

ECPGR Documentation & Information Network meeting  
Workshop 20-22 May 2014, Prague-Ruzyně, Czech Republic

---

## Data exchange: the Darwin Core and other approaches

---

Dag Endresen  
GBIF-Norway, Natural History Museum of the University in Oslo (NHM-UiO)  
Global Biodiversity Information Facility (GBIF)

20<sup>th</sup> May 2014



# Why did we make a Darwin Core extension for germplasm data?

## → Upgrade germplasm data pathways to use web services

The objective (1) was to enable sharing of germplasm information using the standard web-service based biodiversity data publishing toolkits maintained by the Global Biodiversity Information Facility (GBIF) and the Biodiversity Information Standards (TDWG).

## → Upgrade data types to include trait data

The objective (2) was to expand on the germplasm data types published to germplasm data portal from basic passport data to include in particular crop trait information.



## EPGRIS3 Meeting 'Inclusion of C&E data in EURISCO'

### ■ Summary of the meeting

May 2009

- On Thursday May 7th, a EPGRIS3 self funded workshop, hosted by Frank Begemann and Siegfried Harrer, was organised at the Bundesanstalt für Landwirtschaft und Ernährung (BLE) in Bonn, Germany.

### ■ Links

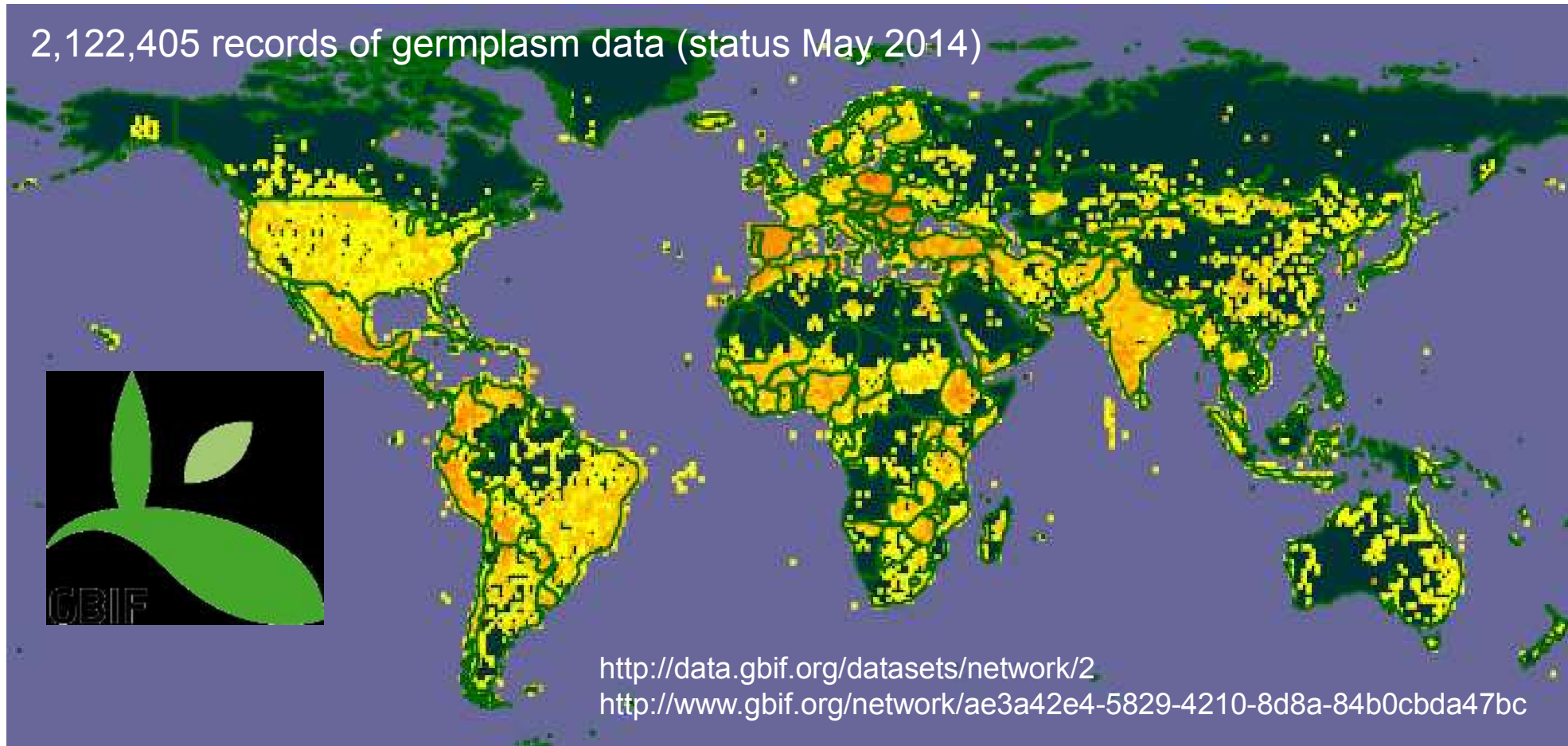
- [EPGRIS3 main page](#)
- [list of EPGRIS3 activities](#)
- [list of EPGRIS3 contacts](#)
- [Proposal 'inclusion of C&E data in EURISCO'](#)





# Potential of the GBIF technology

2,122,405 records of germplasm data (status May 2014)



Using GBIF/TDWG technology (and contributing to its development), the PGR community can more easily establish specific PGR networks without duplicating GBIF's work.

The compatibility of data standards between PGR and biodiversity collections made it possible to integrate the worldwide germplasm collections into the biodiversity community (TDWG, GBIF).

# WHAT IS THE GLOBAL BIODIVERSITY INFORMATION FACILITY?

GBIF enables *free and open access to biodiversity data* online.

We are an international government-initiated and funded initiative focused on making biodiversity data available to all and anyone, for scientific research, conservation and sustainable development.

## Global Biodiversity Information Facility

May 2014

Free and open access to biodiversity data

448,525,599

OCCURRENCES

1,454,695

SPECIES

15,081

DATASETS

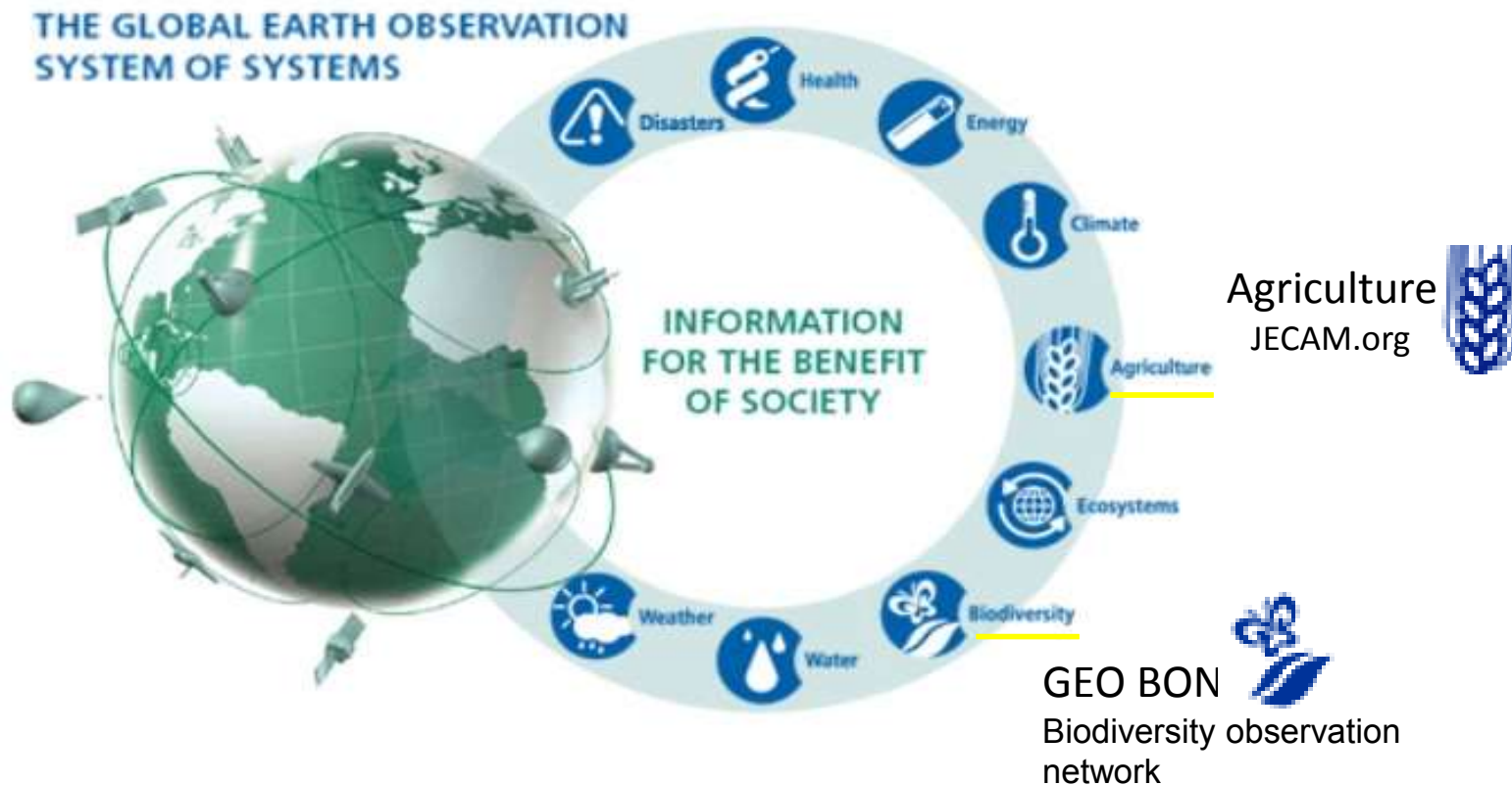
605

DATA PUBLISHERS



# GBIF and GEO

Intergovernmental group on earth observations



## Data Integration & Interoperability

GBIF provides the infrastructure for delivering species occurrence data.

