

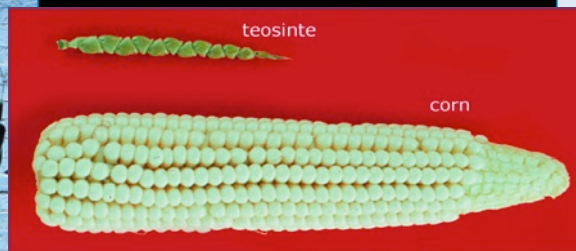
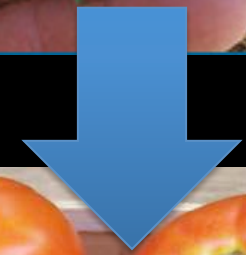
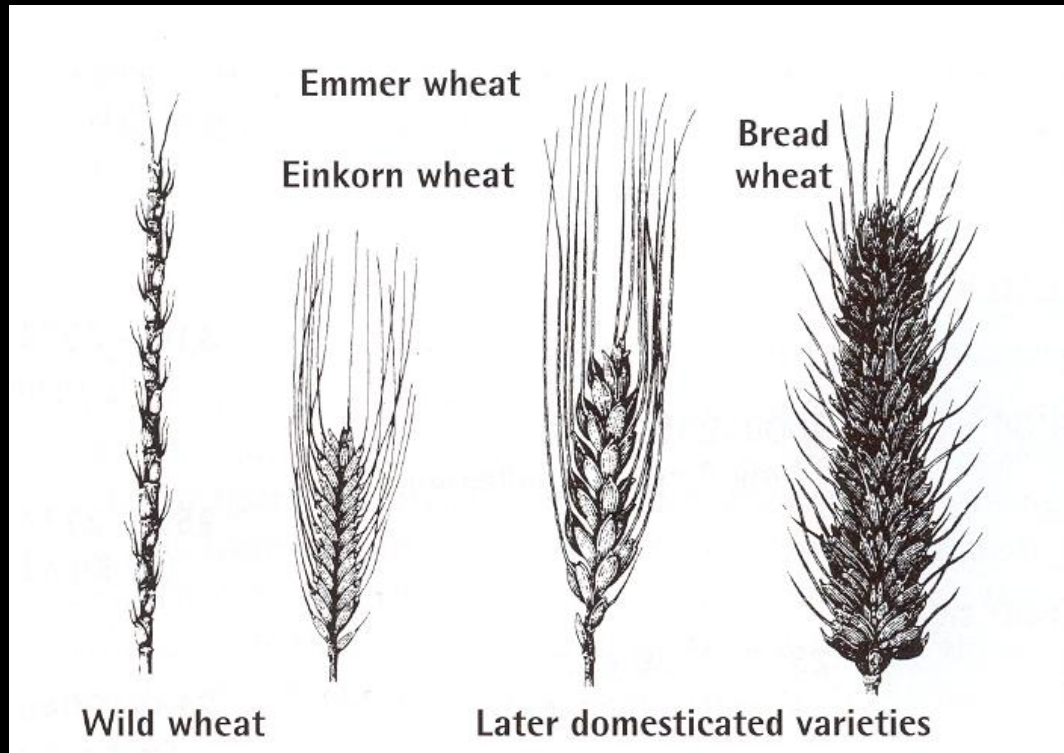
A wooden ruler is placed diagonally across a light-colored wooden surface. Scattered around and on the ruler are numerous small, round beads in three colors: green, red, and black. Some beads are clustered together, while others are isolated. The ruler has visible markings and numbers, including '9' and '10'.

Event core for measurement or fact data

Belspo, Brussel, January 2018

Domestication and cultivated plants:

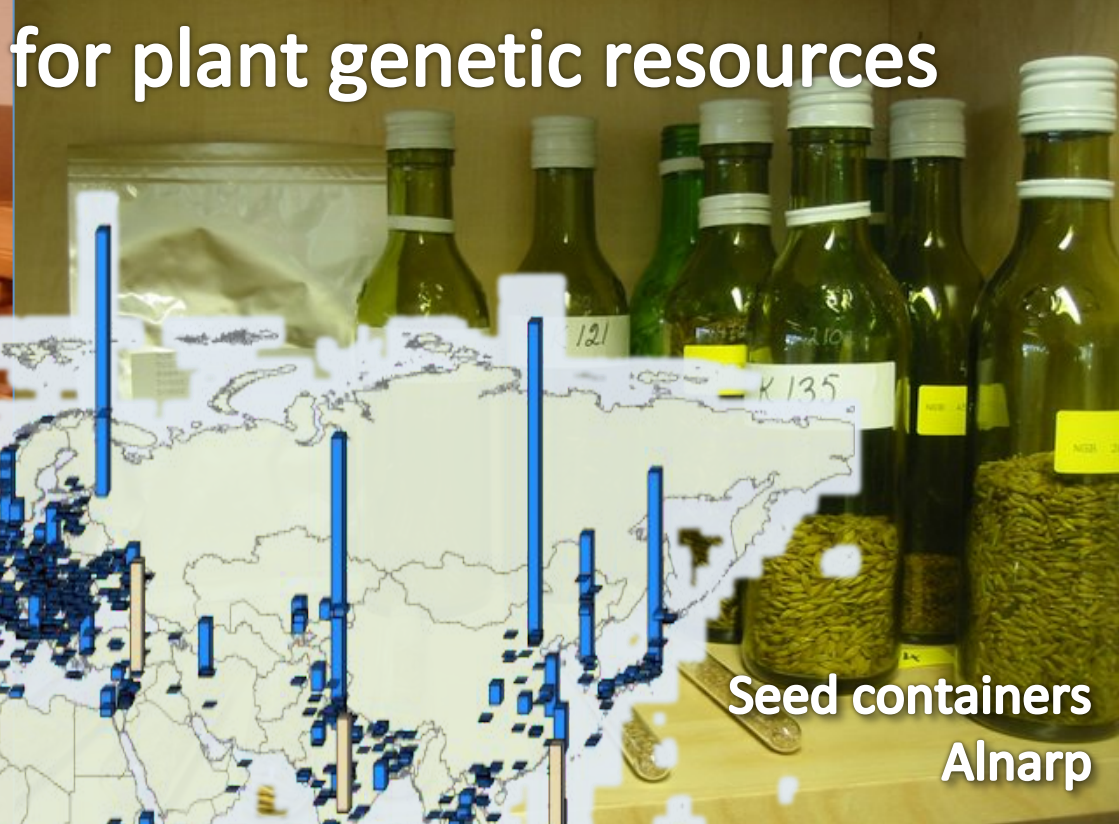
Utilizing genetic potential from the wild



