of all currently recognised wild potato species.
- **DIVA GIS:**[72] tools (downloadable software), georeferenced databases and thematic maps related to genetic resource management.

4.4.4 International Center for Agricultural Research in the Dry Areas (ICARDA)[73]

Aleppo, Syrian Arab Republic.

- **Arid climate cereal and legume varieties:**[74] online data available on barley, bread and durum wheat, kabuli chickpea, lentil, faba bean, peas and forage legumes.

4.4.5 WorldFish Center[75]

Penang, Malaysia. English and native languages.

- **FishBase:**[76] online relational database with information on 28,500 species. Also available in CD-ROM format.
- **ReefBase:**[77] online free and easy access to data and information on the location, status, threats, monitoring and management of coral reef resources in over 100 countries and territories. Includes online GIS maps.
- **TrawlBase:**[78] a system for organising, storing, retrieving and exchanging a huge amount of data from past trawls in the seas of Asia.

68 https://research.cip.cgiar.org/confluence/display/wpa/Home.
69 https://research.cip.cgiar.org/confluence/display/wsa/Home.
70 https://research2.cip.cgiar.org/confluence/setup/setupdbchoice-start.action.
71 https://research.cip.cgiar.org/genebankdb/auto_2list.php?cmd=resetall&id=5.
72 https://research.cip.cgiar.org/confluence/display/divagis/Home.
73 http://www.icarda.cgiar.org.
74 http://www.icarda.cgiar.org/Crops_Varieties.htm.
75 http://www.worldfishcenter.org/wfcms/HQ/Default.aspx.
76 http://www.fishbase.org/search.php.
77 http://www.reefbase.org.
78 http://www.worldfishcenter.org/trawl/index.asp.

4.4.6 World Agroforestry Centre[79]

Nairobi, Kenya.

- **Agroforestree Database:**[80] a species reference and selection guide for agroforestry trees.
- **Useful Tree Species in Africa:**[81] this tool enables users to select useful tree species for planting anywhere in Africa using Google Earth.
- **Botanic Nomenclature:**[82] a compilation of the taxonomic status of over 8000 woody and herbaceous taxa found in agroforest ecosystems including synonyms and common names.
- **Tree Slides Database:**[83] allows to search the collection of agroforestry images.
- **Tree Seed Suppliers Database:**[84] lists suppliers of seeds and microsymbionts for over 5939 tree species. Also available on CD-ROM and in a book version.

4.4.7 International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)[85]

Andhra Pradesh, India.

- **SAT Electronic Library:**[86] an online service to CGIAR's scientific community and the partners from National Agricultural Research Systems (NARS). The SAT Electronic Library consolidates various resources and services available both in-house and on the internet. The various sections are SATSource Database, SRLS Database, SCIRUS Search, SWETSWise Searcher, agricultural sites on the web and full-text publications.
- **infoSAT:** electronic repository of reprints collected and preserved through the project SATCRIS (Semi-Arid Tropical Crops Information Service). While the full-text access to documents in the repository is restricted to ICRISAT researchers/partners, the access to metadata is open to all.

79 http://www.worldagroforestrycentre.org.
80 http://www.worldagroforestry.org/resources/databases/agroforestree.
81 http://www.worldagroforestry.org/our_products/databases/useful-tree-species-africa.
82 http://www.worldagroforestry.org/Sites-old/TreeDBS/Botanic.asp.
83 http://www.worldagroforestry.org/Sites-old/TreeDBS/slides.asp.
84 http://www.worldagroforestry.org/Sites-old/TreeDBS/tssd/treessd.htm.
85 http://www.icrisat.org.
86 http://www.elibrary.icrisat.org/welcome.htm.

4.4.8 International Water Management Institute (IWMI)[87]

English; documents also available in German, French, Dutch, Norwegian, Spanish, Portuguese, Danish, Swedish and Russian.

- **Climate Atlas Web Query (CAWQuer):**[88] service creates online climate summaries for user-specified locations. User registration required.
- **World Water and Climate Atlas:**[89] gives irrigation and agricultural planners rapid access to accurate data on climate and moisture availability for agriculture.
- **Eco-Hydrological Database:**[90] focuses on management of specific information pertaining to various aspects of freshwater ecosystems' functioning, requirements and management. User registration required.
- **Water Data Portal:**[91] an integrated portal providing a one-stop access to all data stored in IWMI's archive.
- **Global Irrigated Area Mapping (GIAM)** and **Global Map on Rainfed Cropland Areas (GMRCA)**[92]
- **African Transboundary Water Law:**[93] contains a searchable database of more than 150 different treaties, amendments and protocols which have been signed to manage the use of Africa's transboundary waters.

4.4.9 International Food Policy Research Institute (IFPRI)[94]

Washington, DC, USA.

Data policy: data is available online per request. Requested data sets are for the use of the requestor only and cannot be used by others without the permission of IFPRI. Proper citation is required; citation information is included with the documentation of each data set. IFPRI encourages the use of these data sets, but emphasises that many of them contain "raw" data files.

