CGIAR Open Access and Open Data – Overview

Medha Devare
(m.devare@cgiar.org)

Arid Cereals, Minneapolis
November 14, 2015
CGIAR R4D Centers

CGIAR is a global research partnership for a food secure future
Why OA/OD??

CGIAR’s decades-rich data trove – availability for value addition via re-use, meta-analyses, decision support...??

• Enhance cross-regional, cross-disciplinary learning
  – Multiple CG centers, partners, countries, hubs
  – Varied data streams (breeding, agronomy...)

• Facilitate internal/external monitoring and evaluation

• Cement institutional memory

• Increase efficiency, RoI: Donor, public, scientist...

• CGIAR outputs are public goods

IMPACT...not impact factor!
Recurrent OA/OD-related themes

- Need to validate with data...
- Sustained info services...
- Big data and ICTs as an opportunity...
- Variety of data needed, in context, aggregated...
- Access to high quality, plug ‘n play data...
### Big Data Scale

<table>
<thead>
<tr>
<th>Little Data</th>
<th>Little Data Shared In a BIG DATA base</th>
<th>Big Data Aggregated, Co-created</th>
<th>Big Data De-personalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
<td>my data</td>
<td>data shared</td>
<td>Aggregated co-created and connectable back to user</td>
</tr>
<tr>
<td>Bob</td>
<td>my data</td>
<td>data shared</td>
<td>Aggregated or combined from multiple sources, de-personalized, but possibly connectable back to user</td>
</tr>
<tr>
<td>Sheri</td>
<td>my data</td>
<td>data shared</td>
<td></td>
</tr>
</tbody>
</table>
BIG DATA HIGHWAY?
CGIAR OA-OD: Overview

**Policy**
- Final policy available at [www.cgiar.org/open](http://www.cgiar.org/open)
- Approved by Consortium Board and all Centers in 2013.

**Funding**
- Obtained $2.35M (BMGF) in Jan 2015.
- Guidelines and support pack developed with Center input.

**Documentation, Monitoring & Evaluation**
- Initial assessment done; in-depth evaluation begun.

**Partner engagement**
- In progress. Working with Centers, FAO, AgMIP, GODAN, WUR, Divseek, private sector...

**Center/CRP OA-OD assessment**
- In progress. Template developed to reduce Center/CRP burden and harmonize.

**Center/CRP Implementation Plans**
- In progress. DMTF, Regional WS, OAWG...

**Center/CRP Implementation**
- In progress. Dspace, Dataverse; harvester use cases, specs developed.

**CG-level metadata, repositories, harvester**
- In progress. Working with Centers, FAO, AgMIP, GODAN, WUR, Divseek, private sector...

**Funding**
- Done. Guidelines and support pack developed with Center input.

*Final policy available at [www.cgiar.org/open](http://www.cgiar.org/open)*
What do we need to accomplish?

Get from this...


To this:

Which requires access to data sets and harmonization on...

standards

metadata

tools/platforms

interoperability

incentives/culture

CGIAR is a global research partnership for a food secure future
Approaches to repository systems

• Biomedical
  → centralized
    @ NCBI PubMed
  → Only ~4% deposition into PubMed until mandated

• Ecology → LTER
• Agriculture → ???
• CGIAR → different approaches, systems

Repercussions for user (FAIR)?

Use case: Joe Schmoe’s team needs 20 years of wheat productivity data for central Asia…

• Where to find?
• Metadata? Annotations?
• Quality controlled?
The bigger picture...

Zea mays cultivar B73 chromosome 4

GenBank: CM000780.3

FASTA  Graphics

LOCUS CM000780 242009974 bp DNA linear CON 24-OCT-2013

DEFINITION Zea mays cultivar B73 chromosome 4.

ACCESSION CM000780 JH967973 JH967981 JH967982

VERSION CM000780.3 GI:552562410

ORIGIN BioProject: PRJNA10769

KEYWORDS .

