EURISCO, the European Search Portal for Plant Genetic Resources

Status quo & planned developments

EURISCO training workshop, 12th to 14th October 2016, Angers, France

Stephan Weise
12 October 2016

IPK GATERSLEBEN

BACKGROUND

Introduction I

- Development of European information system for plant genetic resources

  - Started in 2001 (EU project EPGRIS)

  - EURISCO: Search catalogue for *ex situ* collections; available online since 2003

  - National collections represented by National Inventories (NIs)

  - 43 countries involved (Nordic Countries → NordGen)

  - Network of National Focal Points (NFPs) links NIs ↔ EURISCO
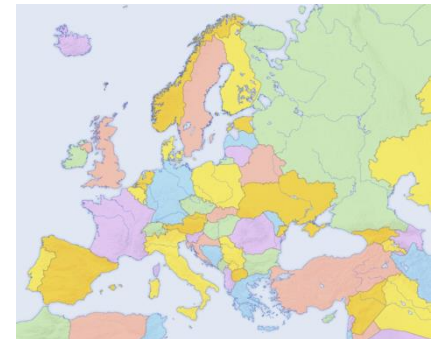
https://upload.wikimedia.org/wikipedia/commons/8/81/Europe_countries_map_2.png
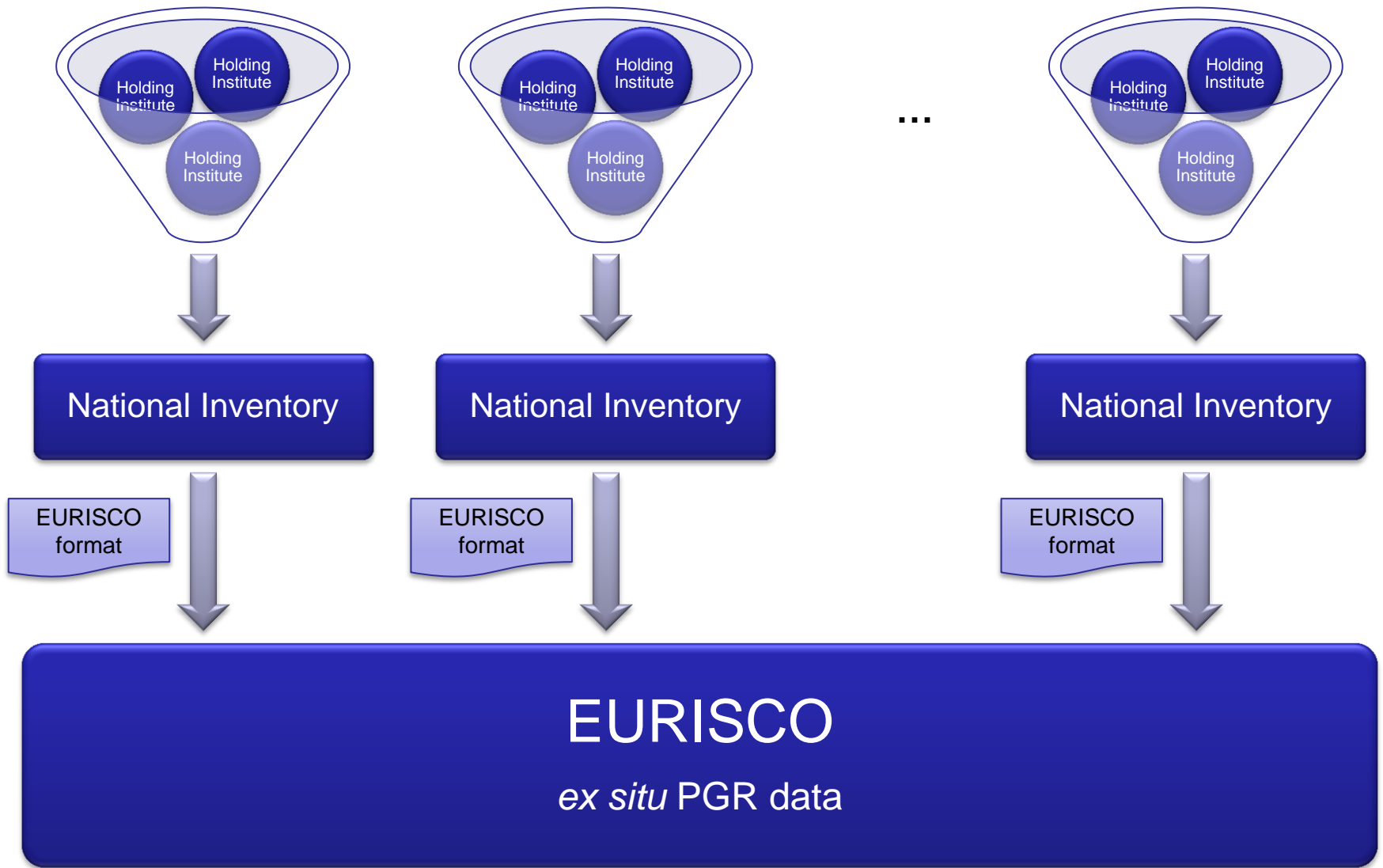
Introduction II

- Development of European information system for plant genetic resources (cont.)