- **IFPRI data sets:**[95] browsing through the data sets is available on household, community and institution-level surveys and social accounting matrices, geospatial data, agricultural investment and expenditure

[87] http://www.iwmi.cgiar.org.
[88] http://www.iwmi.cgiar.org/WAtlas/AtlasQuery.htm.
[89] http://www.iwmi.cgiar.org/WAtlas/Default.aspx.
[90] http://dw.iwmi.org/ehdb/wetland/index.asp.
[91] http://waterdata.iwmi.org.
[92] http://www.iwmigiam.org/info/main/index.asp.
[93] http://www.africanwaterlaw.org.
[94] http://www.ifpri.org.
[95] http://ifpri.catalog.cgiar.org/datasetquery.htm.

and regional data. Data are indexed and available online per request (request form).[96]

- **IFPRI Knowledge Products:**[97] research tools, best practices and services which IFPRI shares as international public goods.
- **Agricultural Science and Technology Indicators (ASTI):**[98] initially managed by the International Service for National Agricultural Research (ISNAR) which has since then been taken over by IFPRI. The ASTI time series database[99] provides access to agricultural research and development indicators for developing countries in tabular format.

4.4.10 International Institute of Tropical Agriculture (IITA)[100]

Ibadan, Nigeria.

- **IITA Catalogues and Databases Directory:**[101] access to publications related to IITA's research.
- **Statistical Database:**[102] access to authoritative agriculture websites based on research related to that of IITA.
- **Genetic Resources Center:**[103] holds plant material (germplasm) of major food crops of Africa. Distributed without restriction for use in research for food and agriculture.

4.4.11 International Livestock Research Institute (ILRI)[104]

Nairobi, Kenya.

- **Library database:**[105] presents all publications which have been published by ILRI staff members, ILRI's corporate documents and extracted records from CABI and AGRIS based on ILRI's mandate.
- **Research outputs:**[106] ILRI institutional repository (Mahider) contains metadata and the link to the full content on an increasing proportion of ILRI's research outputs.
- **GIS data and services:**[107] all spatial data layers generated by ILRI are searchable and downloadable.

[96] http://www.ifpri.org/data/dataform.htm.
[97] http://www.ifpri.org/knowledge-products.
[98] http://www.asti.cgiar.org.
[99] http://www.asti.cgiar.org/timeseries.aspx.
[100] http://www.iita.org.
[101] http://www.iita.org/catalogsanddatabases.
[102] http://www.iita.org/web/iita/statistical-databases.
[103] http://www.iita.org/genetic-resources-center.
[104] http://www.ilri.org.
[105] http://ilri.catalog.cgiar.org/ilribsrc.htm.
[106] http://www.ilri.org/ResearchOutputs.
[107] http://192.156.137.110/gis/default.asp.

4.4.12 Bioversity International[108](formally known as the IPGRI)

Rome, Italy. English; documents also available in Chinese, French, Italian, Spanish, Portuguese and Russian.

- **New World Fruits Database:**[109] provides easier access to some basic, but often difficult to obtain, information on fruits from the New World. Links are provided to additional information, such as experts working on the different species, references and URLs, making the database a useful starting point in a search for more information on the selected species.
- **Species Compendium Database:**[110] a searchable database providing information at taxon level about seed survival during storage, germination requirements and dormancy, reproductive biology, pests and diseases.

4.4.13 International Rice Research Institute (IRRI)[111]

Los Banos, The Philippines.

Data policy: nonexclusive, nontransferable, limited license is granted to view and use information retrieved from their website site, provided solely for personal, informational, noncommercial purposes and provided that copyright notice or other notices are not removed or obscured. In no event shall materials from the website be stored in any information storage and retrieval system without prior written permission from IRRI.

- **Rice Knowledge Bank:**[112] the first comprehensive, digital rice-production library containing an ever-increasing wealth of information on training and rice production.
- **Rice bibliography:**[113] online search for all rice and rice-related articles.

4.4.14 Africa Rice Center (West Africa Rice Development Association)[114]

Cotonou, Benin.

---

108 http://www.bioversityinternational.org.
109 http://www.bioversityinternational.org/databases/new_world_fruits_database/search.html.
110 http://www.bioversityinternational.org/databases/species_compendium_database/index.html.
111 http://irri.org.
112 http://irri.org/knowledge/irri-training/knowledge-bank.
113 http://ricelib.irri.cgiar.org:81/screens/opacmenu.html.
114 http://www.warda.cgiar.org.

- **WAGIS:**[115] contains information on germplasm conserved in AfricaRice's gene bank, procedure to obtain seeds from AfricaRice using the Standard Material Transfer Agreement (SMTA) and many other data sets.
- **WAIVIS:**[116] The West Africa Inland Valley Information System contains databases on the agro-ecosystems of inland valleys in West Africa.
- **Additional data sets:** can be obtained upon request by email on:
  - breeding,
  - INGER Africa (varietal evaluation),
  - physiology, grain quality, drought, iron toxicity, photosynthesis,
  - soil fertility,
  - impact assessment.

4.5 Data portals

⋆ ⋆ ⋆ ⋆ Use URLs to identify things, so that people can point at your data

⋆ ⋆ ⋆ ⋆ ⋆ Link your data to other people's data to provide context

4.5.1 Global Bioversity Information Facility (GBIF)[117]

The GBIF makes digital biodiversity data openly and freely available on the internet for everyone and endorses both open source software and open data access. GBIF provides scientific biodiversity data for decision-making, research endeavours and public use. GBIF is a network of data publishers who retain ownership and control of the data they share. Linked data sets provide a more robust representation of biodiversity than any single data set. GBIF provides access to primary biodiversity data held in institutions in developed and developing countries. Data shared through GBIF are repatriated data. GBIF is a dynamic, growing partnership of countries, organisations, institutions and individuals working together to mobilise scientific biodiversity data.