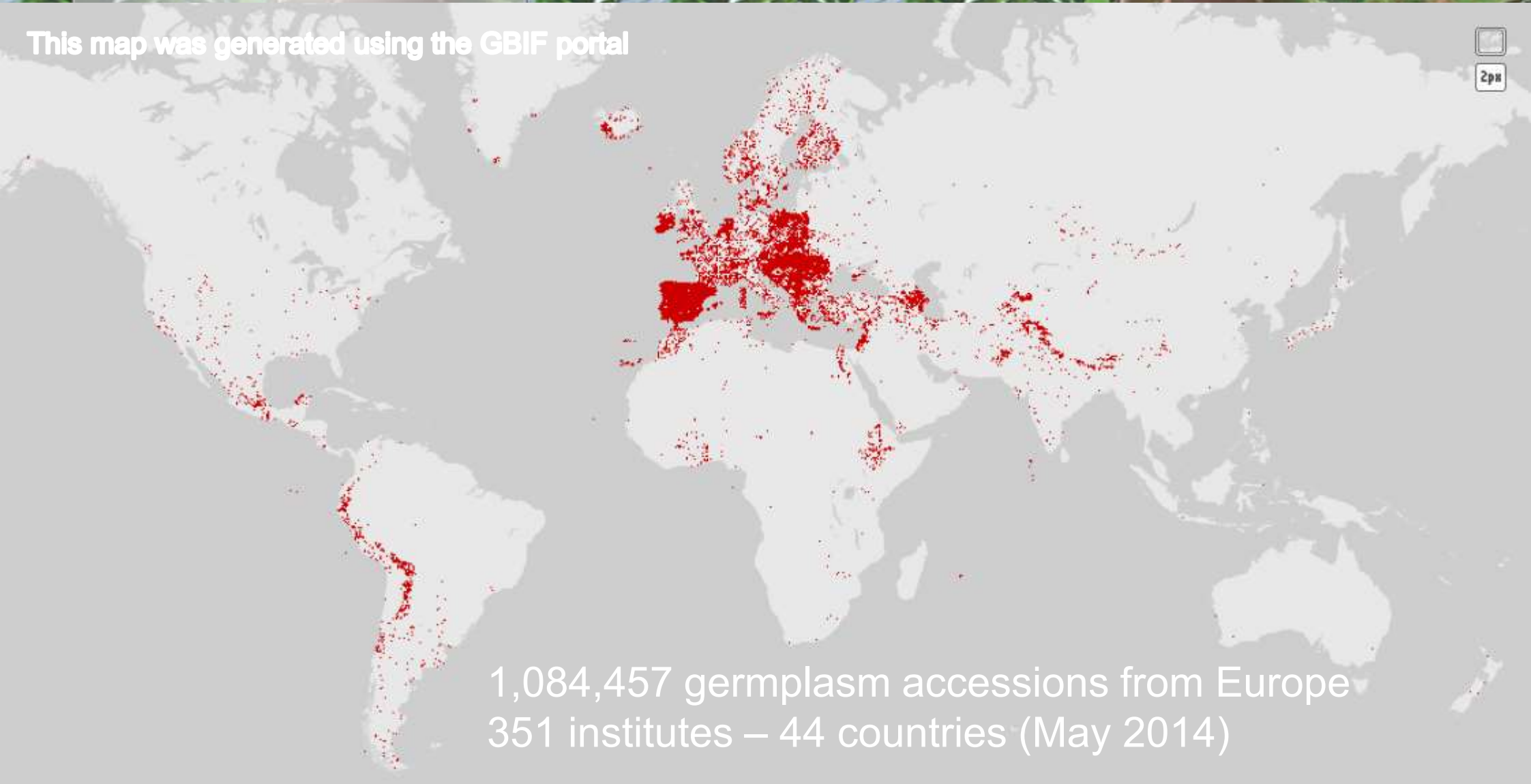




2,773,082 germplasm accessions worldwide  
444 institutes – 252 countries (May 2014)



This map was generated using the GBIF portal



The European Genetic Resources Search Catalogue (EURISCO) receives data from the National Inventories (NI) and provides access to all *ex situ* PGR accessions in Europe, <http://eurisco.ecpgr.org>



*A safety network for our crops*



ECPGR Homepage  Germplasm Databases

Search

About ECPGR

Steering Committee

Executive Committee

Networks

Working Groups

Meetings and quota

Public Awareness

Publications

Germplasm Databases

ECPGR Secretariat

ECPGR Listserver

Links

## ECPGR Central Crop Databases and other Crop Databases

A number of ECPGR Central Crop Databases have been established through the initiative of individual institutes and of ECPGR Working Groups. The databases hold passport data and, to varying degrees, characterization and primary evaluation data of the major collections of the respective crops in Europe. The germplasm accessions are maintained in different European institutes for long-term conservation. To request material included in the databases, please refer to the donor institutes which maintain the germplasm accessions.

You can search for **Crop Databases** in the **ECPGR Networks** listed below or for a more specific search use the search window.

 **Cereals** (8 databases)

 **Forages** (10 databases)

 **Oil and Protein Crops** (10 databases)

 **Fruit** (8 databases)

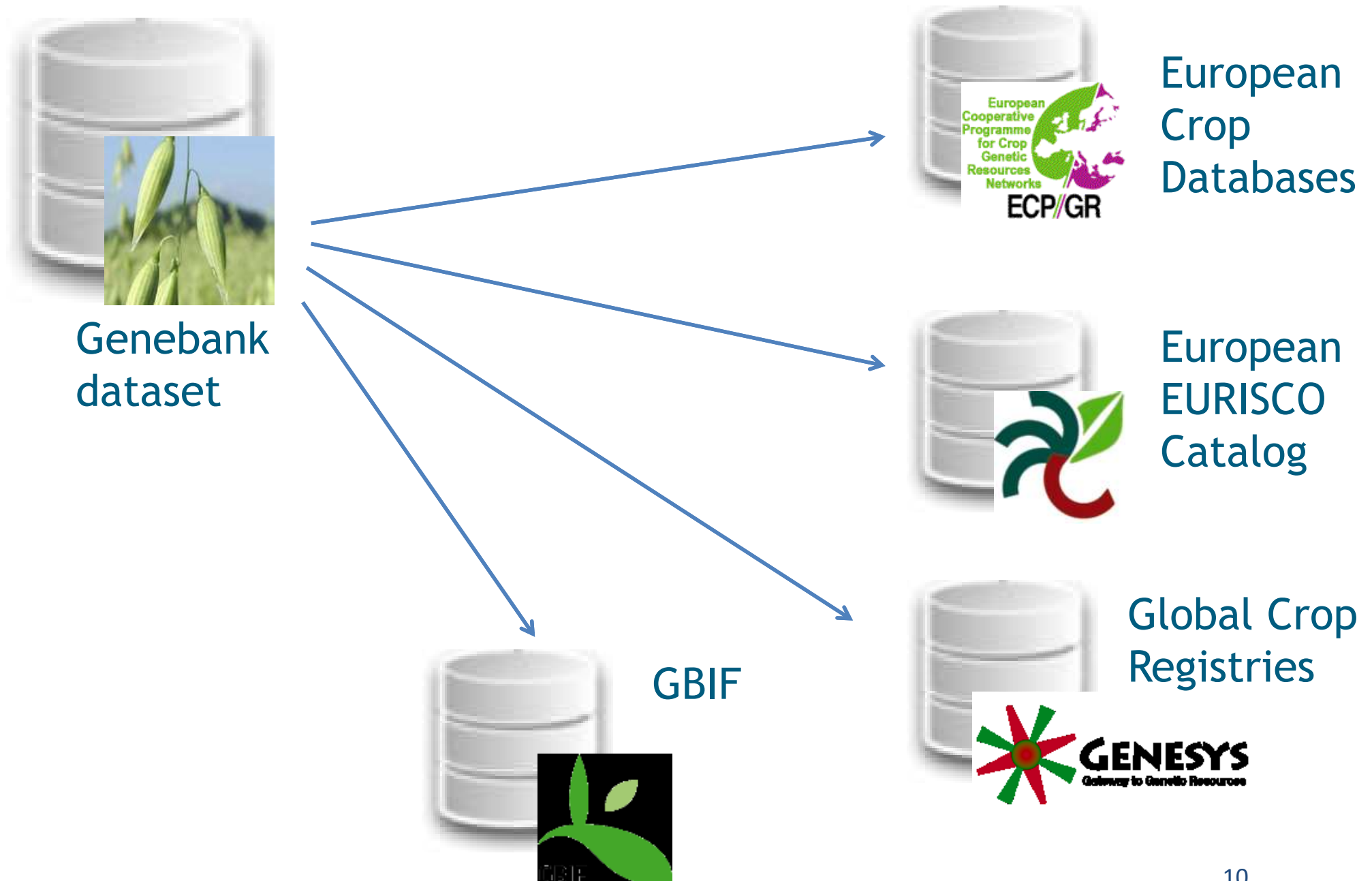
 **Sugar, Starch and Fibre Crops** (6 databases)

 **Vegetables** (22 databases)

A total of **64 ECPGR Central Crop Databases** have been established by individual institutes and the ECPGR Working Groups. The databases hold passport data and, to varying degrees, characterization and primary evaluation data of the major collections of the respective crops in Europe,