  – Accession-level information system

  – Provides passport information about PGR maintained in Europe

  – Assists in meeting national obligations

    • Food and Agriculture Organization of the United Nations (FAO)

    • Convention on Biological Diversity (CBD)

    • International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA)

https://upload.wikimedia.org/wikipedia/commons/8/81/Europe_countries_map_2.png

Data flow

National Inventory

National Inventory

...

National Inventory

EURISCO format

EURISCO format

EURISCO format

EURISCO

ex situ PGR data

Contents of EURISCO

- 1,842,760 accessions

- 376 holding institutes

- 43 countries

- 6,233 genera (including synonyms, spelling variants)

- 41,649 species names (unique combinations genus + species, including synonyms)
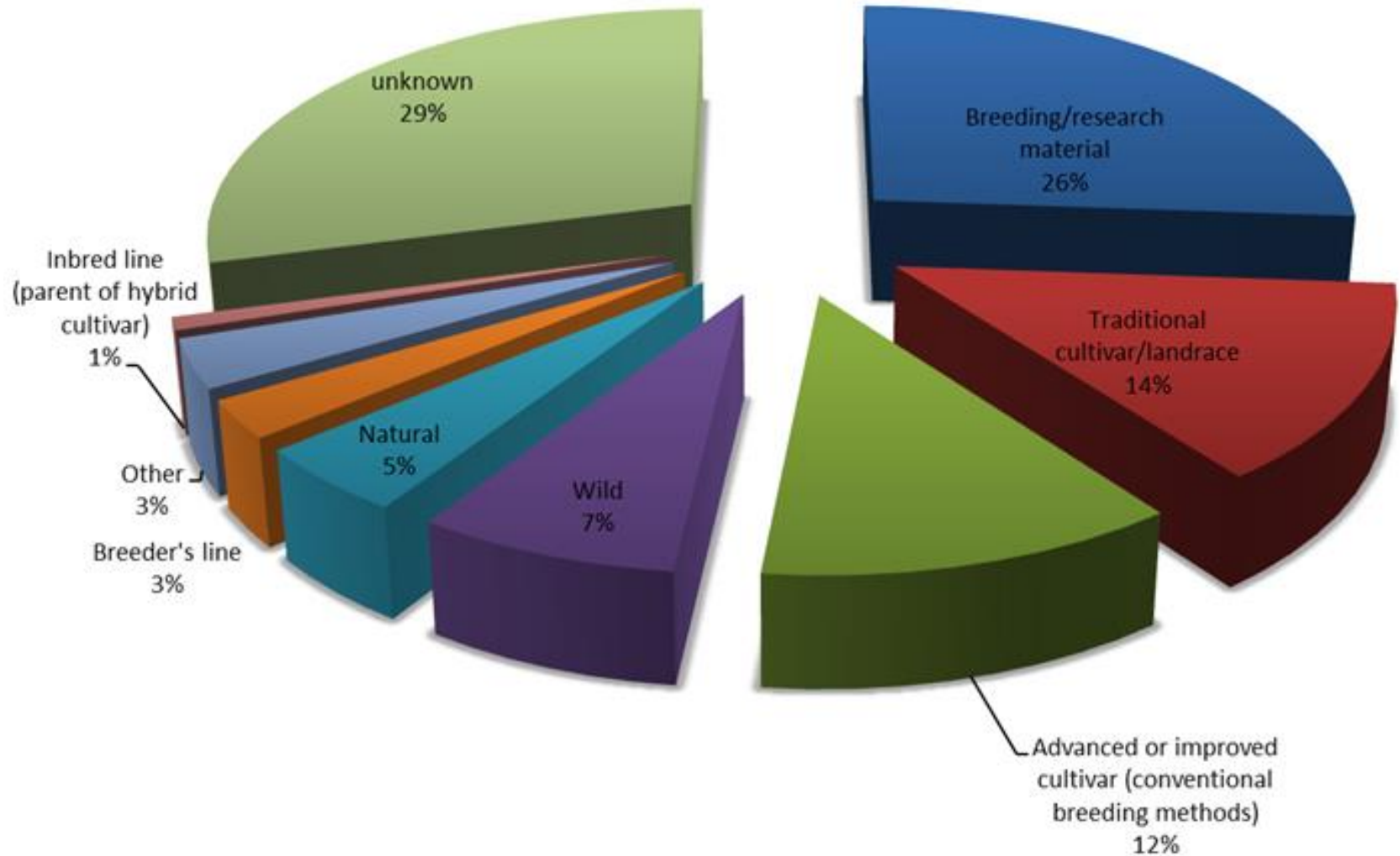
- 368,446 MLS accessions
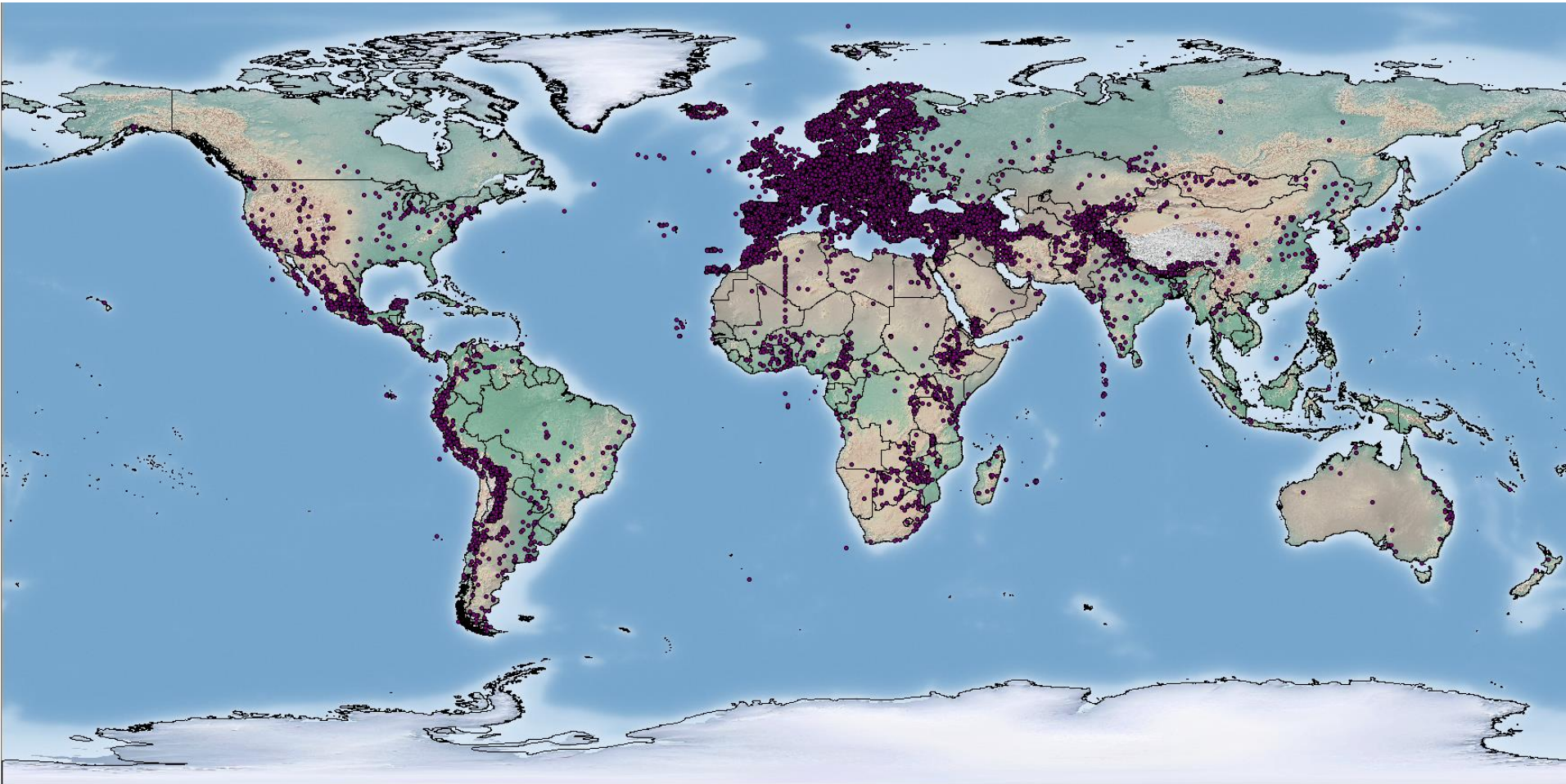
- 28,899 AEGIS accessions