Ex situ genebank collections for plant genetic resources



Seed drying
room
Alnarp



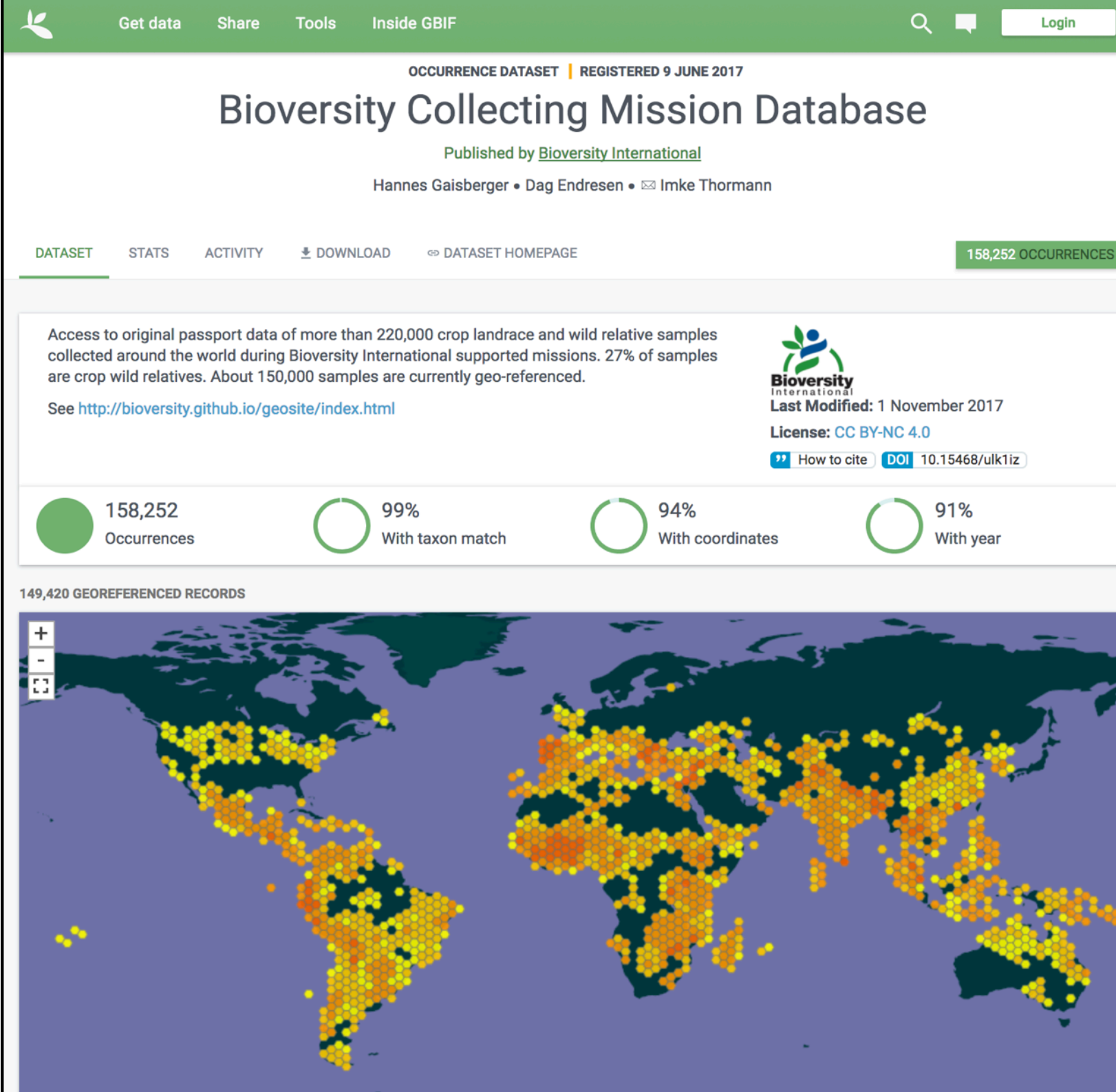
Seed containers
Alnarp



Svalbard Global
Seed Vault



Musa in vitro
Leuven



Biodiversity
Collecting Mission
Database
158 252 occurrences

[doi:10.15468/ulk1iz](https://doi.org/10.15468/ulk1iz)