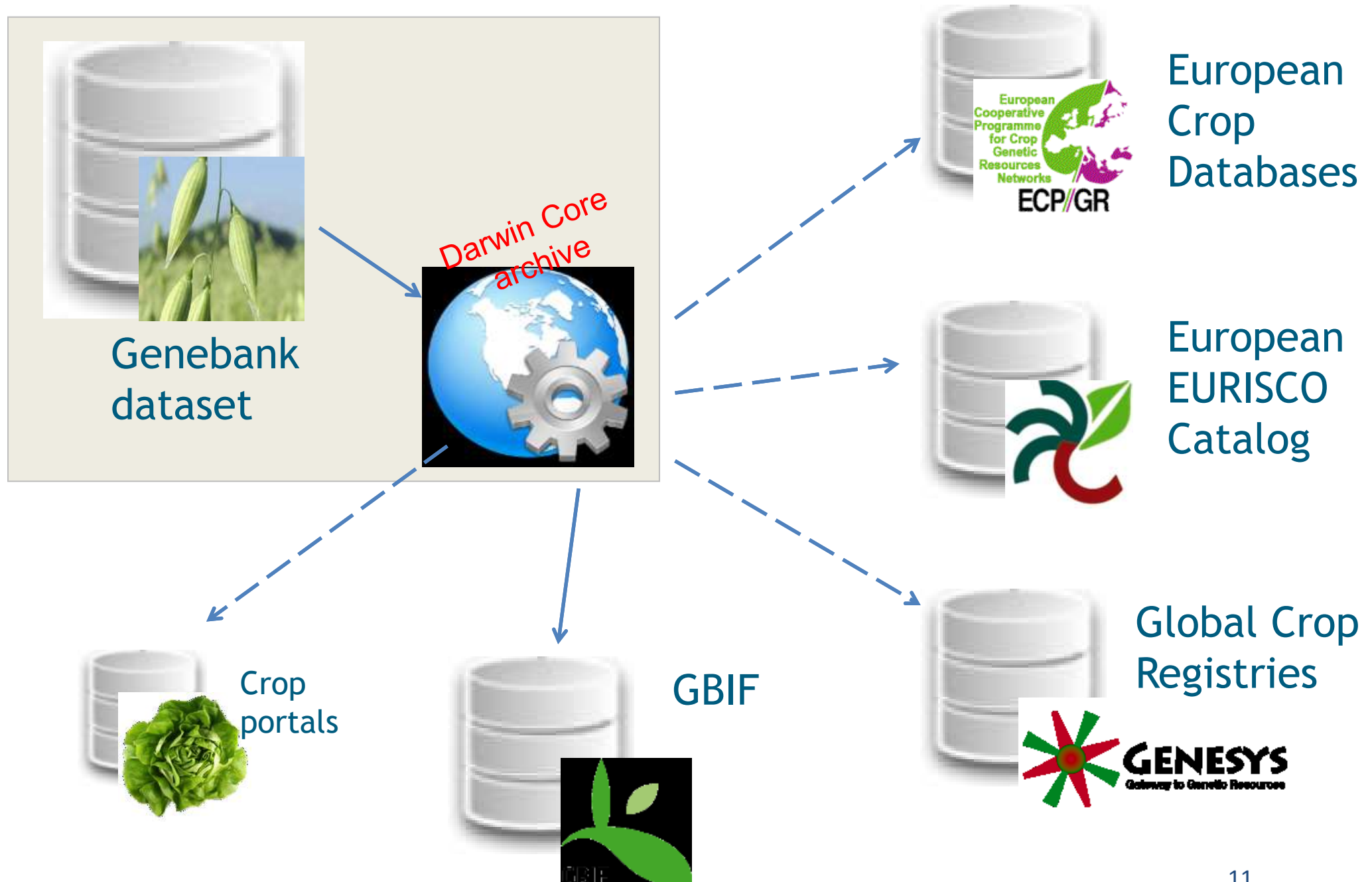
[http://www.ecpgr.cgiar.org/germplasm\\_databases/central\\_crop\\_databases.html](http://www.ecpgr.cgiar.org/germplasm_databases/central_crop_databases.html)

# Multiple data export services for each genebank





# → Multiple-purpose data export services



# Network Model

- ❖ Each dataset is shared from the holding gene bank.
- ❖ The National Inventory (NI) endorse all national gene banks for EURISCO.
- ❖ ECPGR Crop databases can access passport data from EURISCO and additional crop specific data from the gene bank IPT interface.
- ❖ Standard data sharing tools ensure that the genebank dataset is available to other relevant decentralized thematic, regional or global networks.

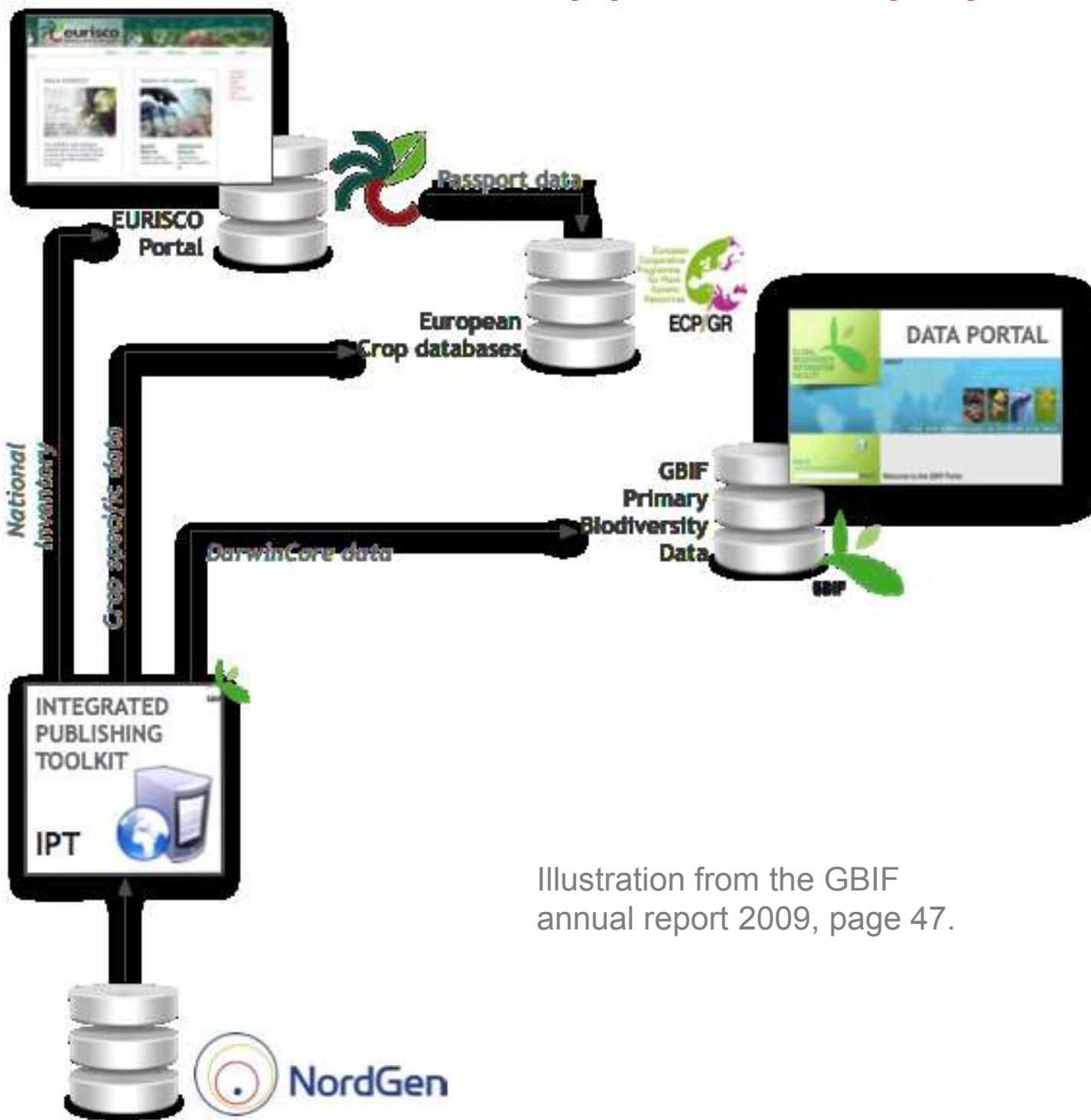


Illustration from the GBIF annual report 2009, page 47.





# Background and context



## FAO/Bioversity Multi-Crop Passport Descriptors V.2 [MCPD V.2].

June 2012



### Related files

- [PDF free file available](#) 1.9 M

The FAO/Bioversity Multi-Crop Passport Descriptors (MCPD V.2) is the result of a thorough revision of the publication originally released by FAO/IPGRI in 2001, which has been widely used as the international standard to facilitate germplasm passport information exchange. In this improved version, the 2001 list of descriptors has been expanded to accommodate emerging documentation needs, derived *inter alia* from the entry into force of the International Treaty on Plant Genetic Resources for Food and Agriculture and its Multilateral System for access and benefit-sharing, and from technological changes such as the broader use of GPS tools.

The descriptors are compatible with Bioversity's crop descriptor lists, with the descriptors used by the FAO World Information and Early Warning System (WIEWS) on plant genetic resources (PGR), and the GENESYS global portal.

**Category:** Descriptors

**Author:** Alercia, A; Diulgheroff, S; Mackay, M.

**Corporate Author:** Food and Agriculture Organization of the United Nations (FAO), Rome (Italy); Bioversity International, Rome (Italy)

**Pages:** 11 p.

**Publication Year:** 2012

**Publication Format:** PDF

**Language:** EN

MCPD  
revisions

1997  
2001  
2012

### FAO/BIOVERSITY MULTI-CROP PASSPORT DESCRIPTORS V.2

June 2012

The FAO/Bioversity Multi-Crop Passport Descriptors (MCPD V.2) is the result of a thorough revision of the original publication released by FAO/IPGRI in 2001, which has been widely used as the international standard to facilitate germplasm passport information exchange. In this improved version the 2001 list of descriptors has been expanded to accommodate emerging documentation needs, derived *inter alia* from the entry into force of the International Treaty on Plant Genetic Resources for Food and Agriculture and its Multilateral System for access and benefit-sharing, and from technological changes such as the broader use of GPS tools. The descriptors and allowed values of the first version form a subset of those in this revision.

These descriptors are compatible with Bioversity's crop descriptor lists, with the descriptors used by the FAO World Information and Early Warning System (WIEWS) on Plant Genetic Resources for Food and Agriculture, and with the GENESYS global portal.

For each multi-crop passport descriptor, a brief explanation of content, coding scheme and, in parentheses, suggested fieldname are provided to assist in the computerized exchange of this type of data. Annex I provides easy access to the historical 'List of major changes' of all descriptor elements.

It is recognized that networks or groups of users may want to further expand this revised MCPD list to meet their specific needs. As long as these additions allow for an easy conversion to the format provided in MCPD V.2, basic passport data can be exchanged worldwide in a consistent manner.

#### Common formatting rules

- If a field allows multiple values, these values should be separated by a semicolon (;) without space (e.g. Accession name: Symphony;Emma;Songino).
- A field for which no value is available should be left empty (e.g. Elevation). If data are exchanged in ASCII format, a field with a missing numeric value should be left empty. If data are exchanged in a database format, missing numeric values should be represented by generic NULL values.
- Dates are recorded as YYYYMMDD. If the month or day are missing, this should be indicated with hyphens or '00' [double zero]. If both (month and day) are missing, two double zeros are needed (e.g. 1975---, 19750000; 197506--, 19750600).