as of 2016-09-07

Genera
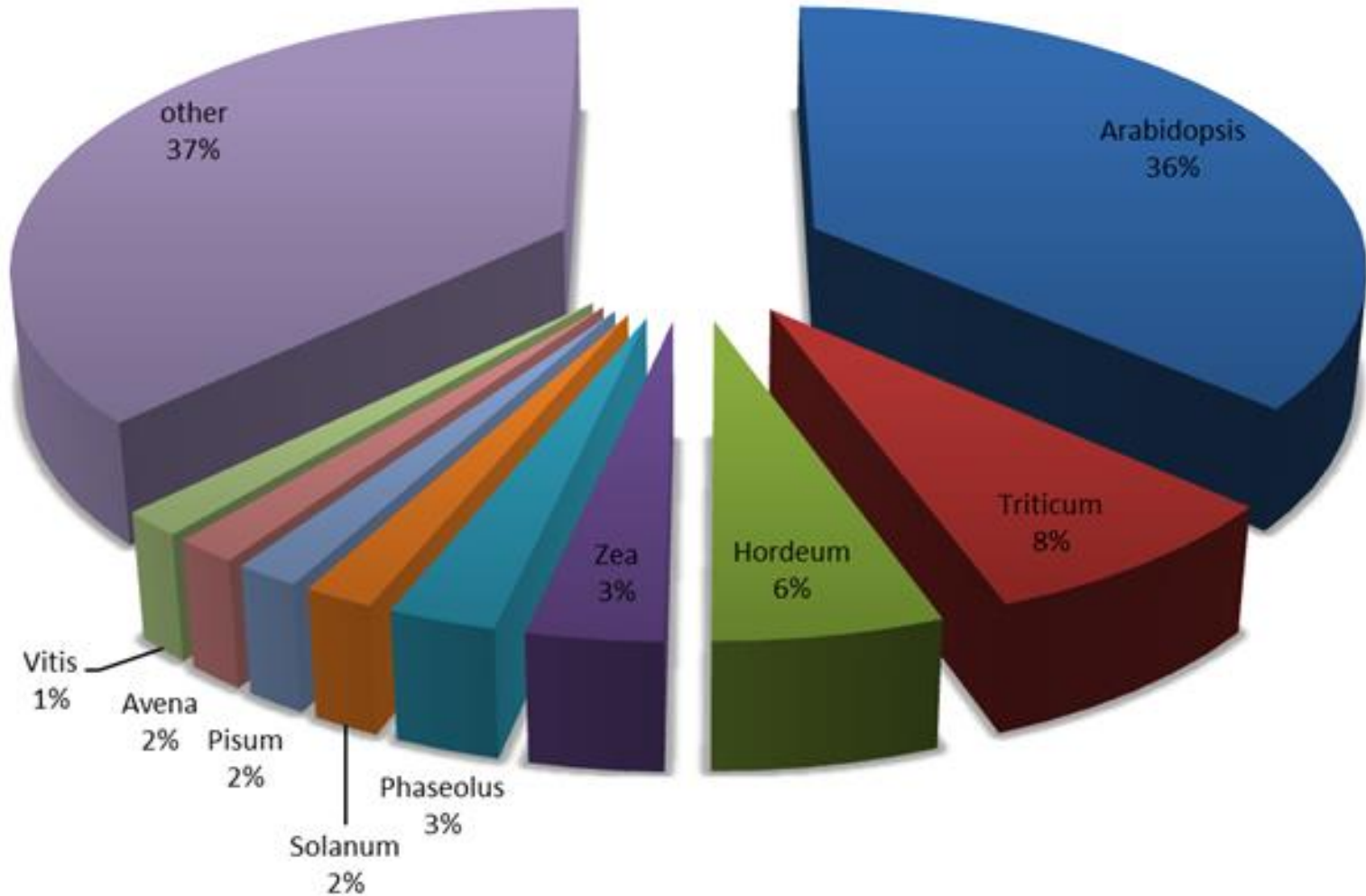
as of 2015-12-07

Biological status

as of 2015-12-07

Regions of origin

as of 2016-02-10

Data quality in EURISCO

- 587,391 accessions with collecting information

  - 91,808 different collecting sites

  - But only 188,456 accessions with coordinates (10% of all accessions; 32% of accessions with collecting information)

- 1,161,631 accessions with donor information

- 960,334 accessions with country of origin

  - 24 different geographic regions

as of 2016-09-07

EURISCO HOSTED BY IPK

IPK became host

- ## Since 2003:
  - EURISCO hosted by Bioversity International, Rome

- ## October 2012:
  - Request for tenders for hosting EURISCO

- ## March 2013:
  - IPK won the bid

- ## May and October 2013
  - Preparatory meetings in Gatersleben and Rome

- ## Since 15th April 2014:
  - Sub-contract between Bioversity International and IPK
  - Transfer of the EURISCO responsibilities to IPK (04-09/2014)

Challenges and decisions

- Challenges faced:

  – Outdated systems

  – Insufficient technical documentation

  – High costs for transfer of as-is status

- Decisions:

  → No transfer of existing solutions

  → New development from scratch

Technical tasks

- Analysis of the former web application

- Reengineering of database schema for web application

- Import (and cleansing) of current data set
  - Migration path MySQL → Oracle RDBMS

http://www.mysql.com

http://www.oracle.com

- New web application for searching EURISCO data
  - PL/SQL for functionality; APEX for rendering
  - First release with basic functionality published 09/2014
  - Continuously improved

- New web application for updating NI data
  - PL/SQL for functionality; APEX for rendering