Data use agreement The goals and principles of making biodiversity data openly and universally available have been defined in the Memorandum of Understanding on GBIF. The Participants who have signed the Memorandum of Understanding have expressed their willingness to make biodiversity data available through their nodes to foster scientific research development internationally and to support the public use of these data.

[115] http://africarice.org/wagis/default.asp.
[116] http://africarice.org/waivis/index.htm.
[117] http://data.gbif.org/welcome.htm.

GBIF data sharing should take place within a framework of due attribution. Therefore, using data available through the GBIF network requires agreeing with the following:

I Data use agreements

(a) The quality and completeness of data cannot be guaranteed. Users employ these data at their own risk.
(b) Users shall respect restrictions of access to sensitive data.
(c) In order to make attribution of use for owners of the data possible, the identifier of ownership of data must be retained with every data record.
(d) Users must publicly acknowledge, in conjunction with the use of the data, the data publishers whose biodiversity data they have used. Data publishers may require additional attribution of specific collections within their institution.
(e) Users must comply with additional terms and conditions of use set by the data publisher. Where these exist they will be available through the metadata associated with the data.

II Citing data Use the following format to cite data retrieved from the GBIF network:
Biodiversity occurrence data published by: (Accessed through GBIF Data Portal, data.gbif.org, YYYY-MM-DD)

III Definitions A series of definitions of data types and approaches are in the full agreement.[118]

4.5.2 Consortium for Spatial Information (CGIAR-CSI) data portal[119]

The CGIAR-CSI is a community of the many geospatial scientists within the CGIAR, linking the efforts of CGIAR scientists, national and international partners and others working to apply and advance geospatial science for international sustainable agriculture development, natural resource management, biodiversity conservation and poverty alleviation in developing countries. The CGIAR-CSI works to facilitate collaboration and capacity building for data sharing, data dissemination and geospatial analysis amongst the 15 CGIAR centers and the broader global research and development communities.

A number of data sets are made available by the CSI on climate, elevation, soil, poverty and others. A few samples:

– **WorldClim:**[120] a set of global climate layers (climate grids) with a spatial resolution of a square kilometer.

[118] http://data.gbif.org/terms.htm?forwardUrl=http3A2F2Fdata.gbif.org%2Fdatasets%2F.
[119] http://www.cgiar-csi.org/data.
[120] http://www.cgiar-csi.org/data/links-to-datasets/56-datasets/7-worldclim.

- **FutureClim:**[121] spatially downscaled climate projection data by Peter Jones (CIAT), Philip Thornton (ILRI) and Jens Heinke (PIK).
- **GADM:**[122] spatial database of the location of the world's administrative boundaries.

4.5.3 International Rice Research Institute (IRRI)[123]

- **World Rice Statistics (WRS):**[124] presents comprehensive time series information related to rice. Data on rice production, trade, consumption, inputs, prices and other related information are compiled from international and national statistical sources, personal communications and responses to questionnaires sent by IRRI's Social Sciences Division.
- **International Rice Information System:**[125] IRIS is the rice implementation of the International Crop Information System (ICIS) which is a database system that provides integrated management of global information on genetic resources and crop cultivars. This includes germplasm pedigrees, field evaluations, structural and functional genomic data (including links to external plant databases) and environmental (GIS) data.[126]

5 Challenges and opportunities

5.1 Challenges

Based on the review of data management within the CGIAR system in the previous sections, a number of challenges to providing data still remain and must be addressed.

Facilities and capacity The term "data curation" refers to a means of collecting, organising, validating and preserving data in such a way that researchers and stakeholders can make best use of the data over time. Data curation and data exchange facilities can be difficult – both technically and organisationally – to integrate into the workflows of research groups. For any system to be successful, however, it must focus on the users' needs, so these hurdles must be overcome and adjustments made.

121 http://www.cgiar-csi.org/data/links-to-datasets/56-datasets/8-futureclim.
122 http://www.cgiar-csi.org/data/links-to-datasets/56-datasets/9-gadm.
123 http://irri.org.
124 http://irri.org/world-rice-statistics.
125 http://irri.org/knowledge/tools/international-rice-information-system.
126 http://www.gosic.org/gtos/cgiar-data-access.htm.

To create a more widely used system of data curation, incentives (e.g. organisational credits or financial incentives) for data-curation efforts should be available to participating researchers. There should also be public recognition and commendation of those who openly publish and share data.

It is also important to keep in mind that, while data sharing is encouraged, there will always be instances where privacy is still warranted and, in fact, essential (e.g. the identities of survey participants must remain confidential).

Ownership and attribution One of the biggest challenges to making data publicly available is the management of intellectual property rights. The fear that there is a lack of control once data are released and a tendency to attribute sources incorrectly (or not at all) can often prevent data sharing entirely.

Similarly, there are complications when data-sharing systems include derived data sets and other components provided by third parties (e.g. satellite images or meteorological data). Derived data sets may be possible to share when source materials are not available.