# Data publishing toolkits



ICIS (Java, 1996 →)



BioMOBY (Perl, 2001 →)



EURISCO (tab-delimited, 2003 →)



DiGIR (PHP, 2001 - 2006)



TapirLink (PHP, 2007 →)

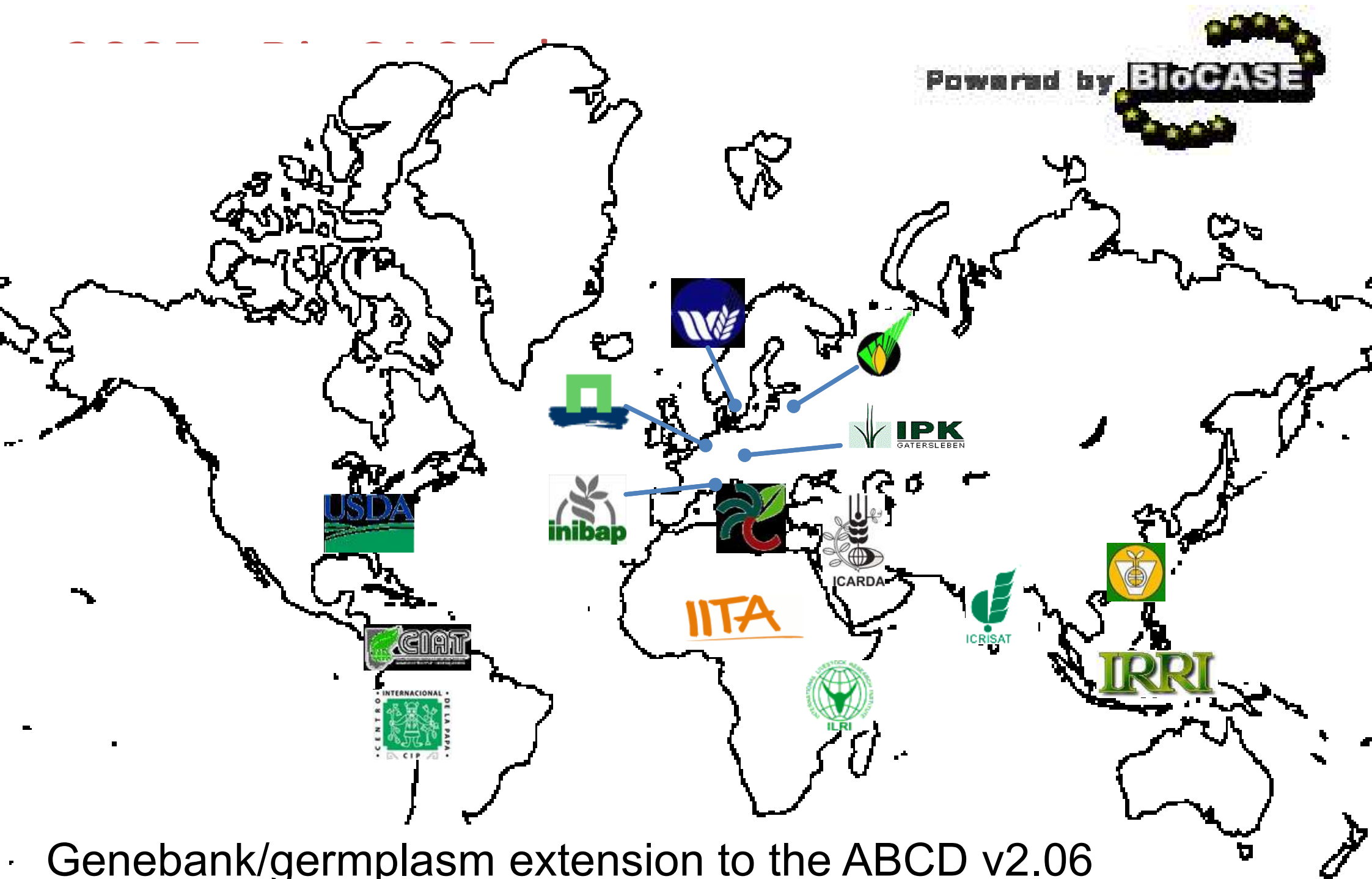


BioCASE (Python, 2001 →)



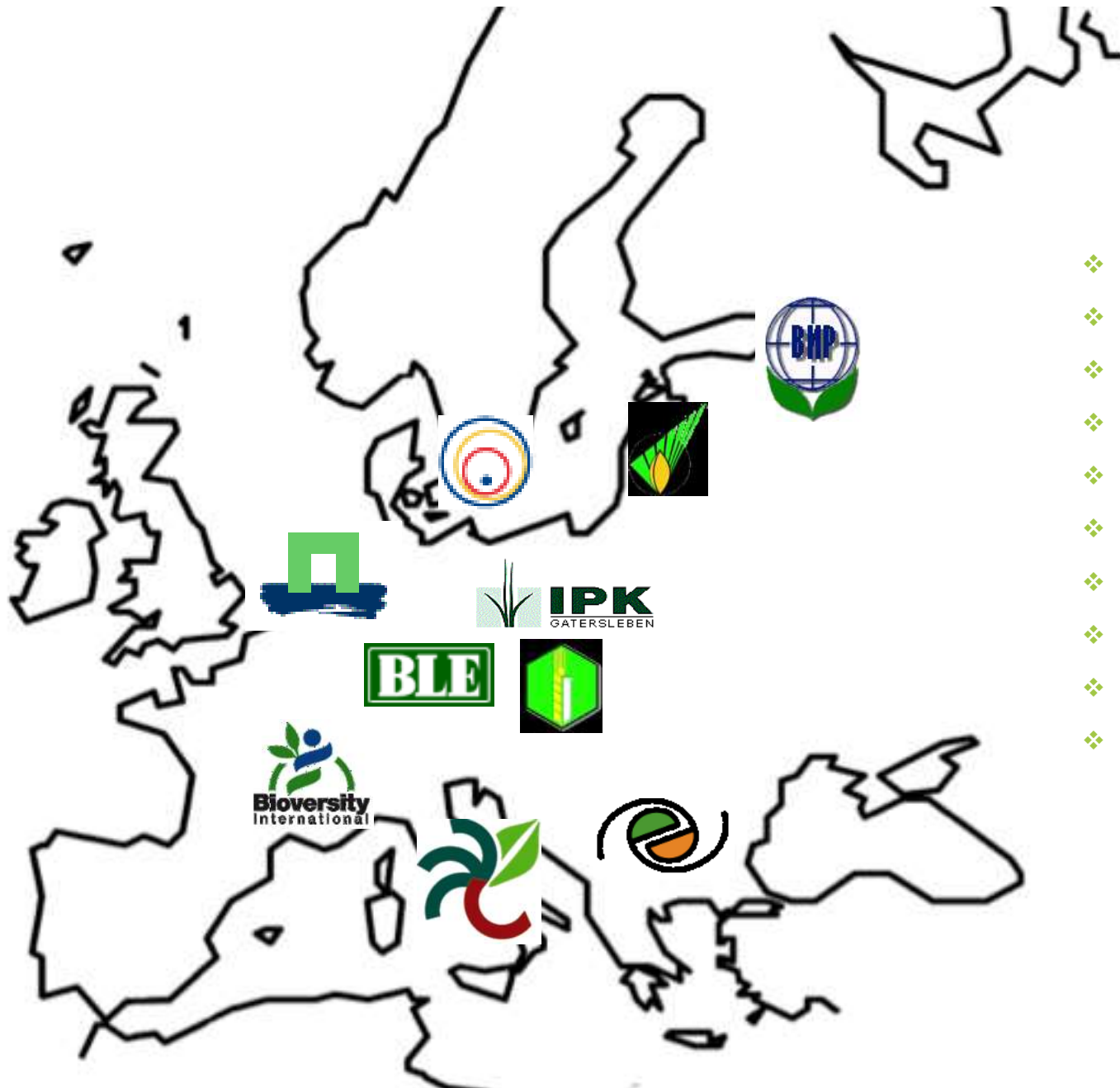
GBIF IPT (Java, 2009 →)





Genebank/germplasm extension to the ABCD v2.06

# 2010 : IPT installations for EURISCO



- ❖ EURISCO
- ❖ NordGen (Nordic countries)
- ❖ Bioversity-Montpellier (France)
- ❖ IPK Gatersleben (Germany)
- ❖ BLE (Germany)
- ❖ WUR CGN (The Netherlands)
- ❖ CRI (Czech Republic)
- ❖ VIR (Russian Federation)
- ❖ SeedNET (Balkan)
- ❖ Baltic (Estonia, Latvia, Lithuania)



Mapping of MCPD → Darwin Core  
was required before using the GBIF IPT

Mostly a mapping of MCPD terms  
to Darwin Core.

The first DRAFT version (0.1) was  
released in August 2009.



# Darwin Core extension for germplasm

The Darwin Core extension for germplasm data is an extension to the Darwin Core standard.

Includes additional terms required for describing germplasm resources that were missing in Darwin Core.

Provides a mapping of MCPD terms and Darwin Core terms.



- Endresen, D., S. Gaiji, and T. Robertson (2009). Darwin Core Germplasm extension and deployment in the GBIF infrastructure. Proceedings of TDWG 2009, Montpellier, France. Bioversity Information Standards (TDWG).
- Endresen, D.T.F. and H. Knüpffer (2012). The Darwin Core extension for genebanks opens up new opportunities for sharing genebank data sets. Biodiversity Informatics 8:11-29.



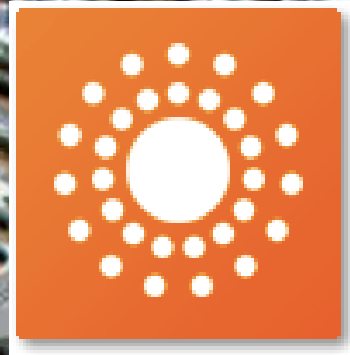
# Darwin Core

*“The Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, and samples, and related information.”*

- a well-defined standard core vocabulary
- a flexible framework to maximize re-usability
- approved as TDWG standard in 2009

<http://rs.tdwg.org/dwc/>

Wieczorek J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, D. Vieglais (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. doi:10.1371/journal.pone.0029715





# Darwin Core - a vocabulary of terms



Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, De Giovanni R, Robertson T, and Vieglaiss D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* 7(1): e29715. (doi:10.1371/journal.pone.0029715)

# Vocabularies/ontologies

- Provide a shared understanding of what we mean when describing biodiversity entities.
- What kind of thing or property.
- A list of things we as a community can agree upon the meaning of.
- “*Concept repository*” with terms identified by URIs.

TDWG Technical Roadmap 2008 (convened by Roger Hyam).



Photo CC-by-3.0 by Hannes Grobe/AWI. Palaeoclimate archives.