New architecture: Overview
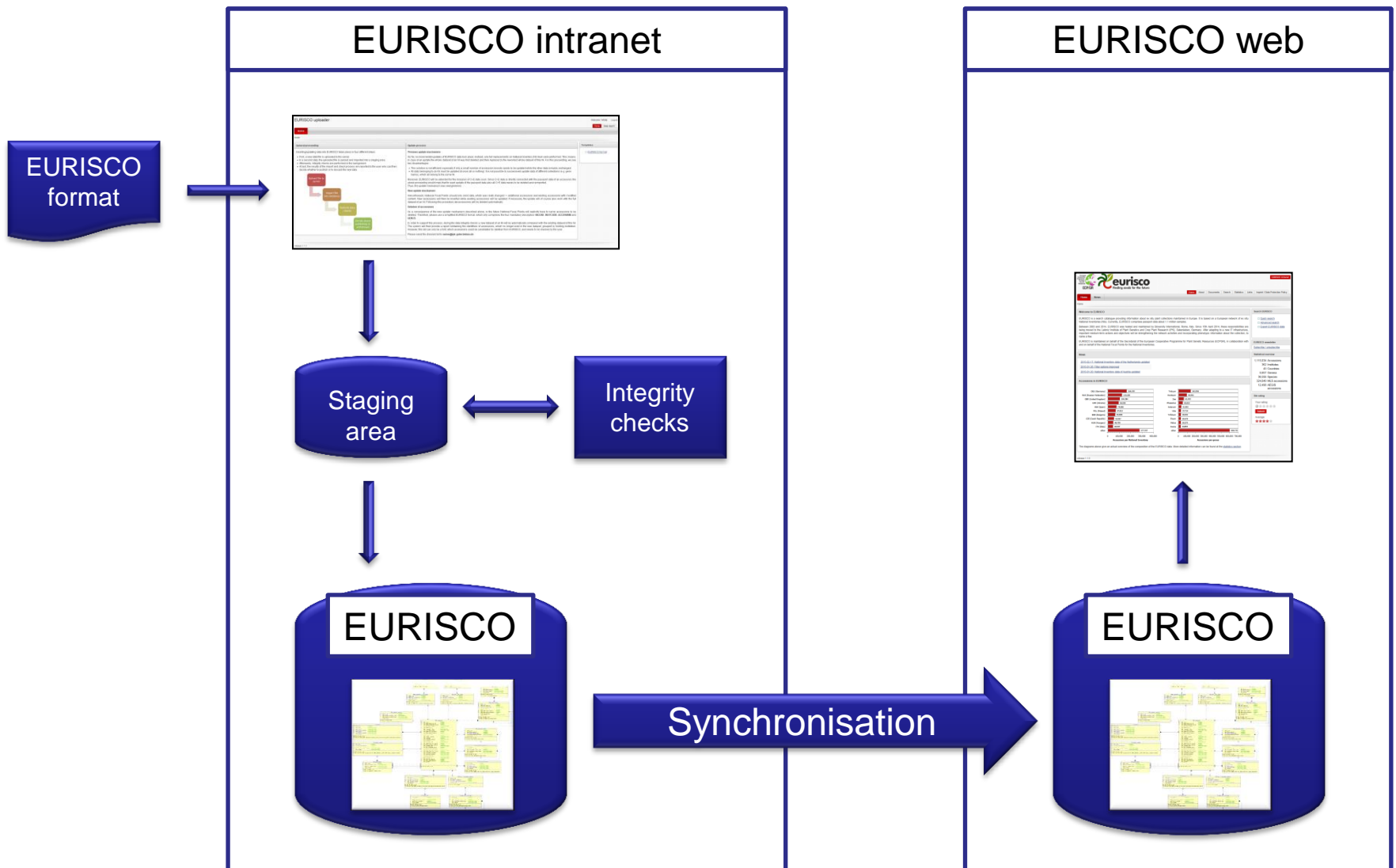
New architecture: DB schema features

- EURISCO intranet
  - 49 tables
  - 460 indexes
  - 90 triggers
  - 9 PL/SQL packages
  - 110 functions and procedures
  - 27 Java classes

- EURISCO web
  - 45 tables
  - 25 materialised views
  - 585 indexes
  - 6 PL/SQL packages
  - 25 functions and procedures
  - In-memory features

EURISCO WEB

New web application

Search form

Four standard searches
- Taxonomy
- Accession
- Biological status
- Collecting site

Search results

Accession details

Advanced search form

User specific export – by species

User specific export – by NI

EURISCO INTRANET

EURISCO intranet

- Upload component for National Inventory Focal Points

- Standardised data exchange formats

    - Passport data

    - Phenotypic (characterisation and evaluation) data

- Details will follow during the hands-on sessions

CHARACTERISATION & EVALUATION DATA

C&E data: Extending EURISCO

- EURISCO backend completely extended for C&E data
  - Exchange format
  - Database schema
  - Upload tools
  - Data integrity checks

- Extension of EURISCO web interface for C&E data – first version available
  - Wizard-based searches for
    - Genus
    - Species and trait
    - Experiment
    - Trait

C&E data: Genus

C&E data: Species and trait

C&E data: Experiment

C&E data: Trait

Filter C&E data by trait

The report below lists the definitions of all phenotypic traits, which are currently available in EURISCO. Please use the search bar below to define filters.

Q ∨ [   ] **Go** Rows 10 [∨] **Actions** ∨

1 - 10 of 1214 (>)

Trait Name	Trait Remark	Trait Method	Trait Group	Details
Petiole and/or midvein enlargement	-	3=narrow, 5=intermediate,7=enlarged (See IPGRI descriptors Brassica and Raphanus 1990 4.2.27)	C&E data (not further specified)	used by experiment(s)
Silique angle	-	1=0°; 2=0-22.5°; 3=22.5°; 4=22.5-45°; 5=45°; 6:45-67.5°; 7=67.5°	C&E data (not further specified)	used by experiment(s)
Leaf anthocyanin content	-	1 = very weak, 3 = weak, 5 = medium, 7 = strong, 9 = very strong	C&E data (not further specified)	used by experiment(s)
Time period until marketable maturity	-	(1=very early, ..., 9=very late)	C&E data (not further specified)	used by experiment(s)
Bremia lactucae	-	Resistance to B. lactucae race Bl 5, scale:1 = susceptible, 9 = resistant	C&E data (not further specified)	used by experiment(s)
Bremia lactucae	-	Resistance to B. lactucae race Bl 11, scale:1 = susceptible, 9 = resistant	C&E data (not further specified)	used by experiment(s)
Bremia lactucae	-	Resistance to B. lactucae race Bl 12, scale:1 = susceptible, 9 = resistant	C&E data (not further specified)	used by experiment(s)
Bremia lactucae	-	Resistance to B. lactucae race Bl 14, scale:1 = susceptible, 9 = resistant	C&E data (not further specified)	used by experiment(s)
Leaf anthocyanin distribution	-	1 = localised, 2 = entire	C&E data (not further specified)	used by experiment(s)
Bremia lactucae	-	Resistance to B. lactucae race Bl 25, scale:1 = susceptible, 9 = resistant	C&E data (not further specified)	used by experiment(s)