In addition to clearly defining the property rights of a data-sharing system, an opportunity for those accessing the data to directly correspond with the researcher(s) who developed the data must be arranged. Since it is often not feasible to provide standalone data, there will be some need to set up correspondence with the original researcher(s).

Cost and duration The financial decision on expenditure in this area is complex. A cost–benefit analysis for data curation is not straightforward and a number of questions regarding what to keep and what to document must be answered. These financial decisions become increasingly complicated upon considering the finite nature of solutions to data-management problems. This brief lifespan is reduced even further because the data produced by CGIAR research changes rapidly. This means that work on updating data-management systems needs to be continuous, which is both time- and cost-consuming.

5.2 Opportunities

On the flipside of challenges, of course, are opportunities, and data sharing offers numerous opportunities to improve the way research findings are spread – and therefore used – throughout the world, which should motivate CGIAR researchers to participate in such activities. There are a variety of benefits to making research more available, accessible and applicable and researchers should be made aware of this.

Increased visibility Data curation and exchange can enhance the visibility of research and, thereby, the renown of its researcher. Thus, one of the benefits of data sharing is that the international community will come to associate a particular researcher (and his or her organisation) as an expert in a particular field. This incentive can boost interest in and resources for a particular project.

Improved access If data and information are more readily available and accessible to others in the field, it follows that studies will be more rigorously scrutinised and evaluated. This will result in more solid findings based on more reliable data.

Greater access to data can also improve partnerships and collaborative efforts by allowing all parties ownership of the material.

Greater impact By sharing data, information, findings and knowledge with more people – namely stakeholders – researchers and their organisations can make a greater contribution to the lives and livelihoods of the people they aim to serve. Often, drawing conclusions from a particular study and publishing within the academic and scientific communities does not translate into making a difference in the lives of farmers or families in developing countries. In order for research to be relevant, it needs to make its way to stakeholders in a format they can use. Data sharing and other interactive methods can help achieve this successful follow through.

6 Future directions and summary

The internet is an ever-evolving medium; as it changes, so too does the potential for its use as a research tool for and within developing countries.

6.1 Provision of data sets with publications

The demand for "the data behind the research" is strong and growing. Thus, there is a clear trend to systematically supply data sets alongside publications. While a certain lag-time still exists between document publication and data set release (in order to accommodate the need to adequately document the data sets), we are working to close that gap. But concerns still exist regarding "squeezing the last drop of usefulness" out of a data set before making it available and accessible and finding the resources to curate and maintain these data sets. Overall, however, the practice of including data sets with print publications serves to enrich the literature and enhance its value. It is important that we start considering data sets as an integral part of the global public goods we provide.

6.2 More work on interactive data visualisations

As software, server performance and bandwidths have improved, so has the ability to provide, within a website, the tools to allow a user to visualise data sets more readily and analyse a data set. The trailblazing work done by Gapminder – a nonprofit organisation that uses animated, interactive graphics to demonstrate statistical time series – has been adopted by some of the CGIAR centres that now use Google charts for visualisations and develop Flash-based visualisations to present data.

6.3 Publishing data in more interactive formats

Metadata is the foundation for information infrastructure and it is found throughout information technology systems: in service registries and repositories, web semantics, configuration management databases (CMDB), business service registries and application development. As the Semantic Web develops, more research organisations are producing linked data (e.g. linked data serialisations) that can be incorporated into linked applications.

In the past, the CGIAR had a strong commitment to metadata and indexed collections, so the challenge for the future is to open up these collections and integrate them in a way that allows their interoperability not only across the agricultural research community but also across the internet community as a whole. For this reason, linked data approaches, harvestable metadata and applications that use them both are clearly a necessity for the CGIAR in the future. We need to match this ambition with the skills sets and resources to deliver these systems.[127]

There are plans to expose data collections from the CGIAR as Linked Open Data and first experiments have been done. The Global Hunger Index (GHI) is available as Linked Open Data[128] and has been integrated into FAO's country profiles.[129] From this experience, a number of lessons have been learned that have been laid down in an online guide:[130]

[127] An example of an effort towards this is described in the blog post "Open Access Agriculture: opening the gates" at http://ictkm.cgiar.org/2010/10/27/open-access-agriculture-opening-the-gates

[128] http://data.ifpri.org/rdf/ghi.

[129] http://aims.fao.org/news/integrating-ifpris-linked-open-data-fao-country-profiles.

[130] http://linkedinfo.ikmemergent.net/content/global-hunger-index.

- Data that makes sense to the human reader is not necessarily sufficiently harmonised to be processed by other computer applications. A number of “cleaning steps” are required (steps 1, 2 and 3 in the online guide).
- Choose an appropriate data model and corresponding ontology for the data collection. The recipient applications that consume the Linked Open Data may have different requirements and the GHI is exposed using as ontologies two alternatives: “Scovo” and “Datacubes”. The important thing is that data provider (GHI) and data consumer (FAO country profile application) refer in the same way to the same things (steps 4 and 5 in the online guide).
- The Linked Open Data needs to be transferred to a server that can expose the data according to a protocol that the data consumer can handle. To encourage to take up the data it is essential to provide good documentation and examples how to use it (steps 6 and 7 in the online guide).