## Record-level Terms

[dcterms:type](#) | [dcterms:modified](#) | [dcterms:language](#) | [dcterms:rights](#) | [dcterms:rightsHolder](#) | [dcterms:accessRights](#) | [dcterms:bibliographicCitation](#) | [dcterms:references](#)

[institutionID](#) | [collectionID](#) | [datasetID](#) | [institutionCode](#) | [collectionCode](#) | [datasetName](#) | [ownerInstitutionCode](#) | [basisOfRecord](#) | [informationWithheld](#) | [dataGeneralizations](#) | [dynamicProperties](#)

## Occurrence

[occurrenceID](#) | [catalogNumber](#) | [occurrenceRemarks](#) | [recordNumber](#) | [recordedBy](#) | [individualID](#) | [individualCount](#) | [sex](#) | [lifeStage](#) | [reproductiveCondition](#) | [behavior](#) | [establishmentMeans](#) | [occurrenceStatus](#) | [preparations](#) | [disposition](#) | [otherCatalogNumbers](#) | [previousIdentifications](#) | [associatedMedia](#) | [associatedReferences](#) | [associatedOccurrences](#) | [associatedSequences](#) | [associatedTaxa](#)

## Event

[eventID](#) | [samplingProtocol](#) | [samplingEffort](#) | [eventDate](#) | [eventTime](#) | [startDayOfYear](#) | [endDayOfYear](#) | [year](#) | [month](#) | [day](#) | [verbatimEventDate](#) | [habitat](#) | [fieldNumber](#) | [fieldNotes](#) | [eventRemarks](#)

## dcterms:Location

[locationID](#) | [higherGeographyID](#) | [higherGeography](#) | [continent](#) | [waterBody](#) | [islandGroup](#) | [island](#) | [country](#) | [countryCode](#) | [stateProvince](#) | [county](#) | [municipality](#) | [locality](#) | [verbatimLocality](#) | [verbatimElevation](#) | [minimumElevationInMeters](#) | [maximumElevationInMeters](#) | [verbatimDepth](#) | [minimumDepthInMeters](#) | [maximumDepthInMeters](#) | [minimumDistanceAboveSurfaceInMeters](#) | [maximumDistanceAboveSurfaceInMeters](#) | [locationAccordingTo](#) | [locationRemarks](#) | [verbatimCoordinates](#) | [verbatimLatitude](#) | [verbatimLongitude](#) | [verbatimCoordinateSystem](#) | [verbatimSRS](#) | [decimalLatitude](#) | [decimalLongitude](#) | [geodeticDatum](#) | [coordinateUncertaintyInMeters](#) | [coordinatePrecision](#) | [pointRadiusSpatialFit](#) | [footprintWKT](#) | [footprintSRS](#) | [footprintSpatialFit](#) | [georeferencedBy](#) | [georeferencedDate](#) | [georeferenceProtocol](#) | [georeferenceSources](#) | [georeferenceVerificationStatus](#) | [georeferenceRemarks](#)

## GeologicalContext

[geologicalContextID](#) | [earliestEonOrLowestEonothem](#) | [latestEonOrHighestEonothem](#) | [earliestEraOrLowestErathem](#) | [latestEraOrHighestErathem](#) | [earliestPeriodOrLowestSystem](#) | [latestPeriodOrHighestSystem](#) | [earliestEpochOrLowestSeries](#) | [latestEpochOrHighestSeries](#) | [earliestAgeOrLowestStage](#) | [latestAgeOrHighestStage](#) | [lowestBiostratigraphicZone](#) | [highestBiostratigraphicZone](#) | [lithostratigraphicTerms](#) | [group](#) | [formation](#) | [member](#) | [bed](#)

## Identification

[identificationID](#) | [identifiedBy](#) | [dateIdentified](#) | [identificationReferences](#) | [identificationVerificationStatus](#) | [identificationRemarks](#) | [identificationQualifier](#) | [typeStatus](#)

## Taxon

[taxonID](#) | [scientificNameID](#) | [acceptedNameUsageID](#) | [parentNameUsageID](#) | [originalNameUsageID](#) | [nameAccordingToID](#) | [namePublishedInID](#) | [taxonConceptID](#) | [scientificName](#) | [acceptedNameUsage](#) | [parentNameUsage](#) | [originalNameUsage](#) | [nameAccordingTo](#) | [namePublishedIn](#) | [namePublishedInYear](#) | [higherClassification](#) | [kingdom](#) | [phylum](#) | [class](#) | [order](#) | [family](#) | [genus](#) | [subgenus](#) | [specificEpithet](#) | [infraspecificEpithet](#) | [taxonRank](#) | [verbatimTaxonRank](#) | [scientificNameAuthorship](#) | [vernacularName](#) | [nomenclaturalCode](#) | [taxonomicStatus](#) | [nomenclaturalStatus](#) | [taxonRemarks](#)

## Auxiliary Terms

### ResourceRelationship

[resourceRelationshipID](#) | [resourceID](#) | [relatedResourceID](#) | [relationshipOfResource](#) | [relationshipAccordingTo](#) | [relationshipEstablishedDate](#) | [relationshipRemarks](#)

### MeasurementOrFact

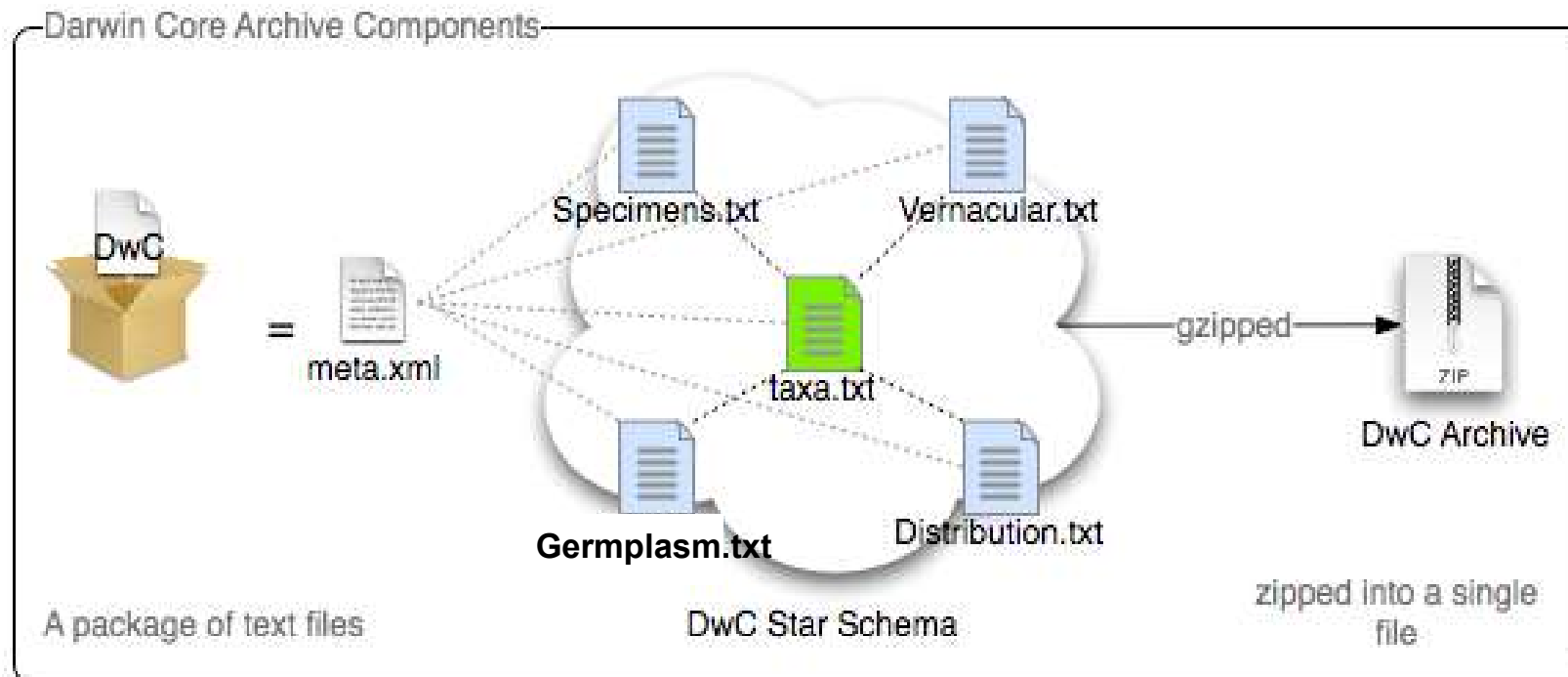
[measurementID](#) | [measurementType](#) | [measurementValue](#) | [measurementAccuracy](#) | [measurementUnit](#) | [measurementDeterminedDate](#) | [measurementDeterminedBy](#) | [measurementMethod](#) | [measurementRemarks](#)

<http://rs.tdwg.org/terms/>



# Darwin Core Archive (DwC-A)

- ❖ DwC-A publish Darwin Core records including extensions
- ❖ Simple text based format
- ❖ Zipped single file archive