1 - 10 of 1214 (>)

0.03 s

eurisco — Finding seeds for the future

IPK GATERSLEBEN

C&E data: C&E scores
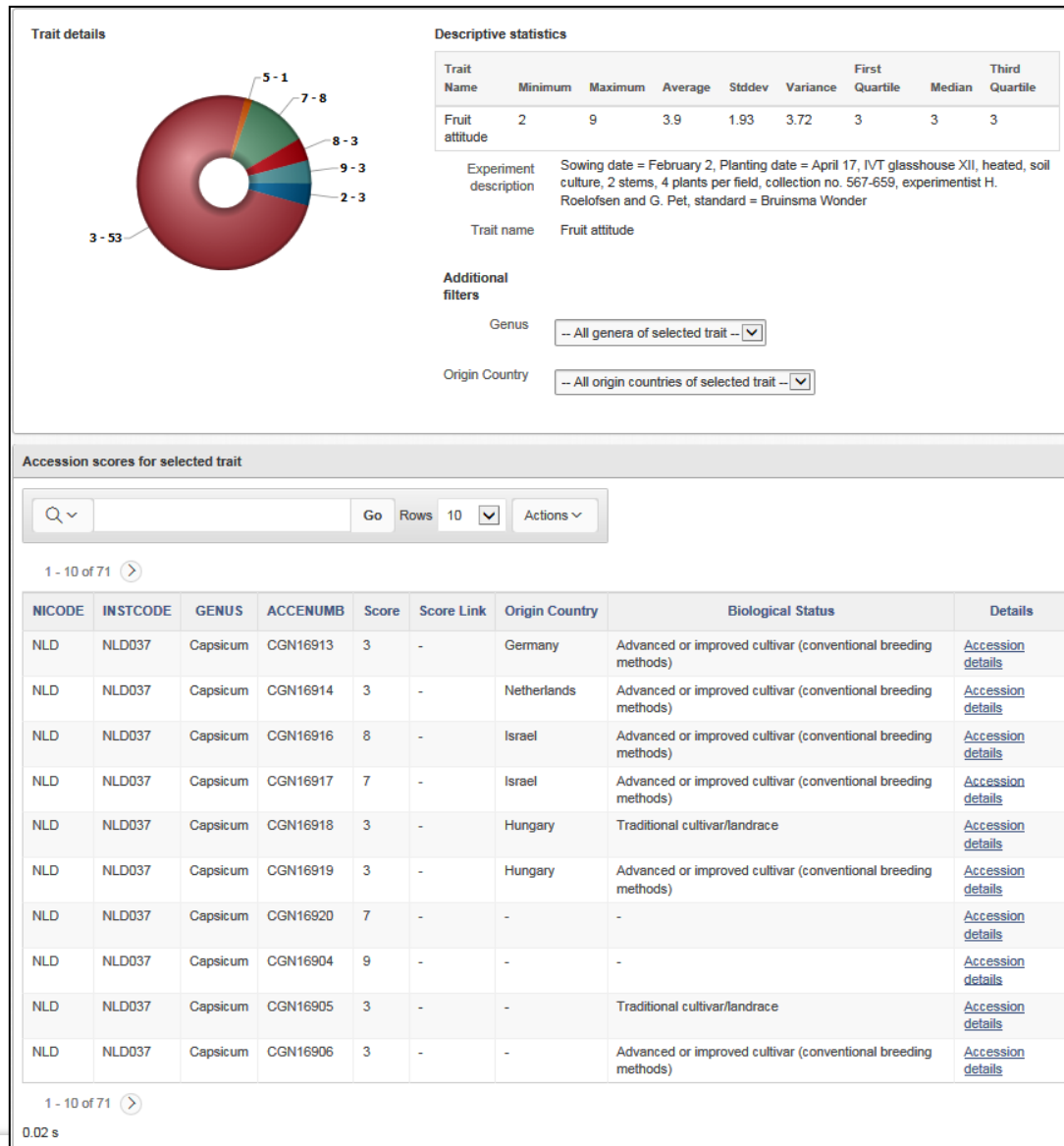
AEGIS

AEGIS

- AEGIS - A European Genebank Integrated System

- ECPGR initiative for improving the coordination of

  - Conservation and management of PGRFA

  - Access to PGRFA

- Aims:

  - Conservation of genetically unique and important accessions

  - Making materials available for breeding and research

  - Safe long-term conservation (with common agreed standards)

  - Reduction of redundancy

  - Clarification of responsibilities for conservation

AEGIS data in EURISCO

- No physical collection → "virtual" genebank
- AEGIS accessions labelled in EURISCO
  - Including tracking of AEGIS status
- 28,899 AEGIS accessions
- 34 member countries