These steps are to degree technical but they require insight in the subject matter and in the way that others may want to use the data. These processes offer opportunities for researchers to communicate more closely with potential beneficiaries of their work beyond the circle of their direct peers.

6.4 Application programming interfaces

Data need to be machine readable so that they can be processed remotely by applications, combined and analysed online. This would be particularly advantageous for time series data in which an initial setup for reading information could be repeated when the data are refreshed.

6.5 Knowledge sharing

As the CGIAR moves ahead with its change process, it is continuously being told that it needs to do a better job at sharing its vast wealth of research-generated knowledge, so that this knowledge can be applied to solving real problems. While written publications – the major output of most projects – are a key source of high-quality information, they are not often widely available or accessible. In fact, for the majority of stakeholders working in agricultural development, they are not even applicable to them. Therefore we need to do a better job sharing our agricultural data, information and knowledge in ways that make them available, accessible and applicable.

Since we have already invested in knowledge sharing and innovative approaches to achieve it, the challenge becomes keeping the momentum going, which begs the question of how to maintain that energy. Universally, there are two types of motivators: positive reinforcement for a job well done and negative reinforcement for the opposite. The anecdotal story of dangling a carrot in front of donkey to entice it to move forward or using a stick to coerce him provides an interesting framework for understanding what motivates behavioural change. Based on the story, the carrot – an appealing treat for the donkey if it moves forward – represents incentives or rewards; the stick – an undesirable repercussion should the donkey refuse to move – represents mandates, policies, enforcements and punishments.

So, which works better when it comes to facilitating better knowledge sharing between researchers and research institutes? The carrot-like incentives or the stick-like mandates and enforcements?

The answer, at this point, is that most people are quick to respond to the question of how to share; they start throwing around resources, tools and methods that can help capture, store and provide access to knowledge. However, despite these active responses, many of these knowledge-sharing approaches are not actually being used widely or comprehensively. So perhaps we need to go back to some more fundamental questions: Why is knowledge sharing important in the first place? What does it aim to achieve?

Why should we share our knowledge? Is it not, after all, the vital capital of a researcher or research institution? Is it our role to share it? Do we have the capacity to do this sharing? What do we (and our institutes) gain by engaging in this time- and resource-consuming work? These are the remarks we often hear on the subject of knowledge sharing and Open Access. And the last question – What do we gain? – is particularly important. In order for researchers and research institutes to carry out knowledge sharing activities, there must be some benefit. This can range from greater visibility, improved fundraising potential, enhanced partnerships or better contribution to impact.

The carrot: incentives and rewards Much has been discussed on the subject of incentivising knowledge sharing; in fact, it has often been proposed that the CGIAR's performance evaluation mechanisms be redesigned to reward staff for going beyond publication requirements to get the research information out to various stakeholders. If researchers go the extra mile to organise workshops, build capacity or disseminate information via radio programmes, for example, how should they be recognised or compensated? In the CGIAR, we are continuously exploring, testing, documenting and celebrating new ways of sharing knowledge so that it can keep growing.

The stick: mandates or enforcements In some cases, organisations and institutions have developed policies and mandates to enforce certain actions amongst staff. These policies and mandates require staff to do particular things in a certain way and staff are evaluated accordingly and can be rewarded or punished if those activities are not carried out. This way of bringing about change has been found by some to be more effective in providing a wider stimulus for change. It also brings a certain consistency to the change desired. For example, CGIAR research staff are required to publish a certain number of journal articles per year, especially in ISI-ranked journals. There is also a reward system enforced by the overall CG system for these centres to increase their publication-per-researcher ratio. Which works better?

6.6 Promotion

The ASTI case study highlighted in this report (see 1.2 Case study) points to the importance of not only collecting and curating data, and subsequently making it available online, but also the necessity of promoting it through different channels. In particular, the study highlights the project's success in attracting media interest through events and press briefings. Therefore, in addition to using online communications, it is key to develop and use the variety of media available, such as print, visual design and face-to-face products, to reach targeted audiences.

While conventional methods for sharing data, information and knowledge, such as conferences, seminars, journal articles and reports, along with the more recent use of institutional repositories, play an important role in the communication of research and development, the way people search for information has been changing, especially in certain countries and amongst certain demographics. Social media has been growing in importance and steadily breaking down barriers to communication, allowing people to connect, engage and share in a more informal way. Social media tools currently used by the CGIAR include blogs, wikis and podcasts, and services such as Facebook, LinkedIn, Twitter and Flickr.

"Social media" has two main components: "social" and "media". Media tools can bring content to a much wider audience and at a much faster rate than previously thought possible. Since the way people search for information is evolving, the tools used must be adapted accordingly. Information overload is now a major concern, with many people not wanting to spend time visiting a website, blog, database or any other resource unless someone they trust points them in that direction. Through social media channels, it is possible to seek out recommendations and suggestions from colleagues, peers and experts.

This has implications for the way research and development organisations communicate. While opening access to information is widely regarded as important, simply pushing it out to target audiences does not guarantee that it will be read and used. Information is useful only when it is received and read by the right person at the right time. Social media tools can help get an organisation's messages to the right people.