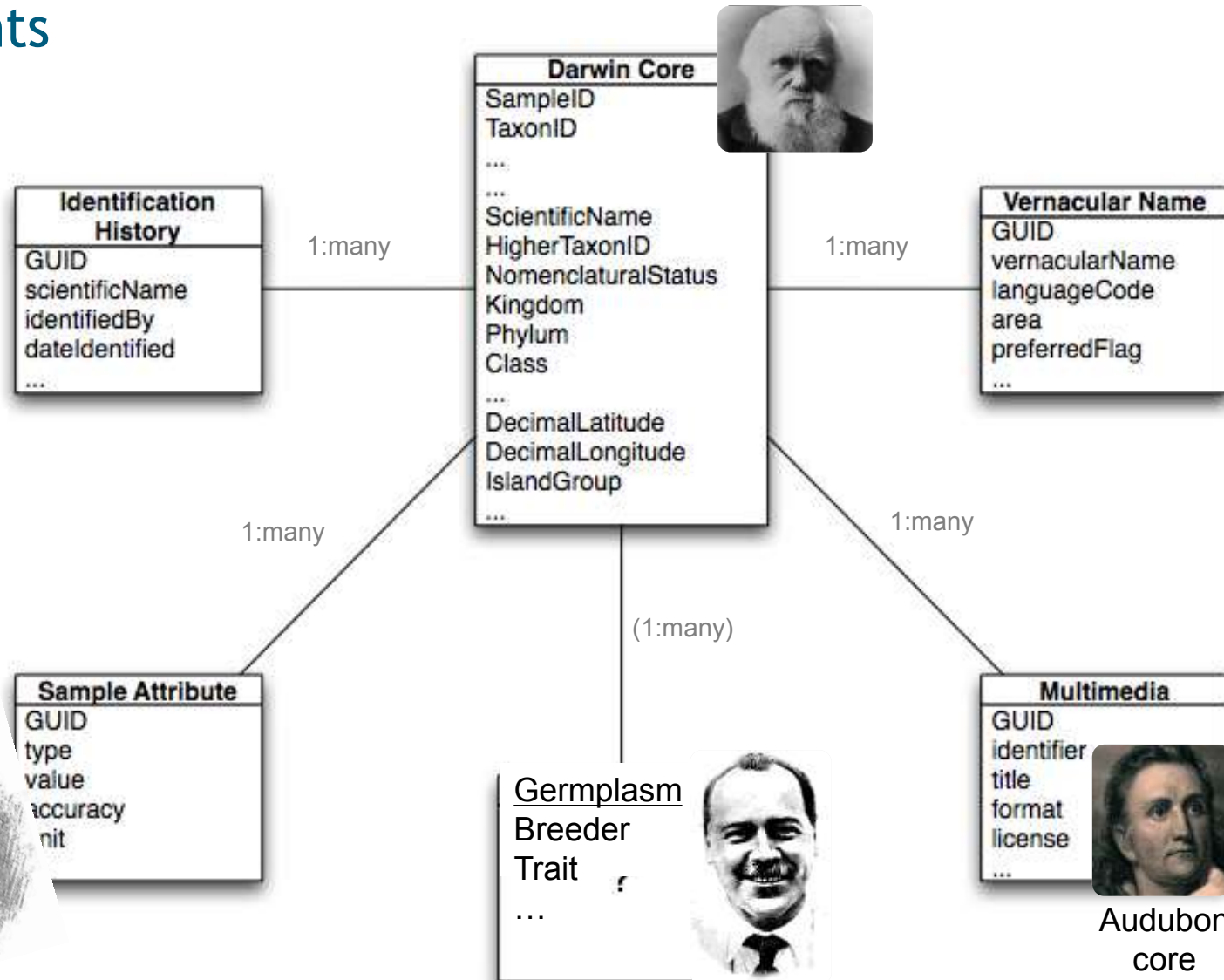




# Darwin Core star schema

Can relate elements

*one-to-one* or  
*one-to-many*.



Audubon  
core



# Darwin Core extension for germplasm

Namespace (SKOS/RDF) (stable version)

<http://purl.org/germplasm/germplasmTerm#>

Code repository (stable

version) <http://code.google.com/p/darwincore-germplasm>

Community discussion (development version)

<http://terms.tdwg.org/wiki/Germplasm>



| MCPD (2012) |                 |  | Darwin Core                               |  |  |
|-------------|-----------------|--|-------------------------------------------|--|--|
|             | (missing)       |  | dwc.datasetID                             |  |  |
|             | (missing)       |  | dwc.occurrenceID                          |  |  |
| 1           | INSTCODE        |  | dwc.institutionCode                       |  |  |
| 2           | ACCENUMB        |  | dwc.catalogNumber                         |  |  |
| 3           | COLLNUMB        |  | dwc.recordNumber                          |  |  |
| 4           | COLLCODE        |  | <a href="#">g.collectingInstituteCode</a> |  |  |
| 4.1         | COLLNAME        |  | dwc.recordedBy                            |  |  |
| 4.1.1       | COLLINSTADDRESS |  | (dwc.recordedBy)                          |  |  |
| 4.2         | COLLMISSID      |  | dwc.collectionCode                        |  |  |
| 5           | GENUS           |  | dwc.genus                                 |  |  |
| 6           | SPECIES         |  | dwc.specificEpithet                       |  |  |
| 7           | SPAUTHOR        |  | dwc.scientificNameAuthorship              |  |  |
| 8           | SUBTAXA         |  | dwc.infraspecificEpithet                  |  |  |
| 9           | SUBTAUTHOR      |  | (dwc.scientificNameAuthorship)            |  |  |
| 10          | CROPNAME        |  | dwc.vernacularName                        |  |  |
| 11          | ACCENAME        |  | <a href="#">g.breedingIdentifier</a>      |  |  |
| 12          | ACQDATE         |  | <a href="#">g.acquisitionDate</a>         |  |  |
| 13          | ORIGCTY         |  | dwc.countryCode                           |  |  |
| 14          | COLLSITE        |  | dwc.locality                              |  |  |
| 15.1        | DECLATITUDE     |  | dwc.decimalLatitude                       |  |  |
| 15.2        | LATITUDE        |  | dwc.verbatimLatitude                      |  |  |
| 15.3        | DECLONGITUDE    |  | dwc.decimalLongitude                      |  |  |
| 15.4        | LONGITUDE       |  | dwc.verbatimLongitude                     |  |  |

| MCPD (2012) |              |  | Darwin Core                                                       |  |  |
|-------------|--------------|--|-------------------------------------------------------------------|--|--|
| 15.5        | COORDUNCERT  |  | dwc.coordinateUncertaintyInMeters                                 |  |  |
| 15.6        | COORDDATUM   |  | dwc.geodetic.Datum                                                |  |  |
| 15.7        | GEOREFMETH   |  | dwc.georeferenceSources                                           |  |  |
| 16          | ELEVATION    |  | dwc.minimumElevationInMeters                                      |  |  |
| 17          | COLLDATE     |  | dwc.eventDate                                                     |  |  |
| 18          | BREDCODE     |  | <a href="#">g.breederInstituteID</a>                              |  |  |
| 18.1        | BREDNAME     |  | <a href="#">g.breedingInstitute</a>                               |  |  |
| 19          | SAMPSTAT     |  | <a href="#">g.biologicalStatus</a>                                |  |  |
| 20          | ANCEST       |  | <a href="#">g.ancestralData</a> , <a href="#">g.purdyPedigree</a> |  |  |
| 21          | COLLSRC      |  | <a href="#">g.acquisitionSource</a>                               |  |  |
| 22          | DONORCODE    |  | <a href="#">g.donorInstituteID</a>                                |  |  |
| 22.1        | DONORNAME    |  | <a href="#">g.donorInstitute</a>                                  |  |  |
| 23          | DONORNUMB    |  | <a href="#">g.donorsIdentifier</a>                                |  |  |
| 24          | OTHERNUMB    |  | dwc.otherCatalogNumbers                                           |  |  |
| 25          | DUPLSITE     |  | <a href="#">g.safetyDuplicationInstituteID</a>                    |  |  |
| 25.1        | DUPLINSTNAME |  | <a href="#">g.safetyDuplicationInstitute</a>                      |  |  |
| 26          | STORAGE      |  | <a href="#">g.storageCondition</a>                                |  |  |
| 27          | MLSSTAT      |  | <a href="#">g.mlsStatus</a>                                       |  |  |
| 28          | REMARKS      |  | dwc.occurrenceRemarks                                             |  |  |

## Mapping of DwC to MCPD

# Darwin Core extension for IPT

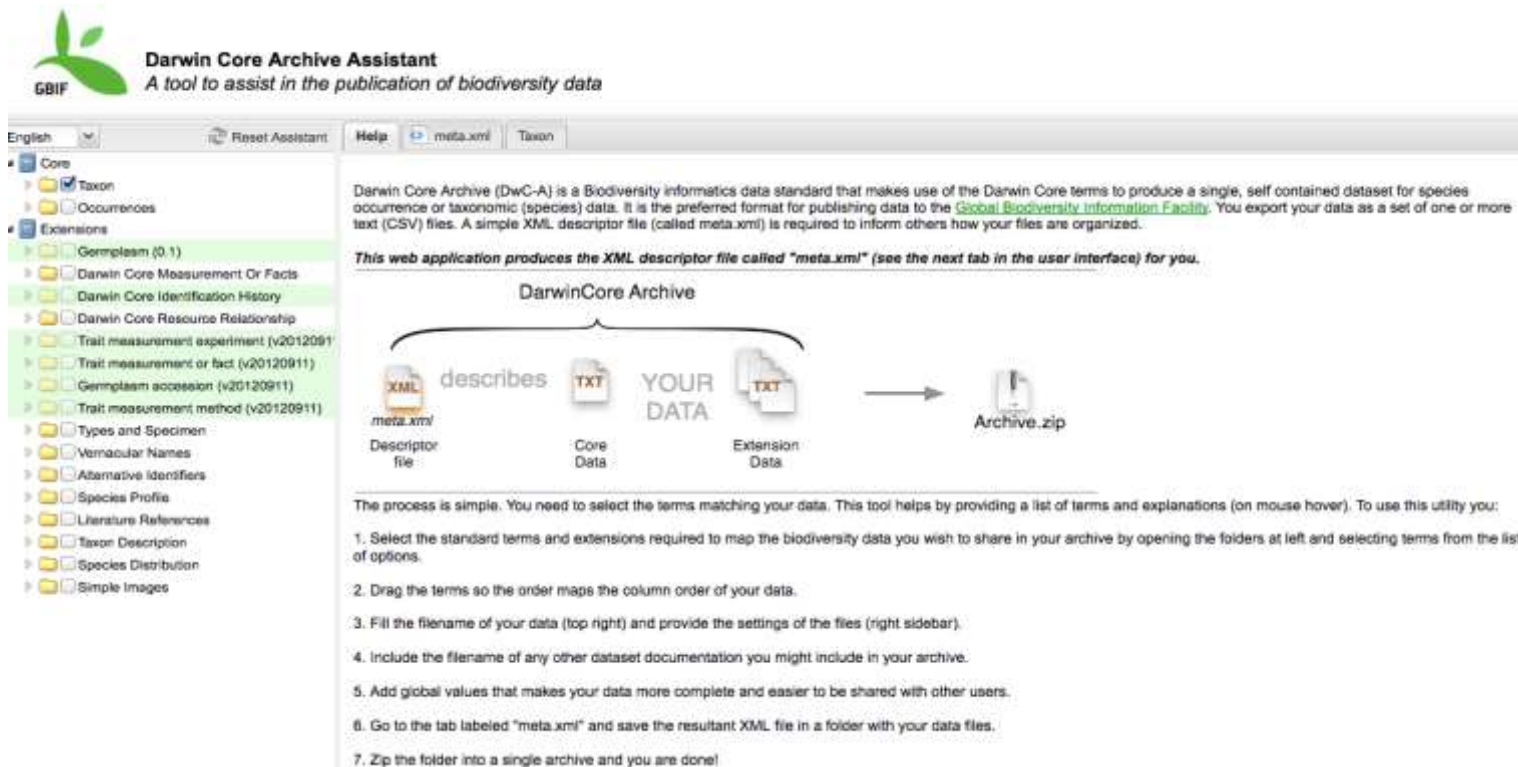
|                                                                                   | Name                   | Last modified     | Size | Description |
|-----------------------------------------------------------------------------------|------------------------|-------------------|------|-------------|
|   | Parent Directory       |                   | -    |             |
|   | GermplasmAccession.xml | 14-May-2014 12:01 | 20K  |             |
|   | MeasurementScore.xml   | 14-May-2014 12:01 | 13K  |             |
|   | MeasurementTrait.xml   | 14-May-2014 12:01 | 6.4K |             |
|  | MeasurementTrial.xml   | 14-May-2014 12:01 | 6.2K |             |

<http://rs.gbif.org/extension/germplasm/>



# Darwin Core Archive Assistant (GBIF, 2010)

The Darwin Core Archive Assistant is a web application that presents a simple interface for describing the data elements a data publisher wishes to serve to the GBIF network as basic text files and composes the appropriate XML descriptor file as defined in the Darwin Core Text Guidelines to accompany them. It communicates with the GBIF registry to provide an up-to-date listing of all relevant Darwin Core terms and available extensions and presents these in a simple checklist format.



