Strengthening comprehensive information system for plant genetic resources

Renata Kowalik

Plant Breeding and Acclimatization Institute – National Research Institute (IHAR-PIB)

Workshop of the Documentation & Information Working Group
18-19 September 2024, Tallinn, Estonia

Background – Plant Genetic Resources Strategy for Europe

Objectives by 2030:

- National Focal Points (NFPs) are supported to collect the passport data of all PGR genebanks in their country and upload them to EURISCO
- Providing publicly available quality phenotypic data to EURISCO
- EURISCO and NFPs applies the FAIR (Findable-Accesible-Interoperable-Reproducible) principles
- EURISCO's data governance and management are improved to reach high standard →EURISCO becomes a trusted European and Global open-access database repository

Targets:

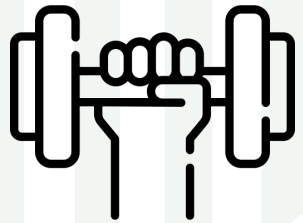
1. **EURISCO network of NFPs is optimally supported.**
2. **High-quality passport data of all European *ex situ* collections + progressively extended of *in situ* CWR populations** and appropiate on-farm landraces data.
3. **NFPs assure access to** all publicly available **quality phenotypic data related to the conserved PGR**. Access provided via inclusion in EURISCO.
4. **Genebanks and other PGR holders** can **improve their data management system practices** (access, use of tools, resources, servives, adopting FAIR principles, part of the open data community).
5. **Data** in EURISCO and associated IT infrastructure are **compliant with FAIR principles**.
6. EURISCO becomes a trustable repository in the arena of European and global open-access databases with high standards.

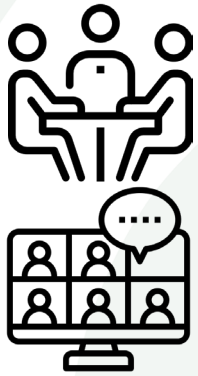
1. Support: EURISCO → National Focal Points

Training, training, and one more training

→ very usefull for new users/new staf (documentation specialists, curators in genebanks, new members of ECPGR WG, etc.)

→ discussion





- In-person meeting
- Online meeting
- Hybrid meeting

→ <u>feedback</u> during/after training

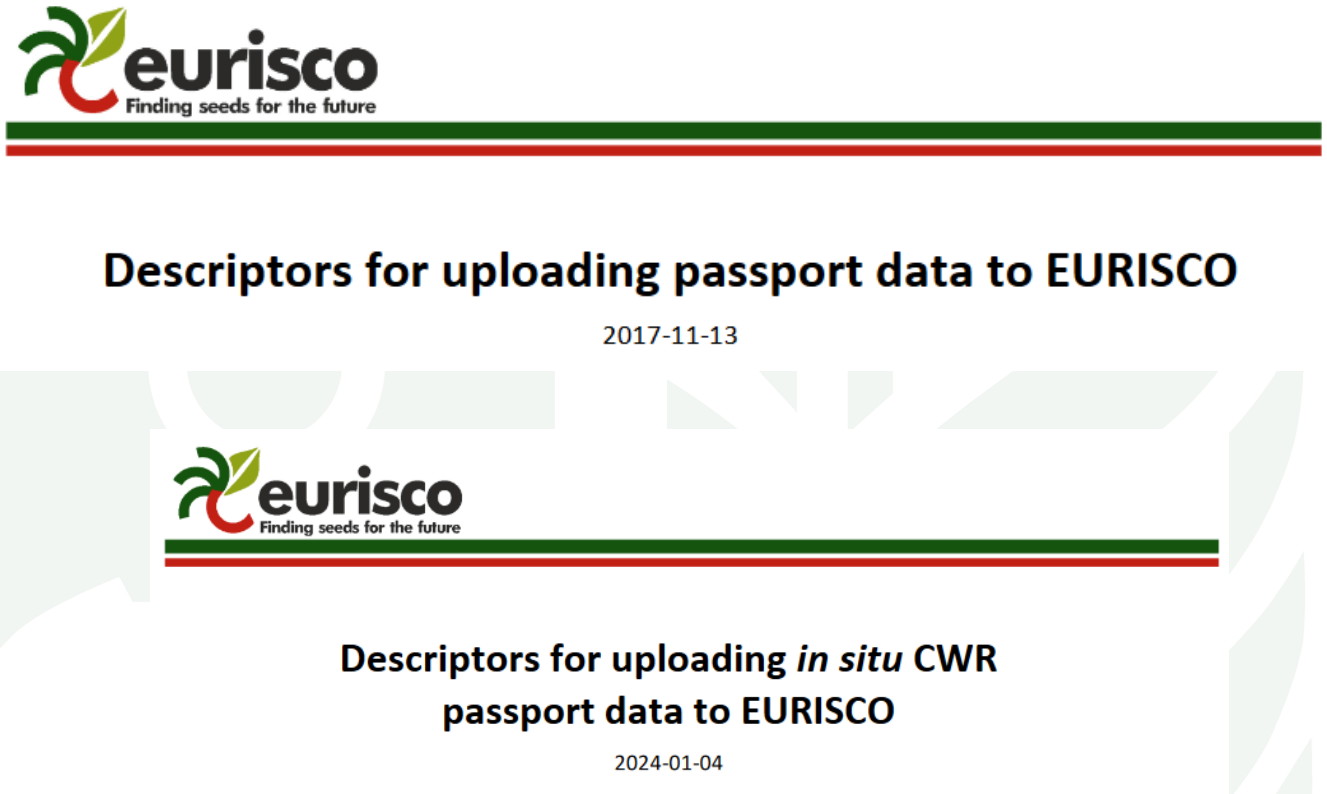
2. High-quality passport data of *ex situ* collections