(Dataset published in GBIF
June 2017)

[Get data](#)[Share](#)[Tools](#)[Inside GBIF](#)[Login](#)

OCCURRENCE DATASET | 22 FEBRUARY 2017

A global database for the distributions of crop wild relatives

Humberto Sotelo • GBIF Norway Helpdesk • Nora Patricia Castañeda-Alvarez • Dag Endresen

[DATASET](#)[TAXONOMY](#)[ORIGIN](#)[METRICS](#)[DOWNLOAD](#)[EXPLORE](#)

This dataset originally held 5 647 442 total records, where 34% of the records corresponded to germplasm accessions and 66% to herbarium samples. A total of 3 231 286 records had cross-checked coordinates (see Figure 2).... [more](#)

Publisher: [Centro Internacional de Agricultura Tropical \(CIAT\)](#)**License:** [CC BY 4.0](#)[Citation](#) [DOI](#) [10.15468/jyrthk](#)**3,403,811**

occurrences

**36%**

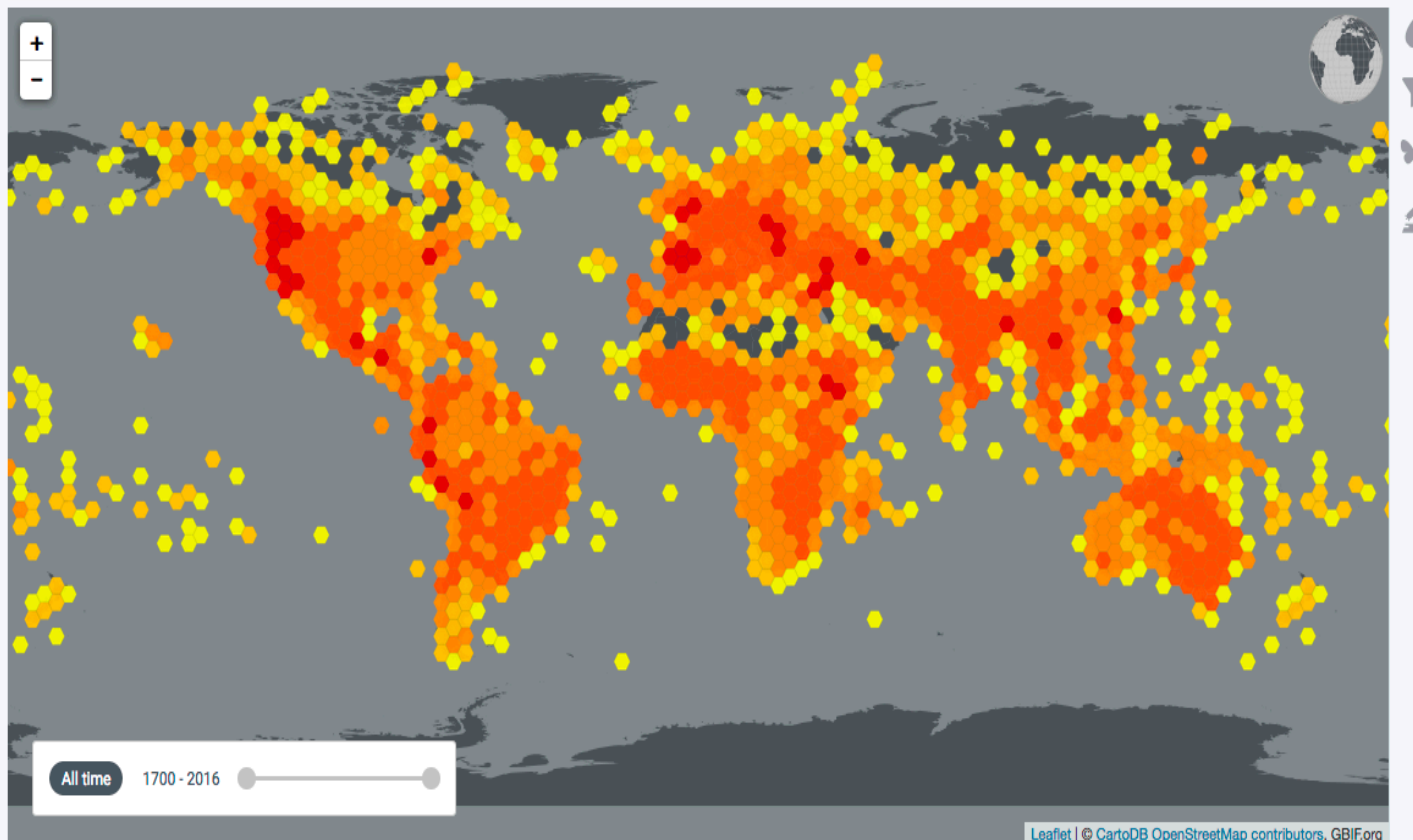
With coordinates

**34%**

With year

**99.9%**

With taxon match



The Global Crop Wild Relative **Occurrence** Database include data from hundreds of data sources – including GBIF

The CWR Database was published in GBIF (2016)
(excluding the data records originating from GBIF)

[doi:10.15468/jyrthk](https://doi.org/10.15468/jyrthk)