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAMAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 242009974)

AUTHORS Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F.,
Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P.,
Reilly, A.D., Courtney, L., Kruhowski, S.S., Tomlinson, C., Strong, C.,
Dolehauty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E.,
Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P.,
Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L.,
Higginbotham, J., Cardenas, M., Waligator, J., Applebaum, E.,
Phefys, L., Falcon, J., Kanchi, K., Thane, T., Scimone, A., Thane, N.,
Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J.,
Ingentron, E., Cordes, M., Kohlbeg, S., Sgro, J., Delgado, B.,
Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K.,
Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T.,
Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotzki, M., Lopez, G.,
Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S.,
Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C.,
Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T.,
Miller, B., Amboise, C., Muller, S., Spooner, W., Narechiano, A.,
Ren, L., Wei, S., Kumaris, S., Faga, B., Levy, M.J., McMahan, L., Van
Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y.
The bigger picture...

Search databases

<table>
<thead>
<tr>
<th>Literature</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Books</td>
<td>28</td>
<td>books and reports</td>
</tr>
<tr>
<td>Policy briefs</td>
<td>0</td>
<td>policy briefs</td>
</tr>
<tr>
<td>Training, extension</td>
<td>0</td>
<td>books, journals, and more</td>
</tr>
<tr>
<td>AgPub</td>
<td>170</td>
<td>abstracts and citations</td>
</tr>
<tr>
<td>AgPub Central</td>
<td>4,580</td>
<td>full-text journal articles</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geographies</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>30</td>
<td>resources for sub-regions and countries in Asia</td>
</tr>
<tr>
<td>Africa</td>
<td>82</td>
<td>resources for sub-regions and countries in Africa</td>
</tr>
<tr>
<td>Middle East</td>
<td>11</td>
<td>resources for countries in the Middle East</td>
</tr>
<tr>
<td>Europe</td>
<td>13</td>
<td>resources for sub-regions and countries in Europe</td>
</tr>
<tr>
<td>N. America</td>
<td>17</td>
<td>resources for sub-regions and countries in N. America</td>
</tr>
<tr>
<td>S. America</td>
<td>25</td>
<td>resources for sub-regions and countries in S. America</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Toolkit</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology tracker</td>
<td>0</td>
<td>technology adoption tracking</td>
</tr>
<tr>
<td>Agri-semantics</td>
<td>5</td>
<td>ontologies, vocabularies</td>
</tr>
<tr>
<td>AMKN</td>
<td>0</td>
<td>climate change adaptation and mitigation knowledge network</td>
</tr>
<tr>
<td>Activity mapper</td>
<td>2</td>
<td>project and activity mapping</td>
</tr>
<tr>
<td>Methods</td>
<td>42</td>
<td>methodologies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subjects</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agroforestry</td>
<td>3</td>
<td>resources related to agroforestry</td>
</tr>
<tr>
<td>Agronomy</td>
<td>22</td>
<td>resources related to agronomy</td>
</tr>
<tr>
<td>Aquaculture/fisheries</td>
<td>0</td>
<td>resources related to aquaculture/fisheries</td>
</tr>
<tr>
<td>Climate change</td>
<td>155</td>
<td>resources related to climate change</td>
</tr>
<tr>
<td>GIS/remote sensing</td>
<td>33</td>
<td>resources related to GIS/remote sensing</td>
</tr>
<tr>
<td>Genebank</td>
<td>53</td>
<td>genebank resources</td>
</tr>
<tr>
<td>Genetic/genomic</td>
<td>5</td>
<td>genetic/genomic resources</td>
</tr>
<tr>
<td>Hydrology/water mgmt.</td>
<td>0</td>
<td>resources related to hydrology/water management</td>
</tr>
<tr>
<td>Livestock/animal breeding</td>
<td>120</td>
<td>resources related to livestock/animal breeding</td>
</tr>
<tr>
<td>Natural resource mgmt.</td>
<td>60</td>
<td>resources related to natural resource management</td>
</tr>
<tr>
<td>Plant breeding</td>
<td>134</td>
<td>resources related to plant breeding</td>
</tr>
<tr>
<td>Plant protection</td>
<td>13</td>
<td>resources related to plant protection</td>
</tr>
<tr>
<td>Socioeconomics/livelihoods</td>
<td>47</td>
<td>resources related to socioeconomics/livelihoods</td>
</tr>
<tr>
<td>Other</td>
<td>56</td>
<td>resources related to other subjects</td>
</tr>
</tbody>
</table>