How to measure data quality?

- Completeness
- Correctness
- Reliability

Completeness

- 45 passport data descriptors for *ex situ* collections
  - 4 mandatory descriptors (NICODE, INSTCODE, ACCENUMB, GENUS)
  - 41 optional descriptors
- 28 passport data descriptors for *in situ* CWR
  - 4 mandatory descriptors (NICODE, INSTCODE, ACCENUMB, GENUS)
  - 24 optional descriptors



Completeness

- Passport data completeness index (PDCI)
  - Van Hintum et al., Quality indicators for passport data in *ex situ* genebanks, *Plant Genetic Resources,* 9(3):478-485, 2011
  - Index between 0 (low quality) and 10 (high quality)

Independent of the population type		Depending on the population type	
INSTCODE	DONORCODE	ORIGCTY	COLLNUMB
ACCENUMB	DONORNUMB	COLLSITE	COLLCODE
GENUS	OTHERNUMB	LATITUDE	COLLNAME
SPECIES	DUPLSITE	LONGITUDE	BREDNAME
SPAUTHOR	STORAGE	ELEVATION	
SUBTAXA	REMARKS	COLLDATE	
SUBAUTHOR	DONORNAME	BREDCODE	
CROPNAME	DUPLINSTNAME	ANCEST	
ACQDATE	ACCEURL	COLLSRC	
SAMPSTAT	MLSSTAT	ACCENAME	

Completness – how to help?

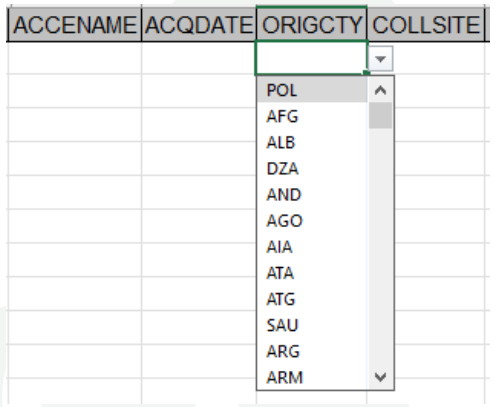
- prepare a template for data providers/translate descriptors
- template with colours
  → example below: divide descriptors for groups: mandatory - red, main information about accession for information system – yellow, collected accession – light grey, bred accession – light green, aquired from other institution – white, etc.)
- fill in as many fields as possible (if possible)