Genesys

ACCESSION MAP

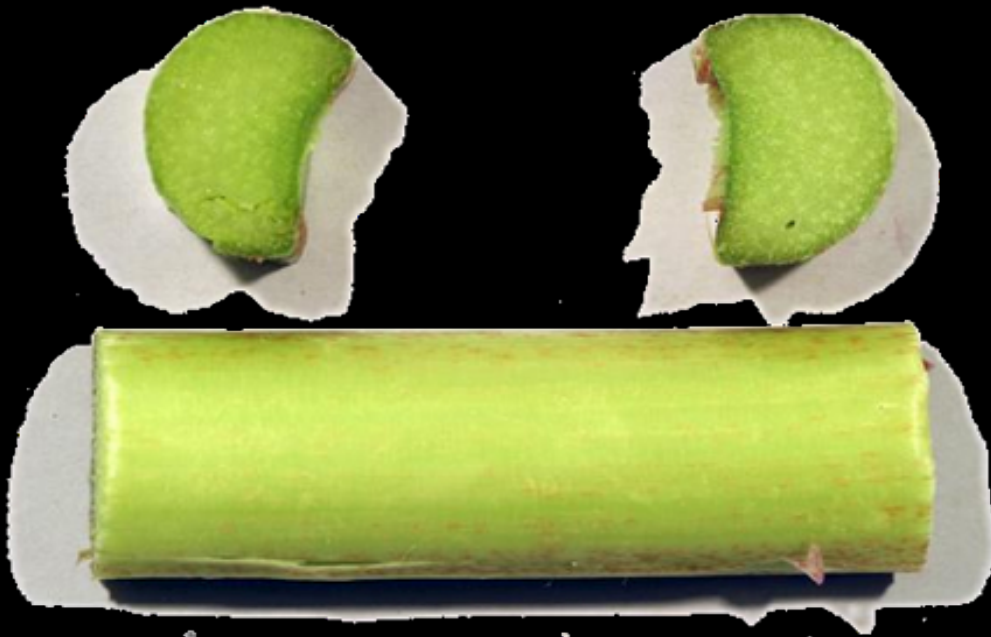
3,615,441 Accessions

435 Institutes

The map shows the collecting site of geo-referenced accessions.



Crop genetic diversity for agricultural traits



Rhubarb, *Rheum x hybridum* Murray, DKRHE43, by Gitte K. Bjørn



Field trials in Horsham Australia for cold tolerance in chickpea (*Cicer arietinum* L.). CC-by D. Endresen



Powdery Mildew, *Blumeria graminis*



Leaf spots *Ascochyta* sp.



Yellow rust *Puccinia striiformis*



Black stem rust *Puccinia graminis*



Apple, *Malus domestica* Borkh. Cultivar "Nanna" by Stein H Hjeltnes



Use case : Measurements for both events and occurrences in the same dataset

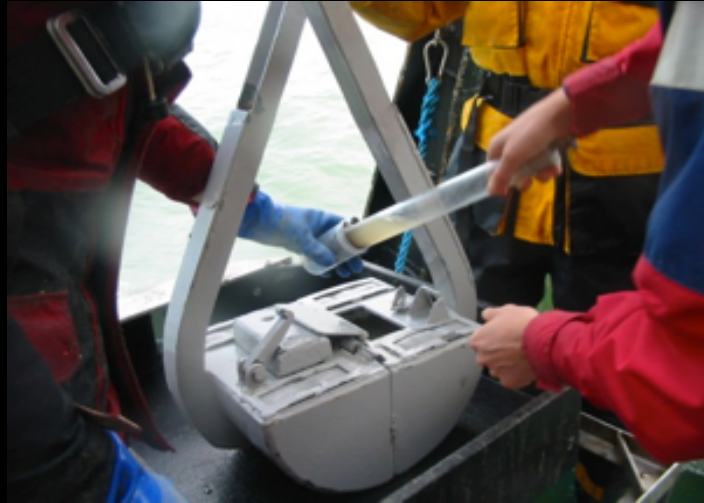
OBIS expeditions are sampling more data than only organism occurrences.

--> development of a new extended measurement and fact extension (for DwC-archives)

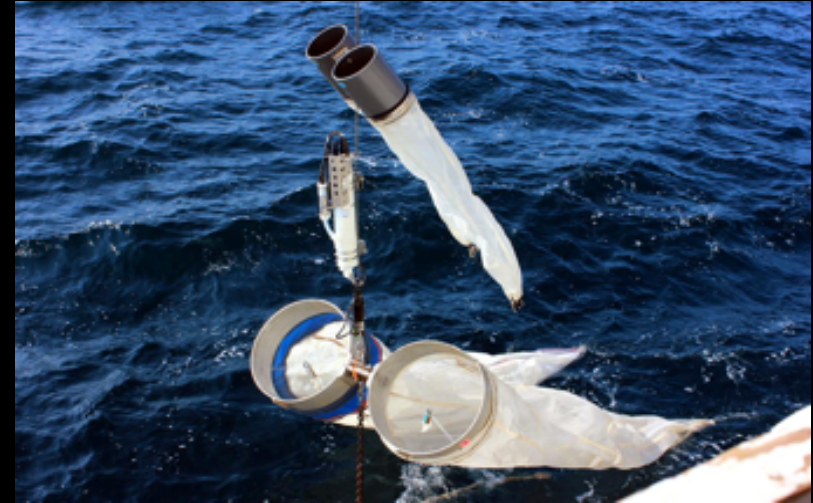
Sampling more data than just species occurrence