Results found in 17 databases for "drought tolerance maize"
Repositories: standards, protocols
Core: metadata
Interoperability:
Vocabularies: GACS (AGROVOC-NAL-CAB)
Ontologies:
- crop
- agronom
- agritech
- livestock
- value chain
- agriVIVO
Harmonized data sets:
BMS
Agronomy field book
ICASA – MD data dictionary
ontology interlinkages
Geospatial:
consistency, quality

Infrastructure:
Analytics/tools:
- Technology mapping - foresight
- Investment mapping - foresight
- M+E
- Research discovery
- Decision support

Aligned CGIAR plans, approaches:
- Overall: implementation plans
- HR: evaluations, KPIs
- HR: contractual language, induction process
- Legal: CC licensing
- Leadership: overall buy-in
- DM/KM: infrastr, support, guidance, messages
- ICT: infrastructure, support
- Partnerships: FAO, CABI, GODAN, AgMIP, Alterra-WUR, DivSeek, INRA, CIRAD...

OA-OD capacity, buy-in:
- CGIAR++: overall buy-in
- Partners
- Govts

OA-OD governance/org...

Definitions (what is OA?), consequences (e.g. 80-20 funds)
Data mgmt. plans, implementation templates, indicators publishing/licensing guidelines
Institutional or community interoperable repositories
OA-OD governance/org...

USAID, Gates, DFID, ACIAR...
Aligned CG-donor policies, guidelines
HR: contractual language, induction process
Legal: CC licensing
Leadership: overall buy-in
DM/KM: infrastr, support, guidance, messages
ICT: infrastructure, support
Partnerships: FAO, CABI, GODAN, AgMIP, Alterra-WUR, DivSeek, INRA, CIRAD...

OA-OD capacity, buy-in:
CGIAR++: overall buy-in
Partners
Govts

OA-OD governance/org...

begun, may be continued in phase II
phase II
Thanks
General approach to R4D

Discovery
- New concept of product, service, or process

Proof of concept
- Testing of proof of concept in real world / controlled conditions (n=1000s)

Pilot
- Multi-location release/trials for smallholder’s benefit (n=100,000s)

Scaling up
- Release for scaling up & adoption in different locations (n= 1,000,000s)

Science partnerships

Development partnerships

Adapted from: P. Ellul

DATA??
CGIAR OA-OD project: Key objectives

1. Conduct a broad inventory and assessment of CGIAR capacity in OA-OD.

Output 1.1. In-depth analysis of the publications landscape across CGIAR. → in progress

Output 1.2. Needs analysis of CGIAR data management and data quality practices over the data life cycle. → in progress

Output 1.3. Analysis of other research products across CGIAR. → in progress

Output 1.4. Identification of gaps in CGIAR human infrastructure and enabling environment for OA-OD. → in progress

Output 1.5. Proof-of-concept indexing tool/portal. → exploration
CGIAR OA-OD proposal: Key objectives

2. Develop a legacy data prioritization framework.

   Output 2.1. Data prioritization framework.
CGIAR OA-OD project: Key objectives

3. Provide coordinated support to Centers and CRPs in their efforts towards OA-OD, and leadership for external efforts.

Output 3.1. Support pack for Open Access and Open Data.
→ Support pack v.1 in place

Output 3.2. Consistent OA-OD implementation plans.
→ in progress

Output 3.3. Improved, interlinked initiatives, tools, and platforms.
→ Ongoing: Agronomy Ontology, AgTrials, IBP-AMS, AATP-VIP

Output 3.4. OA-OD collaboration and leadership beyond CGIAR.
→ AgMIP, agrisemantics harmonization (FAO, CABI, NAL+), WUR, INRA, CIRAD, GODAN...
CGIAR OA-OD proposal: Key objectives

   
   Output 4.1. OA-OD impact assessment framework.