INSTCODE	ACCENUMB	COLLNUMB	COLCODE	GENUS	SPECIES	SPAUTHOR	SUBTAXA	SUBTAUTHOR	ACCENAME	ACQDATE	ORIGCTY	COLLSITE	LATITUDE	LONGITUDE	ELEVATION	COLLDATE	BREDCODE	SAMPSTAT	ANCEST	COLLSRC	DONORCODE	DONORNUMB	OTHERNUMB
Kod instytucji: Młochów kod POL047	Jeśli został nadany	Wypełnić, jeśli obiekt został zebrany					Obligatoryjne, jeśli istnieje dokładniejsze oznaczenie taksonomiczne dla obiektu			Kiedy obiekt został włączony do kolekcji (może być sam rok, jeśli cała data nie jest znana)	Pochodzenie: można słownie wpisać kraj pochodzenia, jeśli nieznane, to pole zostaje puste	Wypełnić, jeśli obiekt został zebrany					Wypełnić, jeśli obiekt został wyhodowany (kod hodowcy wg FAO, jeśli nieznany, to opis w BREEDDESCR)	Jak poniżej:	Informacja o pochodzeniu lub jeśli obiekt został wyhodowany, dane o rodowodzie	Miejsce zbioru lub pozyskania wg kodów	Kod FAO donora, jeśli obiekt pozyskany z innej instytucji , jeśli nieznany, to opis w DONORDESCR	Numer akcesyjny obiektu przypisany przez donora, jeśli obiekt pozyskany z innej instytucji	Inne numery identyfikacyjne dla danego obi Jeśli obiekt został pozyskany z innego banku należy wpisać numer dla tego obiektu pocho tego banku genów poprzedzony kodem inst

Correctness

Most common mistakes:

- ORIGCTY: ~~ROM~~ → ROU, ~~LAT~~ → LVA
  - Data Validation tool (Excel)



- coordinates:
  - non-converted format put into wrong descriptors (LATITUDE:~~0216486E~~ → LATITUDE:0213855E)
  - lack of leading zeros (~~192554E~~ → 0192554E)
- „sea food” – coordinates at the sea area

Reliability

IHAR

Institute	POL003; Instytut Hodowli i Aklimatyzacji Roślin
Accession Number	508596
Collecting Number	
Collecting institute	
Genus	Avena
Species	brevis
Species Authority	Roth
Subtaxa	
Subtaxa Authority	
Common Crop Name	oat
Accession Name	Saia
Breeding institute	
Acquisition Date	
Country of origin	USA
Location of Collecting Site	
Latitude of Collecting Site	
Longitude of Collecting Site	
Elevation of Collecting Site	
Collecting Date of Sample	
Biological Status of Accession	Advanced/improved cultivar (odmiana hodowlana)
Ancestral Data	
Collecting Acquisition Source	Institute, Experimental station, Research organization, Genebank (instytut/stacja eksperymentalna/bank genów)
Donor institute details	
Breeding institute details	
Crop Groups	Spring cereals
Type of Germaplasm Storage	Medium term (przechowalnia średnioterminowa),Long term (przechowalnia długoterminowa)
Donor accession number	AVE 557
Donor institute	DEU146; Genebank, Institute of Plant Genetics and Crop Plant Research; Gatersleben
Other Identification Numbers	DEU146:AVE 557;DEU146:10.25642/IPK/GBIS/29433;DEU077:BBA 2636
Location of Safety Duplicates	
Remarks	spring type
Aegis Status	
Collecting institute details	
Decoded Safety Duplication Location	
Location of Safety Duplicates Code	
MLS Status	1-podlega