Water sample from a Niskin bottle



Core sample from a Van Veen grab



plankton net with CTD



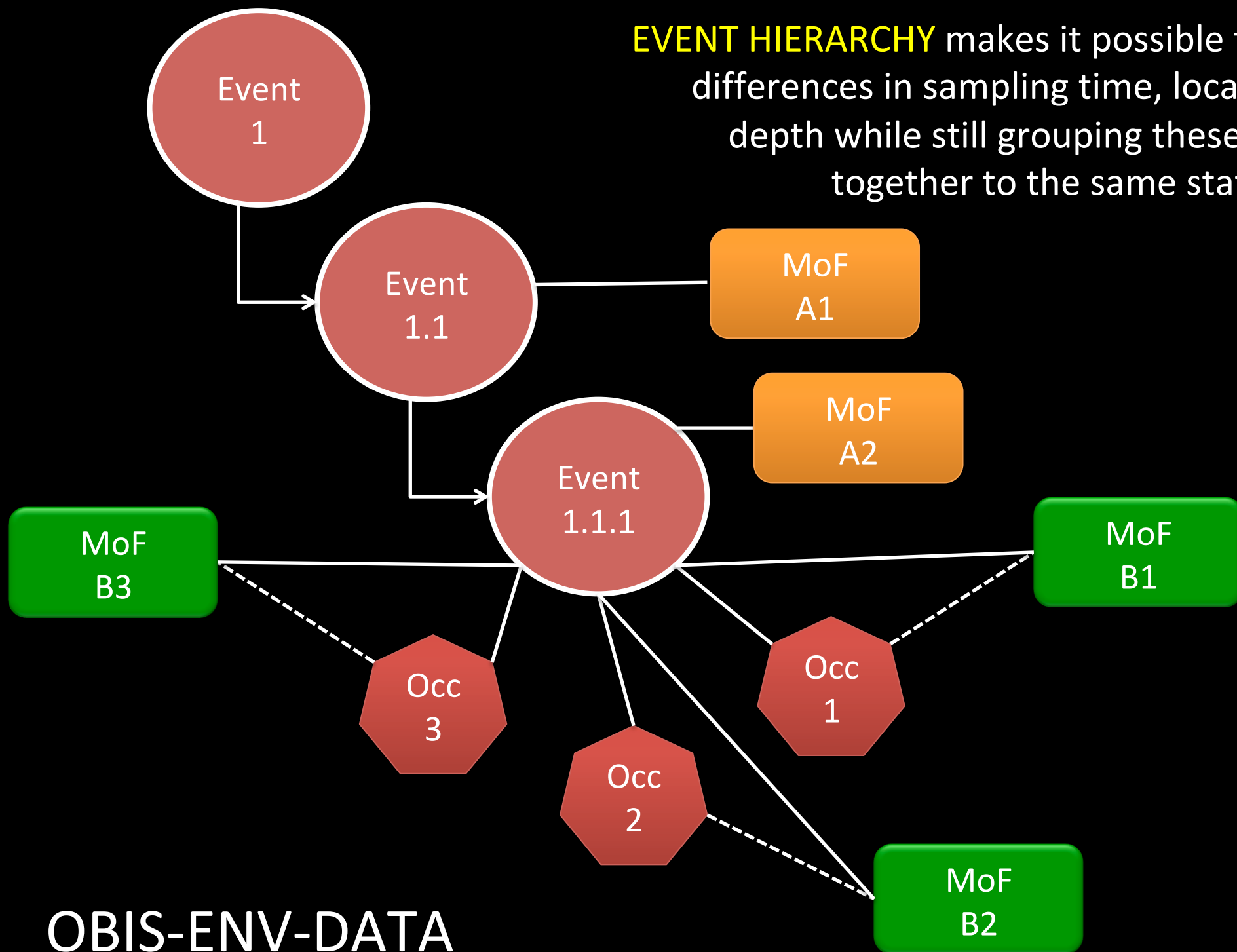
Video plankton recorder



OTN tags



EVENT HIERARCHY makes it possible to record differences in sampling time, location, and depth while still grouping these samples together to the same station visit.