As new media tools are based on being social, users can build a community around their content or particular channel, and this is where social media's true value lies. These communities, or social networks as they are called, are formed around a common interest or shared purpose, and often result in an environment based on trust that facilitates effective collaboration and sharing. This enables a natural flow of peer review and feedback and also enhances transparency. But it is imperative that social media also be used for much more than forming communities or reaching out to those with shared interests. Social media needs to be explored, understood and harnessed to make knowledge available, accessible and applicable to the wide array of audiences participating in the social media arena.

Social media allows for and, in many cases, insists on much more continuous and less formal communication activities. The communication of research usually takes place at a project's conclusion, but social media can facilitate the sharing of ideas, experiences and knowledge throughout the whole research cycle. Social media can give access to the inner workings of research activities, enabling a variety of audiences to learn from them. For example, audiences can read about a project's progress on a project blog, see actual evidence of research activities in the form of Flickr pictures and hear and see testimonials in YouTube videos. Having access to the knowledge being generated throughout the research cycle is not only important to donors, but it can also benefit other target audiences of the research.

Moreover, social media allows audiences to participate and provide almost real-time feedback. By considering an audience's needs and questions, it is possible to keep the research grounded and facilitate better partnerships (as opposed to just having research recipients), both of which enhance the sharing of knowledge, in terms of its use and impact. As such, social media can increase stakeholder involvement and enhance a project's accountability and the overall achievement of its objectives.

Social media can help organisations to reach and involve a wide range of necessary stakeholders such as donors, other research communities, implementing partners, the general public and even the intended beneficiaries, such as farmers. Since social media tools are freely available and internet access is becoming more available throughout developing countries, even in rural areas, research knowledge is now being more readily picked up and demanded directly by farmers. For example, three posts on the new CGIAR Consortium blog received comments from farmers located in rural areas of developing countries asking for more information about getting access to the seeds that one of the posts talked about, among other things. Social media channels can also play a brokering role by connecting people to others who have the knowledge they need, such as linking farmers with extension agents.

Social media, therefore, has the potential to make research outputs much more available, accessible and applicable. It can increase the visibility of, participation in and adaptation of research knowledge. Compatibility among different social media tools provides an added dimension of connectivity so that social networks can be interlinked, creating an audience base that has the potential to expand rapidly.

6.7 Collaborative efforts

The CGIAR system, like many organisations, is adept at generating information. However, the challenge is in knowing how to extract and manage the knowledge buried within the volume of information being produced and then being able to apply that knowledge to emerging needs.

Knowledge sharing is a way of putting information, communities, processes, and tools together to allow the CG to collaborate more effectively and make better decisions. Tools and technologies by themselves cannot ensure successful partnerships, collaboration or teamwork, nor make the CGIAR work as a system; they are necessary but not sufficient. And, while the tools and technologies can contribute to improvements in personal and organisational performance, significant gains require changes in organisational culture and individual behaviour.

People, and the tacit knowledge they have, are central. It is through greater understanding and support to the cross-functional communities that organisational culture can shift towards one of ongoing learning and collaborative sharing of knowledge and expertise – a CGIAR without boundaries. It is now realised that the best knowledge-transfer technique is face-to-face interaction and that the best knowledge repository is a community or group of people, supported by a technology solution.

The CGIAR is committed to strengthening incentive systems that promote knowledge-sharing practices and for communities of scientists to improve the way they work.

6.8 Ubiquitous telecommunications infrastructure

Thanks to the falling costs of all things digital, there has been a steady flow of investment into communications infrastructure around the world. Cell phone networks carrying voice and internet data are being deployed in even the poorest countries and with time will expand to cover most rural areas. These wireless networks are sophisticated and easily managed. Multipurpose public networks will be offered by private telecomm companies and governmental agencies, while self-organising device networks (such as ZigBee, a low-cost, low-power, wireless mesh networking standard) can be installed with minimal planning or oversight. Agriculture and agricultural research can increasingly take communications capacity for granted in the years ahead.

This new infrastructure will enable new applications of communications to both the gathering and dissemination of information by agricultural researchers and practitioners. First, for gathering information, the historical and remotely sensed data that has been gathered to date can be complemented by near-real time, ground-based data. Sensors can transmit the information they detect through increasingly ubiquitous wireless data networks into internet-based servers. Radio-frequency identification (RFID) tags can be attached to vehicles, buildings and selected goods; combined with Geographic Positioning System (GPS) information, objects can be automatically tracked and even audited in real time. The result will be both real-time interpretation of current conditions and longitudinal analyses that reflect up-to-date information.

The costs of the sensors, tags, GPS and RFID devices and the communications between them are dropping so rapidly that new data-gathering applications can be expected to proliferate in the near term. Here are some relevant examples:

- sensors and cameras in fields or on farm equipment,
- sensors of water levels in irrigation or in soils,
- sensors in food storage,
- early detection of pests,
- emissions sensors,
- tagging of livestock,
- tagging of other natural resources,
- tagging trucks and shipping containers,
- market, banking and distribution data.

Like satellite imagery, these new types of data will require considerable processing to ensure their quality and consistency and to make them comparable from one location to the next. The research community will need to establish processes for validation and distribution of these data, as they have with other public information goods.