5. Plan for Phase 2: Implementation.
   
   Output 5.1. Phase II funding proposal.
3.4 Data storage and preservation

- Should raw data be preserved, or processed/normalized/transformed data?
- Description of data to make it relevant in the future?
- Where will data be preserved? Is that location stable?
- Security mechanisms, backup procedures?
- Which file formats are best for long-term preservation?
# Data storage and preservation – file formats

<table>
<thead>
<tr>
<th>Type of data</th>
<th>Acceptable formats for sharing, reuse and preservation</th>
<th>Other acceptable formats for data preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative tabular data with extensive metadata</td>
<td>SPSS portable format (.por) delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information some structured text or mark-up file containing metadata information, e.g. DDI XML file</td>
<td>proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta) MS Access (.mdb/.accdb)</td>
</tr>
<tr>
<td>Quantitative tabular data with minimal metadata</td>
<td>comma-separated values (CSV) file (.csv) tab-delimited file (.tab) including delimited text of given character set with SQL data definition statements where appropriate</td>
<td>delimited text of given character set - only characters not present in the data should be used as delimiters (.txt) widely-used formats, e.g. MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods)</td>
</tr>
<tr>
<td>Geospatial data vector and raster data</td>
<td>ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn) geo-referenced TIFF (.tif, .tfw) CAD data (.dwg) tabular GIS attribute data</td>
<td>ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (KML) (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg), binary formats of GIS and CAD packages</td>
</tr>
<tr>
<td>Qualitative data textual</td>
<td>eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml) Rich Text Format (.rtf) plain text data, ASCII (.txt)</td>
<td>Hypertext Mark-up Language (HTML) (.html) widely-used proprietary formats, e.g. MS Word (.doc/.docx) some proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti</td>
</tr>
</tbody>
</table>
### Data storage and preservation – file formats

<table>
<thead>
<tr>
<th>Digital image data</th>
<th>TIFF version 6 uncompressed (.tif)</th>
<th>JPEG (.jpeg, .jpg) but only if created in this format</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TIFF (other versions) (.tif, .tiff)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adobe Portable Document Format (PDF/A, PDF) (.pdf)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>standard applicable RAW image format (.raw)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Photoshop files (.psd)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Digital audio data</th>
<th>Free Lossless Audio Codec (FLAC) (.flac)</th>
<th>MPEG-1 Audio Layer 3 (.mp3) but only if created in this format</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Audio Interchange File Format (AIFF) (.aif)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Waveform Audio Format (WAV) (.wav)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Digital video data</th>
<th>MPEG-4 (.mp4)</th>
<th>motion JPEG 2000 (.mj2)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Documentation and scripts</th>
<th>Rich Text Format (.rtf)</th>
<th>plain text (.txt)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PDF/A or PDF (.pdf)</td>
<td>some widely-used proprietary formats, e.g. MS Word (.doc/.docx) or MS Excel (.xls/.xlsx)</td>
</tr>
<tr>
<td></td>
<td>HTML (.htm)</td>
<td>XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0</td>
</tr>
<tr>
<td></td>
<td>OpenDocument Text (.odt)</td>
<td></td>
</tr>
</tbody>
</table>

[http://www.data-archive.ac.uk/create-manage/format/formats-table](http://www.data-archive.ac.uk/create-manage/format/formats-table)
3.5 Limited internet connectivity

Internet connectivity: Access, affordability, ability

• Easily accessible information products (e.g. websites, PDFs)
• Alternative versions that require minimal data download
• Mobile versions
• Hosting/server choices based on connectivity – is cloud computing for everyone?
3.2 Interoperability

- Syntactic interoperability: communicate and exchange data
- Semantic interoperability: ascribe meaning to and automatically interpret data (via common vocabularies, trait dictionaries etc)
CGIAR works across centers via 16 CRPs

- MAIZE
- WHEAT
- GRiSP (Global Rice Science Partnership)
- Roots, Tubers & Bananas
- Dryland Cereals
- Grain Legumes
- Livestock & Fish