The screenshot shows the Darwin Core Archive Assistant web application. At the top left is the GBIF logo and the title "Darwin Core Archive Assistant" with the subtitle "A tool to assist in the publication of biodiversity data". Below this is a navigation menu with a tree structure showing "Core" and "Extensions" categories. The "Core" category is expanded, showing sub-items like "Taxon", "Occurrences", "Darwin Core Measurement Or Facts", "Darwin Core Identification History", "Darwin Core Resource Relationship", "Trait measurement: experiment (v20120911)", "Trait measurement: or fact (v20120911)", "Darwin Core accession (v20120911)", "Trait measurement: method (v20120911)", "Types and Specimen", "Vernacular Names", "Alternative Identifiers", "Species Profile", "Literature References", "Taxon Description", "Species Distribution", and "Simple Images". The main content area on the right contains a description of the Darwin Core Archive (DwC-A) and a diagram illustrating the workflow. The diagram shows "XML" (meta.xml) describing "Core Data" (YOUR DATA) and "Extension Data", which are then combined into an "Archive.zip". Below the diagram, a list of seven steps explains the process: 1. Select standard terms and extensions. 2. Drag terms to map column order. 3. Fill filename and settings. 4. Include dataset documentation. 5. Add global values. 6. Save the XML file. 7. Zip the folder.

Darwin Core Archive Assistant  
A tool to assist in the publication of biodiversity data

English | Reset Assistant | Help | meta.xml | Taxon

Darwin Core Archive (DwC-A) is a Biodiversity informatics data standard that makes use of the Darwin Core terms to produce a single, self contained dataset for species occurrence or taxonomic (species) data. It is the preferred format for publishing data to the [Global Biodiversity Information Facility](http://www.gbif.org). You export your data as a set of one or more text (CSV) files. A simple XML descriptor file (called meta.xml) is required to inform others how your files are organized.

This web application produces the XML descriptor file called "meta.xml" (see the next tab in the user interface) for you.

DarwinCore Archive

XML describes TEXT YOUR DATA TEXT  
meta.xml Descriptor file Core Data Extension Data → Archive.zip

The process is simple. You need to select the terms matching your data. This tool helps by providing a list of terms and explanations (on mouse hover). To use this utility you:

1. Select the standard terms and extensions required to map the biodiversity data you wish to share in your archive by opening the folders at left and selecting terms from the list of options.
2. Drag the terms so the order maps the column order of your data.
3. Fill the filename of your data (top right) and provide the settings of the files (right sidebar).
4. Include the filename of any other dataset documentation you might include in your archive.
5. Add global values that makes your data more complete and easier to be shared with other users.
6. Go to the tab labeled "meta.xml" and save the resultant XML file in a folder with your data files.
7. Zip the folder into a single archive and you are done!

<http://tools.gbif.org/dwca-assistant/>



# Darwin Core Archive Assistant

A tool to assist in the publication of biodiversity data

<http://tools.gbif.org/dwca-assistant/>



English

☐ Add Spacer    Filename:

|    | Term                       | Required                            | Default Value |
|----|----------------------------|-------------------------------------|---------------|
| 0  | Core ID                    | <input checked="" type="checkbox"/> |               |
| 1  | measurementValue           | <input type="checkbox"/>            |               |
| 2  | measurementUnit            | <input type="checkbox"/>            |               |
| 3  | measurementAccuracy        | <input type="checkbox"/>            |               |
| 4  | germplasmID                | <input type="checkbox"/>            |               |
| 5  | germplasmIdentifier        | <input type="checkbox"/>            |               |
| 6  | measurementTraitID         | <input type="checkbox"/>            |               |
| 7  | measurementTraitName       | <input type="checkbox"/>            |               |
| 8  | measurementTraitIdentifier | <input type="checkbox"/>            |               |
| 9  | measurementTrialID         | <input type="checkbox"/>            |               |
| 10 | measurementDeterminedBy    | <input type="checkbox"/>            |               |
| 11 | measurementDeterminedDate  | <input type="checkbox"/>            |               |
| 12 | measurementByInstituteID   | <input type="checkbox"/>            |               |
| 13 | measurementMethod          | <input type="checkbox"/>            |               |
| 14 | measurementType            | <input type="checkbox"/>            |               |
| 15 | measurementGrowthStage     | <input type="checkbox"/>            |               |
| 16 | measurementRemarks         | <input type="checkbox"/>            |               |

Core

- Taxon
- Occurrences

Extensions

- Darwin Core Measurement Or Facts
- Darwin Core Identification History
- Darwin Core Resource Relationship
- Trait descriptor (v20140515)
- Trait measurement score (v20140515)**
- Trait measurement trial (v20140515)
- Germplasm accession (v20140515)
- Types and Specimen
- Simple Multimedia
- Vernacular Names
- Alternative Identifiers
- Species Profile
- Literature References
- Taxon Description
- Species Distribution
- Simple Images (deprecated)
- EOL Media Extension 1.0
- EOL References Extension 1.0



#### Directions:

1. Review the available spreadsheet templates
2. Download and complete the templates according to the user guidelines
3. Click the "Choose File" button on the right to upload the completed spreadsheet or
4. Email it to [spreadsheets@tools.gbif.org](mailto:spreadsheets@tools.gbif.org)
5. Receive a validated Darwin Core Archive file or EML metadata document.
6. To publish, put the Darwin Core Archive file and EML metadata document online, with a stable URL.
7. Follow the [GBIF data registration guidelines](#) to register the data files with GBIFS.

Upload File:

No file selected.

or

mail to:

[spreadsheets@tools.gbif.org](mailto:spreadsheets@tools.gbif.org)

**Note:** The processor does NOT publish a datafile to GBIF. It provides a publication-ready file.

Available Spreadsheet Templates – The processor currently only supports a set of pre-configured spreadsheet templates. Each is available in the older XLS format and the new XLSX format, which supports more than 64,000 rows.

### The following templates are available:



**Metadata** - Use this template to describe a database or other data resource. Processor output is an Ecological Metadata Language Document that conforms to a GBIF metadata profile. [\[XLSX format\]](#)

**Species Occurrence** - Use this template to record or store basic species collections or observational data. [\[XLSX format\]](#)

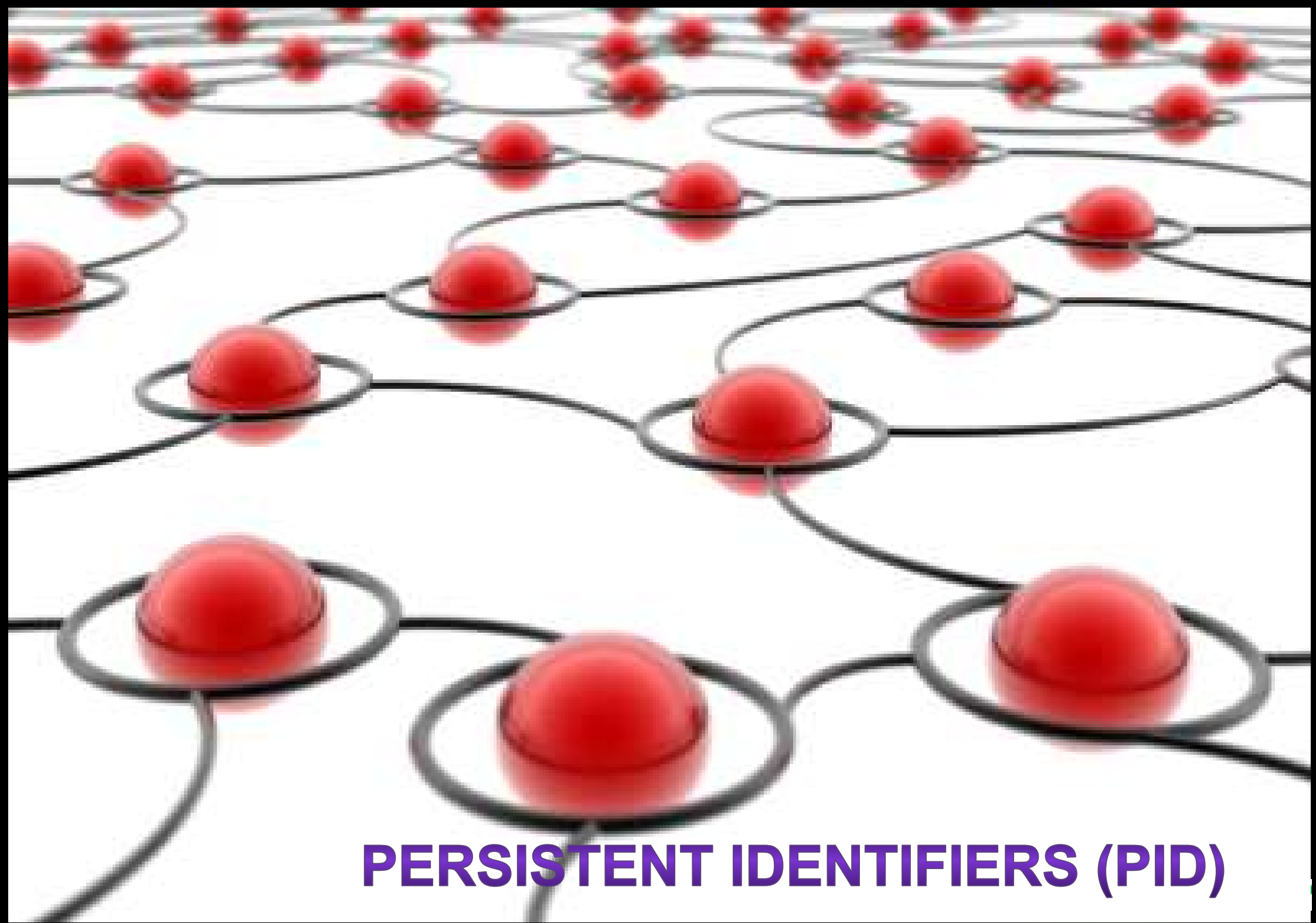
**Species Checklists** – These templates provide different methods for recording and storing simple annotated species checklists. The different versions offer different methods for recording classification information.