The same networks that collect and carry sensor data will also be used to disseminate information into rural areas. Cell phones are already being used at an increasing rate by rural residents. For them, the value of communication is high, and there are many ways to effectively share the fixed costs of phone devices and electric power among numbers of users. As phones get larger screens, touch interfaces and voice recognition, and as new classes of inexpensive and rugged “netbooks” are developed, many new opportunities for agricultural extension will arise. It will start by providing today's information to new audiences. It will grow into provision of new services that are more localised and more up-to-date, building on the data gathering that is enabled by the networks that feed these devices. With new audiences and new services will come new requirements for assuring the quality of the information provided.

6.9 Cloud computing

The combination of progress in system software, computing hardware and internet communications has now enabled the construction of general-purpose data centres that can be reconfigured by command to support any software application in minutes ("virtualisation" software was the key innovation).

There are already data services that allow a user to have many hundreds of computers at their command, and yet pay for them by the hour or minute, without owning or operating the hardware themselves. The costs are far less than even falling hardware prices would suggest, since the cost of the data centre can be shared among many "bursty" users. In effect, the data centre acts like a utility, providing as much computing as requested at just the times when needed. Since these data centres are invariably shared over the internet, they are sometimes called computing "in the Cloud," giving rise to the common term "Cloud computing".

Many observers believe that Cloud computing will soon be the lowest-cost option for nearly all types of data centre computing. Cloud providers are already more cost-effective for "bursty" high-performance computing, like video and image processing, bioinformatics and most types of scientific data analysis. We can expect research centres in agriculture to have accounts on several Cloud providers and to select them at different times for different purposes.

The shift to Cloud computing is a good thing for today's researchers, by cutting the total cost of scientific computing. But it also brings two new opportunities for international agriculture. First, it completely separates the utilisation from the operation of computing facilities. In other words, users of data centres no longer need the capacity to procure and operate them. As long as one has a browser on the internet, one can "order up" essentially any computer software at any scale and pay only for what is used. As a result, many more organisations will be able to take advantage of large-scale advanced computing.

A second implication of Cloud computing is an increased impetus to share data among researchers. It is a common pattern today to move large data sets, such as satellite images or longitudinal data sets, from one data centre to another for use in different projects. The transfers add delay and can be error prone. By contrast, Cloud data centres are a natural repository for public information goods such as shared data sets, so that users in any location or institution can instantly access, analyse and interpret data without the need to move it to their own facilities. This reduces the need for high-speed or high-capacity network connections, since much less data moves between the users and the source of the data. A researcher with a moderate-speed connection to the internet can work with data as well as other researchers regardless of location. In addition, researchers will normally leave the results of a Cloud analysis at the Cloud data centre, allowing potential re-use by others. Properly managed, this can enable new kinds of collaboration and project organisation.

Implications Leading institutions in agricultural research have an opportunity to flesh out these possibilities today and thereby create templates for future models of progress. Here is one illustration.

A research centre that works with a crop could choose a group of similar varieties that have been cultivated for a long period at one of their facilities. The centre will already have basic long-term data across many seasons, along with much bioinformatic data. These data could be stored in one or more Cloud computing facilities and could be supplemented from now on by extensive sensor data, collected and made available in near real-time from the fields where the varieties are cultivated. In effect, these fields become a "bio-observatory" for those varieties. In addition, one or more regions where those varieties are currently cultivated by farmers could also be instrumented with some sensors, and the markets in those regions could employ tags or other methods for continuous data collection. Once a data collection like this is available in a Cloud data centre, a series of analytical studies could be commissioned at various developing country institutions around the world.

These institutions would be chosen for having familiarity with the varieties but currently lacking the facilities to do their own extensive data analysis or interpretation. In addition, adaptation and extension projects could be commissioned at additional national organisations to produce materials for delivery into the areas where these varieties are grown.

Like any collaborative research, this kind of project would have to confront issues of data harmonisation, accessibility and ownership. Part of the value of this project would be the demonstration of solutions to these issues, as a pattern for future projects to follow.

Naturally, this entire scenario could be adapted in many ways to the other agricultural research topics. For example, a project could treat livestock instead of crops or could extend a system like Fishbase[131] for a class of fish. Genetic studies of crop pathogens, patterns of water supply and utilisation in a watershed, forest growth and production patterns: all lend themselves to this sort of project. There will be limitations to the effectiveness of any single project, but the first projects are likely to provide key lessons to light the way for the research community in utilising the next wave of technological changes.

6.10 Increased use of spatial analysis and GIS

As more data sets become available, the opportunities to compare agricultural data across similar agro ecological zones leads to an increased need to accurately geo-locate data. This has lead to an increasing use of geographical information systems and spatial analysis. One example of this is presented in the case studies. We see this trend developing in the future with increased interest in spatial analysis and the development of new models and tools in this area. Agriculture is inextricably tied to the physical environment and the unpredictability of nature. Factors such as climate, soil and water availability play more of a defining role in agriculture than in any other economic sector. And nowhere is this more evident than in Africa. If smallholder farmers are to be consistently successful, from one season to another, and from one year to another, they need to have access to essential geospatial (location-specific) information.

[131] An online database with information on 28,500 species developed by the WorldFish Center http://www.fishbase.org.

7 List of figures

8 List of tables

9 Bibliography

ASTI. *ASTI Toolkit. Monitoring Agricultural R&D. Capacity and Investment Indicators: A Practitioner's Guide to ASTI's Methodologies & Data Collection Standards*. Washington, DC, International Food Policy Research Institute, 2011.