IPK

Genebank Information System of the IPK Gatersleben

You are not logged in >Your wish list is empty

Detailed information

Back to list

AVE 557

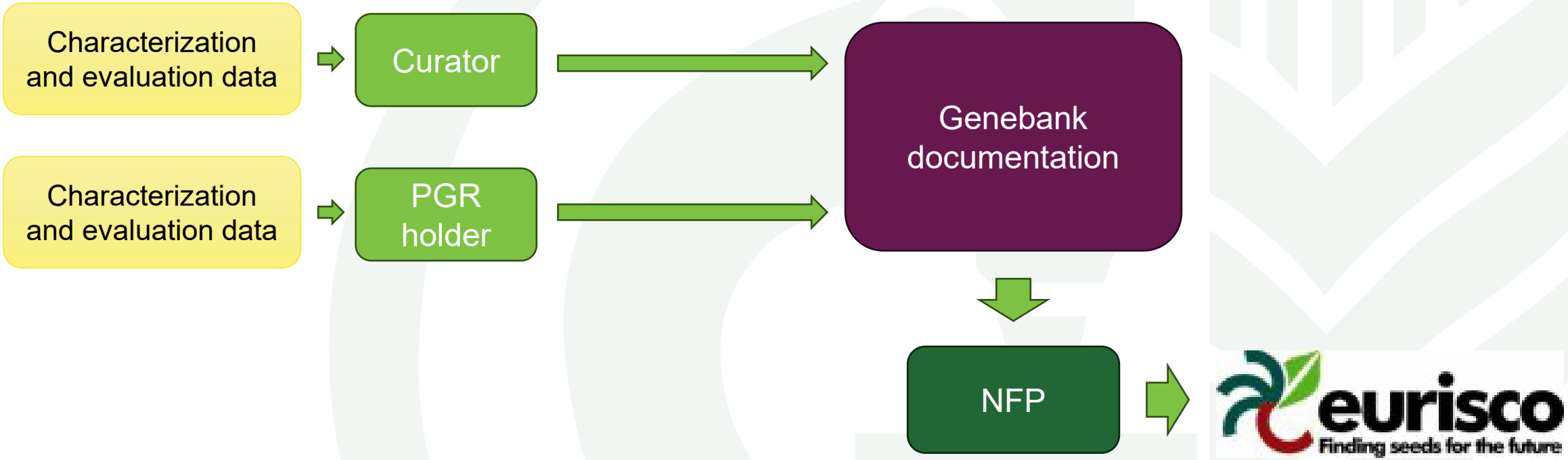
Passport data

DOI: *10.25642/IPK/GBIS/29433*
Accession numbers: AVE 557
Current scientific name: Avena brevis Roth
Biol. status: Advanced/improved cultivar
Life form: spring type
Availability: unrestricted available
ITPGRFA Annex1 crop: Oat
Acquisition form: unknown
Acquisition date: 1957
Genebank subcollection: GTR

Publications to scientific name
Accession names
Geographic data
Donor
Expedition
Characterisation and Evaluation Data