AEGIS status auditing

NICODE	INSTCODE	AEGISSTAT old	AEGISSTAT current	Change timestamp	DML operation	No of accessions
SVK	SVK001	-	1	2016-08-26	Existing EURISCO accession labelled as part of AEGIS	5 (click for details)
SVK	SVK001	0	1	2016-08-26	Existing EURISCO accession labelled as part of AEGIS	47 (click for details)
LVA	LVA009	0	1	2016-06-10	Existing EURISCO accession labelled as part of AEGIS	12 (click for details)
POL	POL003	-	1	2016-06-02	Existing EURISCO accession labelled as part of AEGIS	149 (click for details)
SVK	SVK001	-	1	2016-03-01	Existing EURISCO accession labelled as part of AEGIS	247 (click for details)
LVA	LVA009	0	1	2016-02-15	Existing EURISCO accession labelled as part of AEGIS	9 (click for details)
ROU	ROM007	-	1	2016-02-02	Existing EURISCO accession labelled as part of AEGIS	196 (click for details)
EST	EST001	0	1	2016-01-12	Existing EURISCO accession labelled as part of AEGIS	15 (click for details)
POL	POL003	-	1	2016-01-11	Existing EURISCO accession labelled as part of AEGIS	152 (click for details)
BGR	BGR001	-	1	2015-12-08	Existing EURISCO accession labelled as part of AEGIS	261 (click for details)
NLD	NLD037	1	-	2015-11-06	Existing AEGIS accession completely removed from EURISCO	9 (click for details)
GBR	GBR016	-	1	2015-10-09	Existing EURISCO accession labelled as part of AEGIS	7 (click for details)
GBR	GBR016	0	1	2015-10-09	Existing EURISCO accession labelled as part of AEGIS	2517 (click for details)
NGB	SWE054	0	1	2015-09-02	Existing EURISCO accession labelled as part of AEGIS	3597 (click for details)
NGB	SWE054	-	1	2015-09-02	Existing EURISCO accession labelled as part of AEGIS	111 (click for details)

AEGIS Member Countries

AEGIS Member Countries with European accessions
1   2500+
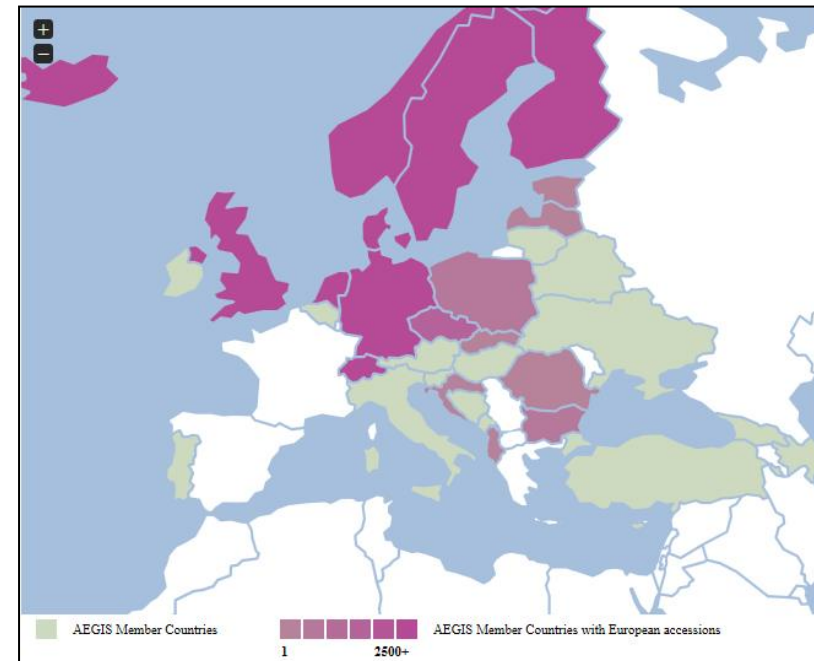
as of 2016-09-08

FUTURE OF EURISCO

More data

- Data quantity
  - Inclusion of additional passport data
    - ~250,000 accessions (van Hintum, 2014)

- Data quality
  - Increase frequency of updates
    - Avg. age 1.16 years
    - Oldest 10% avg. 4.96 years
      (van Hintum, 2014)

The documentation of Plant Genetic Resources in Europe

Theo van Hintum, Centre for Genetic Resources, The Netherlands (CGN)

April, 2014

This is a personal view as input for the discussions at the workshop of the Documentation and Information Working Group 'Tailoring the Documentation of Plant Genetic Resources in Europe to the Needs of the User' to be held 20-22 May '14 in Prague, Czech Republic. It provides a conceptual background to the issues to be discussed and formulates a number of draft resolution.

Focus

The current *ex situ* PGR documentation landscape consists of very many 'data sources', i.e., documentation systems of germplasm collections. The passport information from these systems is to a large extent collected by National Focal Points (NFPs) to create National Inventories (NIs). The information in NIs is expected to be regularly uploaded to EURISCO that thus should always provide an overview of the content of these NIs and of the genetic resources in Europe. Parallel to this data-flow and depository, Central Crop Data Bases (CCDBs) have been created since the early days of the ECPGR Crop Working Groups, to collect passport and sometimes additional data on a crop specific basis, however many of these databases do not appear to be up to date or to provide information and features that are not already present in EURISCO.

On a global level, the GeneSys initiative tries to create an entry point to data on all PGR maintained in the world. Data providers and database managers operate in an environment with rapidly evolving technologies and policies. It is therefore expected that developments in information technology, sequencing technology and policies on access and benefit sharing (ABS) will have a large impact on PGR documentation.

The combined data from EURISCO and 46 accessible Central Crop Data Bases originate from 506 data sources in 43 countries. The largest data source, according to EURISCO is IPK in Germany with 128k accessions, followed by the Vavilov Institute in Russia with 123k accessions[1]. The number of accessions currently documented in EURISCO is 1065766, and the total number of accessions in Europe is expected to be around 1.3 million.