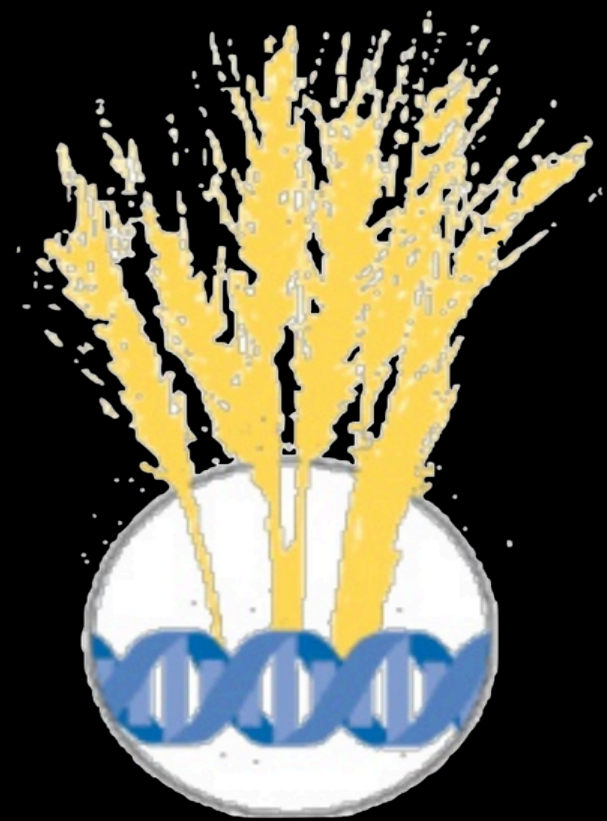


OBIS-ENV-DATA

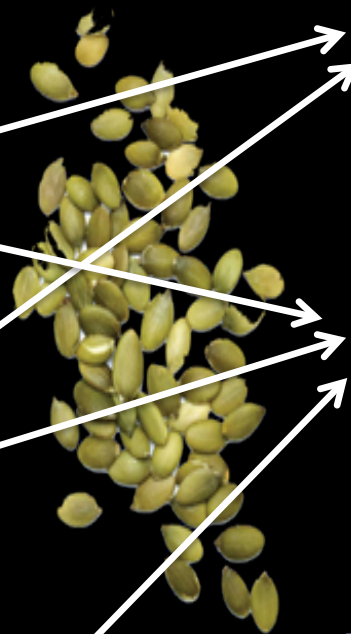
Proposing a new use case:

Measurement events as dwc:Event

Experiments with living specimens at a location and date that is different from the collecting event.

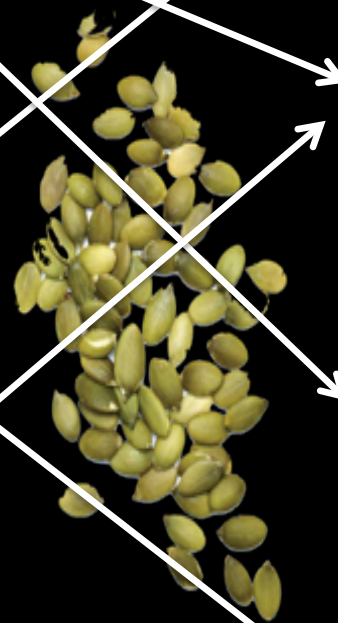


Collecting expeditions



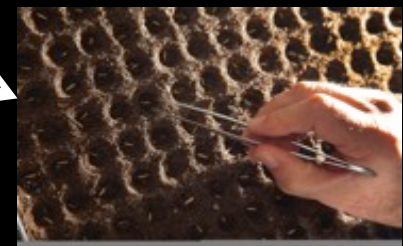
Seeds from
collecting trip
expeditions are
deposited in
genebank
collections.

Genebank, collections



Living seeds from
genebank
collections are
used for field or
laboratory
experiments.

Field/lab experiments



Germplasm material flow

- Material samples from many different collecting missions are deposited in different genebanks.
- Material from public genebanks are included in other genebank collections (and breeders collections).
- Material from different genebanks (and thus many different collecting events) are included in each field trial experiment.
- **Collecting Event $\neq$ Measurement Event (MoF)**

European meeting on C&E ("MoF")

- Aim to publish trait data (characterization and evaluation, C&E) from online crop portals.
 - European genebank meeting in **Bonn May 2009** (ECPGR ICT Group).
 - Developed guidelines & data exchange models.
 - Some harmonization with Darwin Core and DwC Measurement or Fact (MoF).
-
- DwC germplasm extension, released 2012.
 - GBIF data fitness for use in agrobiodiversity, 2016.
 - EURISCO C&E data models, implemented 2017.

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

May 2009



ECPGR C&E trait data model (1/2)

ECPGR trait data

- **Genotype** : Identity of the germplasm material
- **Trait** : Type of measurement, protocol
 - **Method** : How the trait was scored
- **Experiment** : Location and date of the experiment (*dwc:Event*)
- **Score**: Measurement value (*dwc:MoF*)

ECPGR C&E trait data model (2/2)

- DATASET
 - uploadercode (FK)
 - dataset_remark
 - (dataset_id)
- GENOTYPE
 - genotype_number (PK)
 - genotype_nicode
 - genotype_instcode
 - genotype_accenumb
 - genotype_genus
- TRAIT
 - trait_number (PK)
 - trait_name
 - trait_remark
 - trait_method
- EXPERIMENT
 - experiment_number (PK)
 - experiment_description
 - experiment_year
 - experiment_longitude
 - experiment_latitude
 - experiment_report
- SCORE
 - genotype_number (FK)
 - experiment_number (FK)
 - trait_number (FK)
 - **score**

Darwin Core Germplasm extension (1)

- dct = <http://purl.org/dc/terms/>
- dctype = <http://purl.org/dc/dcmitype/>
- dsw = <http://purl.org/dsw/>
- dwc = <http://rs.tdwg.org/dwc/terms/>
- geo = http://www.w3.org/2003/01/geo/wgs84_pos#
- g = <http://purl.org/germplasm/germplasmTerm#>
- gtype = <http://purl.org/germplasm/germplasmType#>
- mcpd = Alercia *et al.* (2015, 2012, 2001, 1997)
- epgris3 = van Hintum (2009)

Darwin Core Germplasm extension (2)