**Checklist 1** - Supports a 'normalised' classification where the classification is stored in two spreadsheet columns: one stores the ID of the taxon and another stores the ID of the parent taxon. [\[XLSX format\]](#)

**Checklist 2** - Classification in columns where the user can define the specific higher taxon groups. Useful when using super- or sub- ranks or other ranks outside the core Linnaean ranks. [\[XLSX format\]](#)

**Checklist 3** - Classification in columns is limited to the basic Linnaean ranks of Kingdom, Phylum, Class, Order, Family, Genus, etc. [\[XLSX format\]](#)





**PERSISTENT IDENTIFIERS (PID)**

# The purpose of identifiers

...is to name things,  
making it possible to refer to them.

## What is an identifier:

“Each identifier refers to one and only one thing” (Coyle 2006).

“An **association** between a string and a thing” (Kunze 2003).

“A **stated association** between a symbol and a thing; that the symbol may be used to **unambiguously** refer to the thing within a given **context**” (Campbell 2007).

# http – PURL – UUID

<http://purl.org/nhmuio/id/41d9cbb4-4590-4265-8079-ca44d46d27c3>

<http://purl.org/cgngenis/accenumb/CGN00001>



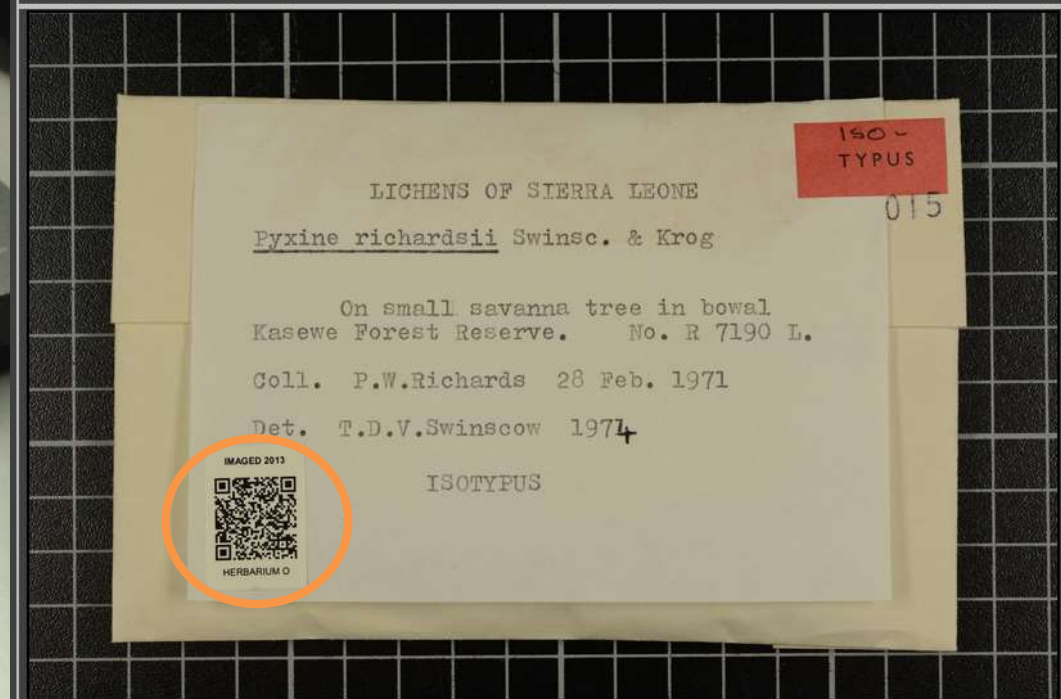
UUID QR codes for all museum objects at NHM-UiO would provide:

- **Machine-readable** using an ordinary smart phone (or PDA).
- Allows for new and efficient **workflows** for collection management.
- Deployment for **stable identifiers** appropriate for data-basing.





## Images of Lichen Types in Herbarium O (Oslo)

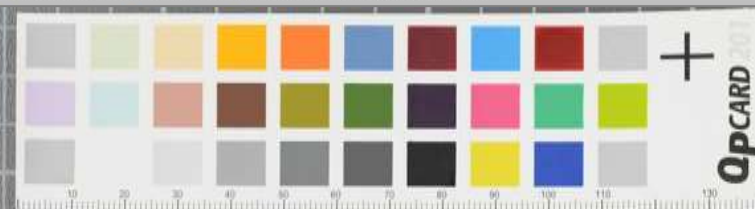


**Pyxine richardsii Swinscow & Krog** ISOTYPE O-L-000015 [Click on image for full size \(36 MPx\)](#)

**Protolog:** Swinscow, TDV/Krog, H, Norw. J. Bot. 22: 127 (1975)

**Locality:** SIERRA LEONE: bowal Kasewe Forest Reserve, 1971.02.28, Richards, P.W. R 7190 L

**Currently accepted name:** *Pyxine richardsii* Swinscow & Krog





Catalog number: O-L-000014

PID: <http://purl.org/nhmuio/id/41d9cbb4-4590-4265-8079-ca44d46d27c3>





# Resolver service

<http://purl.org/nhmuio/id/UUID> → <http://gbif.no/resolver/UUID>

**gbif.no resolver**

Identifier

*42791 records*

41d9cbb4-4590-4265-8079-ca44d46d27c3

d91e8253-0ac1-4681-ac69-e50070af86a2

|                          |                                                                |
|--------------------------|----------------------------------------------------------------|
| <b>ID:</b>               | 41d9cbb4-4590-4265-8079-ca44d46d27c3                           |
| <b>Occurrence ID:</b>    | http://purl.org/nhmuio/id/41d9cbb4-4590-4265-8079-ca44d46d27c3 |
| <b>Institution code:</b> | O                                                              |
| <b>Collection code:</b>  | L                                                              |
| <b>Catalogue number:</b> | 14                                                             |
| <b>Basis of record:</b>  | Specimen                                                       |

Including machine  
readable formats

## Event

|                     |            |
|---------------------|------------|
| <b>Recorded by:</b> | Tapper, R. |
| <b>Year:</b>        | 1971       |
| <b>Month:</b>       | 6          |
| <b>Day:</b>         | 23         |

## Taxon

|                         |                      |
|-------------------------|----------------------|
| <b>Scientific Name:</b> | Anaptychia ethiopica |
| <b>Kingdom:</b>         | Fungi                |
| <b>Phylum:</b>          | Ascomycota           |
| <b>Class:</b>           | Ascomycetes          |
| <b>Order:</b>           | Lecanorales          |
| <b>Family:</b>          | Physciaceae          |
| <b>Genus:</b>           | Anaptychia           |
| <b>Type Status:</b>     | Isotype              |

## Location

|                        |          |
|------------------------|----------|
| <b>Continent:</b>      | Africa   |
| <b>Country:</b>        | Ethiopia |
| <b>State/Province:</b> | Simen    |
| <b>Locality:</b>       | Buahit   |

## Images



# Status for genebank datasets in GBIF, May 2014

| Publisher                                               | Dataset                                                                               | Updated      | Records          | Georef         |
|---------------------------------------------------------|---------------------------------------------------------------------------------------|--------------|------------------|----------------|
| Bioversity International                                | <i>The European Genetic Resources Search Catalogue (EURISCO)</i>                      | Dec 2, 2009  | 976 457          | 87 776         |
| Bioversity International                                | <i>The System-wide Information Network for Genetic Resources (SINGER)</i>             | Oct 25, 2008 | 683 018          | 171 493        |
| US National Plant Germplasm System                      | <i>United States National Plant Germplasm System Collection (USDA GRIN)</i>           | Apr 29, 2009 | 313 949          | 66 267         |
|                                                         |                                                                                       |              |                  |                |
| Plant Breeding and Acclimatization Institute (IHAR)     | Polish gene bank passport data of plants accessions which are important in human life | May 8, 2013  | 59 948           | 16 344         |
| Plant Breeding and Acclimatization Institute (IHAR)     | Seed collection Dead seeds for evaluation and observation purposes                    | May 8, 2013  | 10 597           |                |
| Plant Breeding and Acclimatization Institute (IHAR)     | Polish seed gene bank historical passport data of accessions                          | May 8, 2013  | 8 462            | 3 476          |
| Nordic Genetic Resource Center (NORDGEN)                | Nordic Genetic Resources                                                              | Jun 6, 2012  | 37 641           | 5 237          |
| Centre for Genetic Resources, The Netherlands           | Centre for Genetic Resources, the Netherlands, PGR passport data                      | Apr 18, 2014 | 22 579           | 20 159         |
| Dep Plant Biology, Agronomy, Univ Politécnica de Madrid | Universidad Politécnica de Madrid, Dpto. Biología Vegetal, Banco de Germoplasma       | Sep 27, 2012 | 9 754            |                |
| <b>7 data publishers</b>                                | <b>9 datasets</b>                                                                     |              | <b>2 122 405</b> | <b>370 752</b> |

This list only includes the datasets classified to the PGR network:  
<http://www.gbif.org/network/ae3a42e4-5829-4210-8d8a-84b0cbda47bc>



*“Things can happen in a band, or any type of collaboration, that would not otherwise happen” (Jim Coleman, Jazz-musician).*





Thank you for  
listening!

GBIF-Norway

Dag Endresen  
dag.endresen@nhm.uio.no



UiO : Universitetet i Oslo