In this document some important issues related to the current situation and developments regarding the documentation of PGR in Europe will be discussed, and resolutions will be formulated. These resolutions are aimed at either the EURISCO Management (currently, since beginning of 2014 at IPK) that coordinates the network of NFPs and runs the EURISCO database and web interface, or the ECPGR Doc/Info Working Group, that acts as the steering committee for EURISCO and oversees the documentation activities of ECPGR.

The issues that are to be discussed include: quality and coverage of the passport data in EURISCO, characterisation and evaluation data in EURISCO, the future of CCDBs in relation to EURISCO, PGR Portals, the relation EURISCO – GeneSys, the relation

1

eurisco
Finding seeds for the future

IPK
GATERSLEBEN

Better data

- Data quality
  - Improve taxonomic backbone of EURISCO
    - Management of taxon synonyms
    - Improvement of checks during import
      - GRIN, Catalogue of Life webservice

  - Increase completeness of information
    - Often limited information about certain accessions
    - Some descriptors only sparsely populated
      - SPECIES: 97%
      - BREDCODE: 40%
      - ELEVATION: 11%

http://www.ars-grin.gov

http://www.catalogueoflife.org

Challenges: Sea food?

Taxonomy

Genus **Medicago**
Species **murex**

Acquisition/storage

Acquisition Source **Roadside**

Collection

Collecting Institute Code
Collecting Date **1987**
Collecting Latitude **41.4264**
Collecting Longitude **-9.0831**
Collecting Elevation **635**
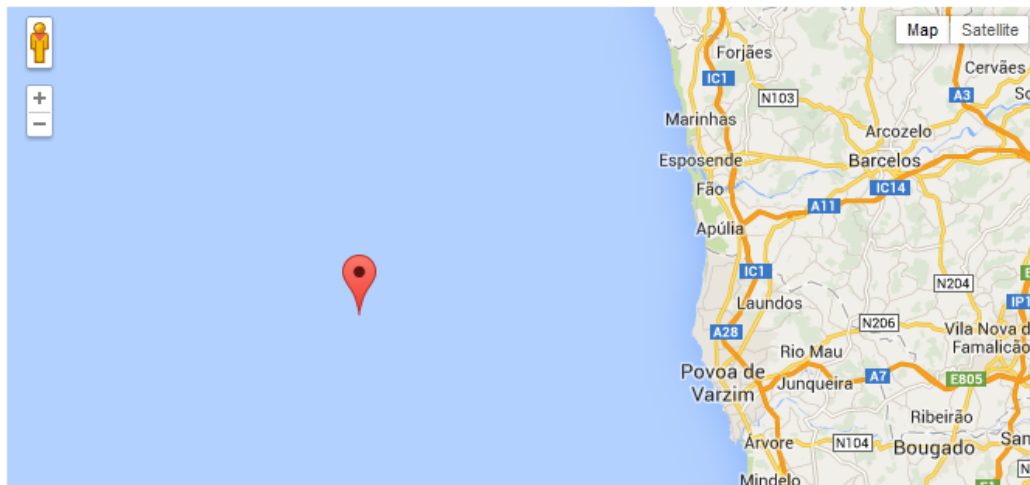
Challenges: Inconsistencies

Challenges: Babylonian aspects

Accession Na

OMAR

ELMAR

òåë-800

òåë-135

òåë-798

Accession Name

nagykallo 47ý52n, 21ý51'e

Collecting Locality , Slobozia Mare, Region Vulconețti , Slobozia Mare

Collecting Elevation

Collecting Locality   �. ���������, ������������� ���

Increase functionality

- Improve import mechanism
  - Nested user accounts
    - Data upload also by holding institutes
    - Check and approval by NFPs

- Improve public web interface
  - Additional reports and download facilities
  - Extend filter possibilities by additional fields
  - Enable case insensitive search for taxa (for standard searches)
  - Provide a full-text search

Extend for *in situ* data

- Will be documented in EURISCO

- Sufficient specification needed

- Data exchange formats need to be agreed

Additional means of access

- Web services

    - Additional means of access

        - Data exchange with Genesys, GBIF etc.

    - Improvement of upload mechanism

        - Selective updates of certain accessions

- Mobile version

DISSEMINATION

Dissemination

- Presentations on several workshops

- Journal article

  - Nucleic Acids Research (database issue), DOI: 10.1093/nar/gkw755

- EURISCO posters on several conferences

- Regularly short information in ECPGR bulletin

- EURISCO newsletter twice a year

- Preparation of project proposals (H2020 and others)

- In preparation: Update of fact sheets

Acknowledgements

- Bioversity International
    - Sónia Dias
    - Simone Mori
    - Milko Škofič
    - Dario Valori

- Global Crop Diversity Trust
    - Matija Obreza

- IPK Gatersleben
    - Andreas Graner
    - Helmut Knüpffer
    - Markus Oppermann
    - Uwe Scholz

- ECPGR
    - Theo van Hintum (Doc&Info group)
    - Lorenzo Maggioni, Elinor Lipman, Lidwina Koop, Jan Engels (ECPGR Secretariat)
    - ECPGR Executive Committee

THANK YOU FOR YOUR ATTENTION