- CRP for Managing & Sustaining Crop Collections
- Policies, Institutions & Market
- Agriculture for Nutrition & Health
- Humid Tropics
- Aquatic Agricultural Systems
- Dryland Systems

- Climate Change, Agriculture and Food Security (CCAFS)
- Forests, Trees and Agroforestry (FTA)
- Water, Land and Ecosystems (WLE)
Poverty: 1.2 billion on less than USD 1 per day

Each dot represents 100,000 people living on less than 1 dollar per day
# CGIAR repository systems - publications

- Almost all Centers have/are finalizing publication repositories

<table>
<thead>
<tr>
<th>Center</th>
<th>Item-Level Access</th>
<th>Launch Year</th>
<th># Items</th>
<th>Repository Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa Rice</td>
<td>Metadata only</td>
<td>2009</td>
<td>579</td>
<td>Other - Mendeley</td>
</tr>
<tr>
<td>Bioversity</td>
<td>Metadata only</td>
<td>2013</td>
<td>70</td>
<td>DSpace</td>
</tr>
<tr>
<td>CCAFS</td>
<td>Metadata &amp; full text</td>
<td>2012</td>
<td>643</td>
<td>DSpace</td>
</tr>
<tr>
<td>CIAT</td>
<td>Metadata &amp; full text</td>
<td>2012</td>
<td>6700</td>
<td>DSpace</td>
</tr>
<tr>
<td>CIFOR</td>
<td>Metadata only</td>
<td>2009</td>
<td>3000</td>
<td>Other - InMagic</td>
</tr>
<tr>
<td>CIMMYT</td>
<td>Full text</td>
<td>2010</td>
<td>3500</td>
<td>DSpace</td>
</tr>
<tr>
<td>CIP</td>
<td></td>
<td></td>
<td></td>
<td>(Dspace)</td>
</tr>
<tr>
<td>ICARDA</td>
<td>Full text</td>
<td>2011</td>
<td>500</td>
<td>Other – SharePoint 2013</td>
</tr>
<tr>
<td>ICRAF</td>
<td>Full text</td>
<td>2014</td>
<td>6240</td>
<td>Other - Invenio</td>
</tr>
<tr>
<td>ICRISAT</td>
<td>Metadata &amp; full text</td>
<td>2009</td>
<td>7000</td>
<td>EPrints</td>
</tr>
<tr>
<td>IFPRI</td>
<td>Metadata &amp; full text</td>
<td>2011</td>
<td>10000</td>
<td>OCLC/ContentDM</td>
</tr>
<tr>
<td>IITA</td>
<td>Full text</td>
<td>2009</td>
<td>8000</td>
<td>Other - Aigaion</td>
</tr>
<tr>
<td>ILRI</td>
<td>Metadata &amp; full text</td>
<td>2009</td>
<td>12000</td>
<td>DSpace</td>
</tr>
<tr>
<td>IWMI</td>
<td>Full text</td>
<td></td>
<td>2390</td>
<td>DSpace</td>
</tr>
<tr>
<td>WLE</td>
<td>Metadata &amp; full text</td>
<td>2013</td>
<td>450</td>
<td>DSpace</td>
</tr>
</tbody>
</table>
### CGIAR repository systems - data

- Several Centers have data repositories – several developing...