- Germplasm type (class terms)
 - dctype:Dataset
 - dwc:Taxon
 - gtype:GermplasmAccession (dsw:LivingSpecimen)
 - gtype:CollectingEvent (dwc:Event)
 - gtype:BreedingEvent [dwc:Event]?
 - gtype:AcquisitionEvent [dwc:Event]?
 - gtype:SafetyDuplication [dwc:Event]?
 - gtype:TreatyOrRegulation
 - gtype:MeasurementTrait
 - gtype:MeasurementTrial (dwc:Event)
 - gtype:MeasurementScore (dwc:MeasurementOrFact)

Darwin Core Germplasm extension (3)

- **gtype:GermplasmAccession** (see also: dsw:Specimen and dsw:LivingSpecimen)
 - dwc:occurrenceID
 - dwc:catalogNumber (mcpd:ACCENUMB)
 - *g:germplasmID* (**epgris3:GENOTYPE_NUMBER**)
 - *g:germplasmIdentifier* (mcpd:ACCENAME)
 - *g:biologicalStatus* (mcpd:SAMPSTAT)
 - *g:storageCondition* (mcpd:STORAGE)
 - dwc:otherCatalogNumbers (mcpd:OTHERNUMB)
 - dwc:occurrenceDetails
 - dwc:occurrenceRemarks

Darwin Core Germplasm extension (4)

- **gtype:MeasurementTrait** (measurement methodology, descriptor)
 - *g:measurementTraitID*
 - *g:measurementTraitIdentifier* (**epgris3**:TRAIT_NUMBER)
 - *g:measurementTraitName* (**epgris3**:TRAIT_NAME)
 - *g:measurementTraitCategory*
 - *g:measurementTraitScale*
 - *g:measurementTraitSource*
 - *g:measurementTraitRemarks* (**epgris3**:TRAIT_REMARK)
 - *dwc:measurementType*
 - *dwc:measurementMethod*

Darwin Core Germplasm extension (5)

- **gtype:MeasurementTrial** (experiment trial season, see **dwc:Event**)
 - *g:measurementTrialID*
 - *g:measurementTrialIdentifier* (**epgris3:experiment_number**)
 - *g:measurementTrialYear* (**epgris3:experiment_year**)
 - *geo:location*
 - *geo:lat* (**epgris3:experiment_latitude**)
 - *geo:long* (**epgris3:experiment_longitude**)
 - *geo:alt*
 - *g:measurementTrialReport* (**epgris3:experiment_report**)
 - *g:measurementTrialRemarks* (**epgris3:experiment_description**)

Darwin Core Germplasm extension (6)

- **gtype:MeasurementScore** (see **dwc:MeasurementOrFact**)
 - dwc:measurementID
 - dwc:measurementType
 - dwc:measurementValue
 - dwc:measurementAccuracy
 - dwc:measurementUnit
 - dwc:measurementDeterminedDate
 - dwc:measurementDeterminedBy
 - dwc:measurementMethod
 - dwc:measurementRemarks
 - *g:measurementByInstituteID*
 - *g:measurementGrowthStage*
 - dwc:occurrenceID
 - dwc:catalogNumber
 - *g:germplasmID*
 - *g:germplasmIdentifier*
 - *g:measurementExperimentID*
 - *g:measurementExperimentIdentifier*
 - *g:measurementMethodID*
 - *g:measurementMethodIdentifier*

Name	Last modified	Size	Description
 Parent Directory		-	
 GermplasmAccession.xml	2015-08-27 17:01	22K	
 MeasurementScore.xml	2015-08-27 17:01	14K	
 MeasurementTrait.xml	2015-08-27 17:01	6.5K	
 MeasurementTrial.xml	2015-08-27 17:01	6.2K	

dwc:MeasurementOrFact
g:MeasurementScore (2013)
 iode:eMeasurementOrFact (2016)

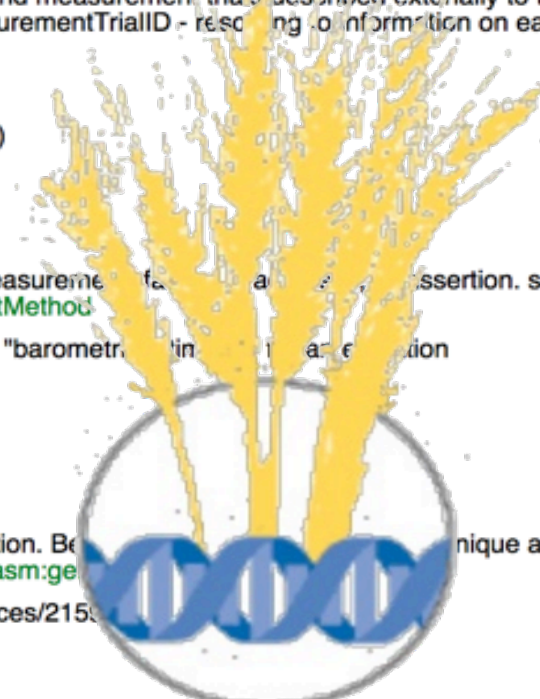
Core Extension ([index](#))

Trait measurement score (v20140515)

Title	Trait measurement score (v20140515)
Name	MeasurementScore
Issued	2014-05-15
Namespace	http://purl.org/germplasm/germplasmTerm#
RowType	http://purl.org/germplasm/germplasmTerm#MeasurementScore
Description	Trait measurements or the so-called Characterization and Evaluation (C and E) data for plant genetic resources for food and agriculture (PGRFA). The germplasm vocabulary term g.germplasmID is a persistent identifier for the germplasm material (genebank accession, specimen, cultivar, etc that was the subject of the measurement experiment. NB! please report stable and resolvable persistent identifiers for g.germplasmID term when linking the trait measurement score to a germplasm occurrence described in an external Darwin Core archive. When linking to trait measurement descriptors and measurement trials described externally to the Darwin Core archive, please include stable and resolvable persistent identifiers for g.measurementTraitID and g.measurementTrialID - resolvable information on each entity.
Link	http://purl.org/germplasm/germplasmTerm (This is an HTML view of the definition. Use View-Source to see the underlying XML.)