Alercia, A, Diulgheroff, S, & Metz, T. FAO/IPGRI Multi-crop Passport Descriptors. Rome, Bioversity, 2001. http://www.bioversityinternational.org/fileadmin/bioversity/publications/pdfs/124.pdf?cache=1318460131.

Arivananthan, M, Ballantyne, P, & Porcari, E. Benchmarking CGIAR outputs for availability and accessibility. Agricultural Information Worldwide 2010, 3(1), 17–22. http://journals.sfu.ca/iaald/index.php/aginfo/article/view/35/51.

Norton, G. W. (2010). Impact Assessment of the IFPRI Agricultural Science and Technology Indicators (ASTI) Project. Washington, DC, International Food Policy Research Institute, 2010. http://www.ifpri.org/sites/default/files/publications/ia32.pdf.

Vargas, R. (2010). Breaking the Norm: An Empirical Investigation into the Unravelling of Good Behaviour (no. 00948). Washington, DC, International Food Policy Research Institute, 2010. http://www.ifpri.org/publication/breaking-norm.

Vargas, R, & Viceisza, A. An Experiment on the Impact of Weather Shocks and Insurance on Risky Investment (no. 01100). Washington, DC, International Food Policy Research Institute, 2010. http://www.ifpri.org/publication/experiment-impact-weather-shocks-and-insurance-risky-investment.

10 Further reading

CIAGR blog The CGIAR has promoted a number of knowledge-sharing approaches through a series of blog articles, which it sees as a useful vehicle to continue these developments in the future.

Create awareness by raising the profile of your organisation on social networking sites. Cultivate long-term support for your organisation by creating your own network of scientists, research partners and interested individuals. http://ictkm.wordpress.com/2009/05/06/social-networks-friend-or-foe. Use social media tools to promote your projects, events and activities. Announce time-sensitive, newsworthy items by microblogging. Microblogging involves posting short sentences (max. 140 characters) that can be used to promote your journal article or a useful website, act as a reminder for an activity or even ask questions. http://ictkm.wordpress.com/2009/04/02/microblogging. Promote your name. Use social media to establish your reputation in the research and development arena. Blogging is a good way for researchers to share their research ideas with others and gain feedback from a wider, online audience. Well-thought-out blogs attract people with similar thoughts and queries, people who can validate your ideas and also challenge you by sharing varying opinions. http://ictkm.wordpress.

com/2009/04/23/blogging-for-impact. There are many ways you can engage with others and share knowledge using social networking sites.

Engaging people

- Promote issues that resonate with people to encourage involvement and gather support for your cause.
- Form strategic alliances with influential people and institutions that help boost your organisation's profile.
- Bring together expertise and talent, whether potential research partners, service providers or other experts.

Sharine knowledge Social media transcends geographic boundaries. Test your research ideas by sharing them with your colleagues globally. You can collaborate at a fraction of the time and cost associated with face-to-face meetings. Collaborative sharing sites also come with security options that allow secure knowledge sharing. http://ictkm.cgiar.org/2009/05/29/wikis-sites-docs-and-pads-the-many-flavours-of-collaborative-writing. Create an environment where people recognise your expertise and you can establish your organisation as the expert in your field of research. Whether you are a researcher who is new to a field and eager to learn more, or the resident expert, share your knowledge and experiences by contributing to insightful blogs.

- Communicate your research outputs better by adjusting your content to fit different social media tools. Think of social media as strategic communication lines that branch outward to several different networks, which in turn branch into other networks.
- Reach out to interested people outside your regular circle and gain valuable ideas/feedback from your pool of social networks. Pay attention to conversations that are already ongoing on social media sites. Sharing is a two-way process, and you should take the time to interact with others in a similar fashion.

Share resources within interested communities. Social bookmarks and news feeds are great online organisation tools that keep track of what's being published on useful websites and blogs you frequent. Share this with others and then see the favour being returned. http://ictkm.wordpress.com/2009/05/18/social-bookmarking-storm-brewing. http://ictkm.wordpress.com/2009/06/19/newsfeeds-delivering-the-latest-news-to-your-virtual-doorstep.

11 Glossary

Cloud computing

A shared Cloud data centre will typically have over 1000 computers, which can support at least 100,000 user "virtual" computers. This is super-computer scale by any standard, so most research centres will not own one but rather will share one with hundreds of other customers. Commercial Cloud providers, such as Amazon, Google and Microsoft, already offer services and some government-run research Clouds exist. Shared by many thousands of customers, these are extremely cost-efficient. They employ a relatively small staff of system managers, keep a low budget for electric power, can survive routine equipment failure without service interruption and adopt continuous modular upgrades of new types of hardware. There are choices in many countries, which allow for flexibility where there are legal restrictions.

Social media

The following definition is given by Wikipedia (http://en.wikipedia.org/wiki/Social_media): "Social media is online content created by people using highly accessible and scalable publishing technologies. Social media is a shift in how people discover, read and share news, information and content; it supports the human need for social interaction with technology, transforming broadcast media monologues (one to many) into social media dialogues (many to many). It supports the democratisation of knowledge and information, transforming people from content consumers into content producers. Social media has become extremely popular because it allows people to connect in the online world to form relationships for personal, political and business use."