<table>
<thead>
<tr>
<th>Center</th>
<th>Data Repository</th>
<th>Repository Platform</th>
<th>Launch Year</th>
<th># Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa Rice</td>
<td>Yes</td>
<td>Dataverse</td>
<td>2012</td>
<td>24</td>
</tr>
<tr>
<td>Bioversity</td>
<td>Yes</td>
<td>Dataverse</td>
<td>2013</td>
<td>4</td>
</tr>
<tr>
<td>CCAFS</td>
<td>Yes</td>
<td>Dataverse</td>
<td>2012</td>
<td>108,103</td>
</tr>
<tr>
<td>CIAT</td>
<td>Yes</td>
<td>Dataverse</td>
<td>2013</td>
<td>6</td>
</tr>
<tr>
<td>CIFOR</td>
<td>Yes</td>
<td>Dataverse</td>
<td>2013</td>
<td>55</td>
</tr>
<tr>
<td>CIMMYT</td>
<td>Under development</td>
<td>Dataverse</td>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>CIP</td>
<td>Under development</td>
<td>Dataverse/Biomart</td>
<td>2008</td>
<td>100,000</td>
</tr>
<tr>
<td>ICARDA</td>
<td>Under development</td>
<td></td>
<td>2009</td>
<td></td>
</tr>
<tr>
<td>ICRAF</td>
<td>Yes</td>
<td>Dataverse</td>
<td>2013</td>
<td>209</td>
</tr>
<tr>
<td>ICRISAT</td>
<td>Yes</td>
<td>Dataverse</td>
<td>2013</td>
<td>450</td>
</tr>
<tr>
<td>IFPRI</td>
<td>Yes</td>
<td>Dataverse</td>
<td>2008</td>
<td>107</td>
</tr>
<tr>
<td>IITA</td>
<td>Under development</td>
<td>CKAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILRI</td>
<td>Under development</td>
<td>CKAN</td>
<td>2014</td>
<td></td>
</tr>
<tr>
<td>IWMI</td>
<td>Yes</td>
<td>Dataverse?</td>
<td>2009</td>
<td>3000</td>
</tr>
<tr>
<td>WLE</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Getting “Big”…

http://informationcatalyst.com/

CGIAR is a global research partnership for a food secure future
Guiding principles for repositories

• Adherence to spirit of “openness” – CGIAR Policy; donor policies

• Repositories compliant with industry standards for semantic interoperability
  – OAI-PMH (metadata)
  – XML, RDF, JSON, O-data…

• Common minimum metadata schema → CG core
  – CG core application profile – v. Sep 2015
  – CG core – AgTrials mapping
Federated search across centers & high precision and recall based on controlled vocabulary

Controlled vocabulary
Search expansion based on relationship of ontologies

Categorized content type
Browsed by content type first to narrow down results

Geo-referencing
Any contents referenced via standard geo-coordination (ISO) and easy to map corresponding region/countries

Collection of tools
Toolkits for analysis and further researchers

Machine-readable
Human and machine readable contents

Inference based on machine agents
Provide better customized search

Top Search

Search databases

Results found in 17 databases for "drought tolerance maize"

Subjects
- Agroforestry: resources related to agroforestry
- Agronomy: resources related to agronomy
- Aquaculture/fisheries: resources related to aquaculture/fisheries
- Climate change: resources related to climate change
- GIS/remote sensing: resources related to GIS/remote sensing
- Genebank: plant resources
- Genetic/genomic: genetic/genomic resources
- Hydrology/water management: resources related to hydrology/water management
- Livestock/animal breeding: resources related to livestock/animal breeding
- Natural resource mgmt.: resources related to natural resource management
- Plant breeding: resources related to plant breeding
- Plant protection: resources related to plant protection
- Socioeconomic/livelihoods: resources related to socioeconomic/livelihoods
- Other: resources related to other subjects

Geographies
- Asia: resources for subregions and countries in Asia
- Africa: resources for subregions and countries in Africa
- Middle East: resources for subregions and countries in the Middle East
- Europe: resources for subregions and countries in Europe
- N. America: resources for subregions and countries in N. America
- S. America: resources for subregions and countries in S. America

Georeferencing
Any contents referenced via standard geo-coordination (ISO) and easy to map corresponding region/countries

Collection of tools
Toolkits for analysis and further researchers

Inference based on machine agents
Provide better customized search
Building a “cyberinfrastructure for agriculture”

- Infrastructure (Linked Open Data-enabled)
- Semantics + interoperability
- Toolkit (analysis, visualization, discovery, foresight, M+E…)
- Culture (advocacy, education, support to facilitate data-sharing)
- Partners (FAO, CABI, USDA, AgMIP, WUR, INRA, CIRAD…)

- agronomic++ e.g. CSISA/TAMASA
- socioeconomic e.g. LSMS-ISA/DHS
- Technology/investment mapping?

CGIAR is a global research partnership for a food secure future