Properties

measurementMethod	A description of or reference to (publication, URI) the method or protocol used to determine the measurement fact and/or assertion. see also http://rs.tdwg.org/dwc/terms/index.htm#measurementMethod <i>Examples:</i> "minimum convex polygon around burrow entrances" for a home range area, "barometric pressure measurement" Qualname http://rs.tdwg.org/dwc/terms/measurementMethod Namespace http://rs.tdwg.org/dwc/terms/ Group Data Type Required false
germplasmID	A persistent identifier for a germplasm accession maintained as part of an ex situ genebank collection. Be a persistent identifier (PID). see also http://terms.gbif.org/wiki/germplasm:germplasmID <i>Examples:</i> URI, DOI, LSID or similar; http://data.gbif.org/occurrences/215 Qualname http://purl.org/germplasm/germplasmTerm#germplasmID Namespace http://purl.org/germplasm/germplasmTerm# Group GermplasmAccession Data Type string Required false



THE DARWIN CORE EXTENSION FOR GENEbanks OPENS UP NEW OPPORTUNITIES FOR SHARING GERMPLASM DATA SETS

DAG T.F. ENDRESEN*

Global Biodiversity Information Facility (GBIF), Copenhagen, Denmark

AND

HELMUT KNÜPFFER

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

Abstract. – Darwin Core (DwC) defines a standard set of terms to describe the primary biodiversity data. Primary biodiversity data are data records derived from direct observation of species occurrences in nature or describing specimens in biological collections. The Darwin Core terms can be seen as an extension to the standard Dublin Core metadata terms. The new Darwin Core extension for genebanks declares the additional terms required for describing genebank data sets, and is based on established standards from the plant genetic resources community. The Global Biodiversity Information Facility (GBIF) provides an information infrastructure for biodiversity data including a suite of software tools for data publishing, distributed data access, and the capture of biodiversity data. The Darwin Core extension for genebanks is a key component that provides access for the genebanks and the plant genetic resources community to the GBIF informatics infrastructure including the new toolkits for data exchange. This paper provides one of the first examples and guidelines for how to create extensions to the Darwin Core standard.

Keywords. – Darwin Core; Darwin Core extension; GBIF; genebank collections; germplasm; plant genetic resources.



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).

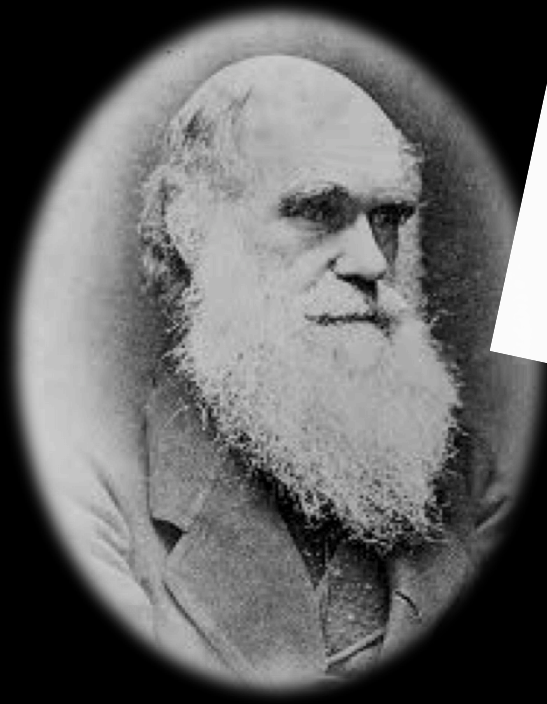
→ Upgrade data types to include trait data

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.



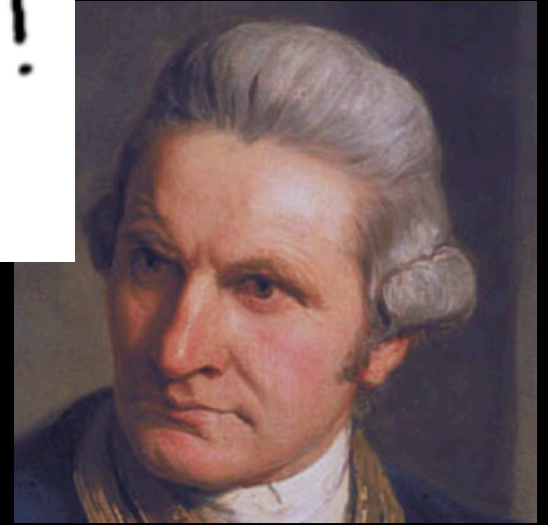
DwC Event Core + eMoF

- DwC Event Core + eMoF, as proposed by iOBIS, opens up promising possibilities for publishing also crop trait /C&E) data!!
 - Using the GBIF infrastructure for data exchange.
 - With collaboration on data exchange schema between thematic domains (terrestrial crops, marine, etc).



Hi!
/

Hi!
/



Hi!
/