3. NFPs assure access to all publicly available quality phenotypic data related to the conserved PGR

Data flow

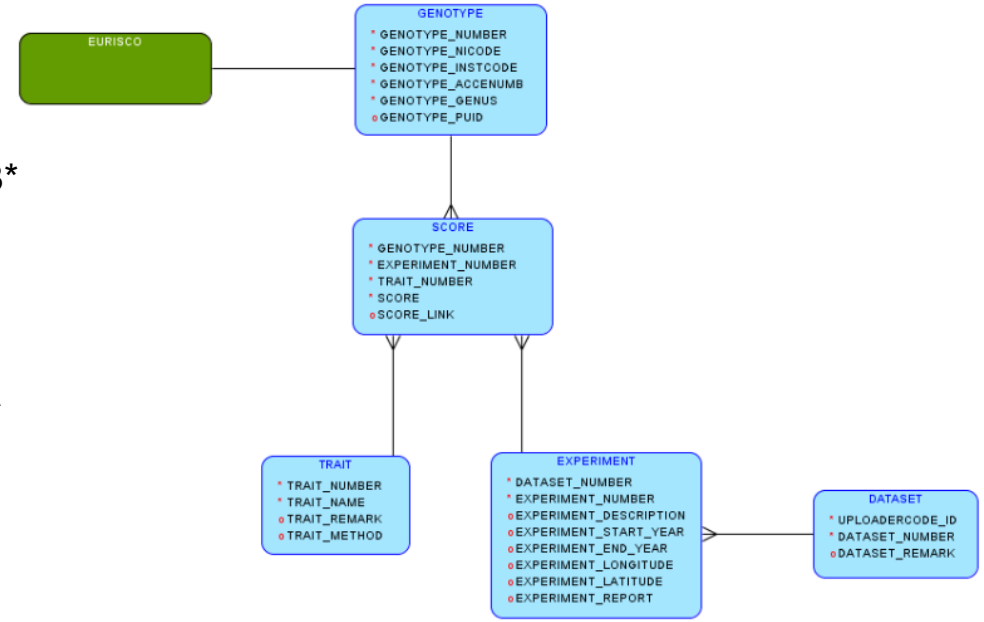


Data exchange standard for uploading C&E data to EURISCO

- 5 templates:
  - DATASET
    - UPLOADERCODE_ID*
    - DATASET_NUMBER*
    - DATASET_REMARK
  - EXPERIMENT
    - DATASET_NUMBER*
    - EXPERIMENT_NUMBER*
    - EXPERIMENT_DESCRIPTION
    - EXPERIMENT_START_YEAR
    - EXPERIMENT_END_YEAR
    - EXPERIMENT_LONGITUDE
    - EXPERIMENT_LATITUDE
    - EXPERIMENT_REPORT
  - TRAIT
    - TRAIT_NUMBER*
    - TRAIT_NAME*
    - TRAIT_REMARK
    - TRAIT_METHOD
  - GENOTYPE
    - GENOTYPE_NUMBER*
    - GENOTYPE_NICODE*
    - GENOTYPE_INSTCODE*
    - GENOTYPE_ACCENUMB*
    - GENOTYPE_GENUS*
    - GENOTYPE_PUID
  - SCORE
    - GENOTYPE_NUMBER*
    - EXPERIMENT_NUMBER*
    - TRAIT_NUMBER*
    - SCORE*
    - SCORE_LINK

eurisco
Finding seeds for the future

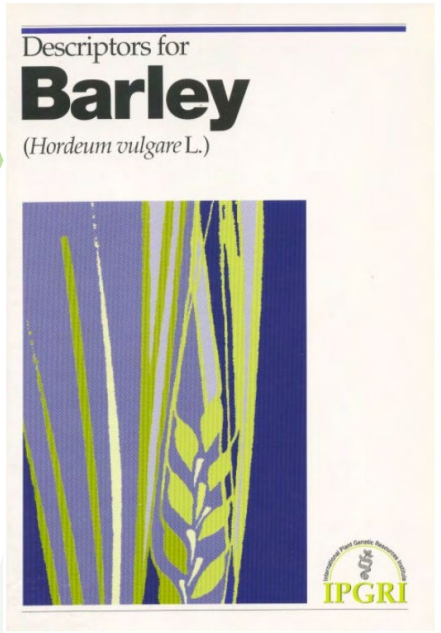
Data exchange standard for uploading characterisation and evaluation data from National Inventories to EURISCO

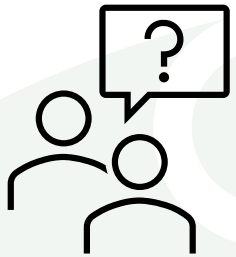


ECP/GR

Quality of phenotypic data - what descriptors should be used for characterization and evaluation?

1994
Perhaps update needed?





E

UPOV

TG/3/12 Rev.
ORIGINAL: English
DATE: 2017-04-05
+ 2022-10-25

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
Geneva

WHEAT
UPOV Code(s): TRITI_AES
Triticum aestivum L. emend. Fiori et Paol.

GUIDELINES
FOR THE CONDUCT OF TESTS
FOR DISTINCTNESS, UNIFORMITY AND STABILITY

Alternative names:*

Botanical name	*English*
Triticum aestivum L. emend. Fiori et Paol.	Wheat

ECP/GR

ECPGR Characterization and Evaluation
Descriptors for Apple Genetic Resources

Apple (*Malus x domestica*)

M. Lateur, E. Dapena, D. Szalatnay, M. E. Gantar, A. Guyader, I. Hjalmarsson, M. Höfer, L. Ikase, M. Kellerhals, G. Lacis, M. Militaru, C. Miranda Jiménez, G. Osterc, J-B. Rey, A. Rondia, K. Volens, M.K. Zeljković, M. Ordidge.

Crop Ontology
for agricultural data

Phenotype and Trait Ontology

Ontology name	Ontology description	Type
Apple Created on Tuesday 23rd of August, 2022. 13:13:10 53 variables	Apple (Malus domestica) defines crop traits and variables to support the standardisation of apple breeding databases providing description of agronomic, morphological, physiological, quality traits, its methods and scales.	Trait
Bambara groundnut Created on Friday 20th of December, 2019. 08:42:44 163 variables	version Dec 2019	Trait
Banana Created on Thursday 18th of April, 2019. 10:45:39 390 variables	Banana Trait Dictionary in template 5 - Bioversity & IITA - April 2019	Trait
Barley Created on Tuesday 9th of June, 2020. 07:45:15 148 variables	ICARDA - TDv5 - Sept 2018	Trait

E

UPOV

TG/14/10
ORIGINAL: English
DATE: 2023-10-24

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
Geneva

APPLE
(Fruit Varieties)
UPOV Code(s): MALUS_DOM
Malus domestica (Suckow) Borkh.

GUIDELINES
FOR THE CONDUCT OF TESTS
FOR DISTINCTNESS, UNIFORMITY AND STABILITY

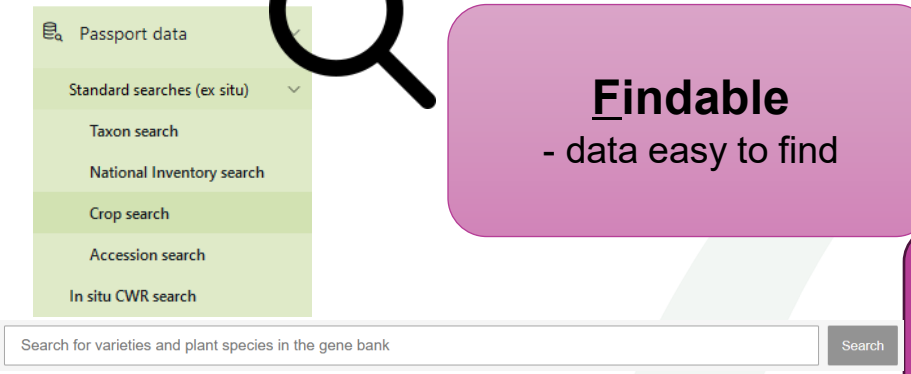
ECP/GR

4. Genebanks and other PGR holders can improve their data management practices

- Training workshops for genebank information system officers
  - general training (online or in-person) or national level?
- Genebank level: training workshops for curators and other data providers, translating passport data templates,

5. Both data in EURISCO and the associated IT infrastructure are compliant with the FAIR principles

https://eurisco.ipk-gatersleben.de/
https://wyszukiwarka.ihar.edu.pl/en



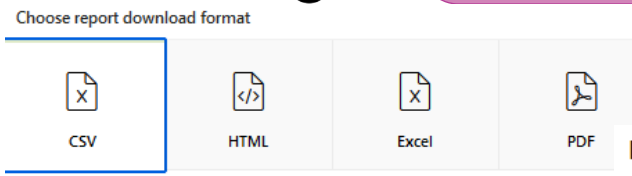
Findable
- data easy to find

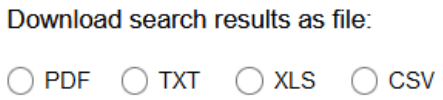
Accessible
- how to get data?

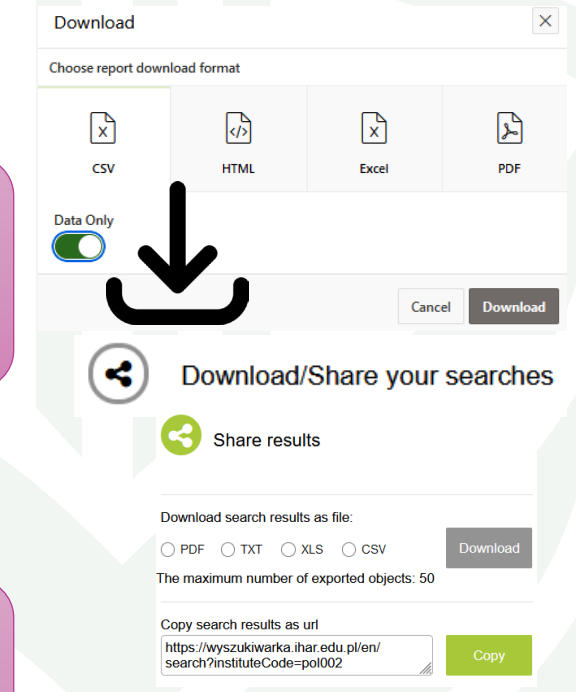
FAIR principles

Interoperable
- data format needs to interoperate with applications for analysis

Reusable
well described data →can be replicated/combined in different settings







Thank you for your attention!

Mgr inż. Renata Kowalik
Contact information:
phone number: +48 22 733 46 40
e-mail: r.kowalik@ihar.edu.pl