UPOV
International Union for the Protection of New Varieties of Plants

E

Technical Working Party on Testing Methods and Techniques **TWM/3/23**

Third Session
Beijing, China, April 28 to May 1, 2025

Original: English
Date: April 17, 2025

SHARED MOLECULAR DATABASE - WHAT DUS MOLECULAR DATABASE FOR THE FUTURE? - GEVES OPINION

Document prepared by an expert from France

Disclaimer: this document does not represent UPOV policies or guidance

The annex to this document contains a copy of a presentation “Shared molecular database - What DUS molecular database for the future? - GEVES opinion”, to be made by an expert from France, at the third session of the TWM.

[Annex follows]

Shared molecular database

What DUS molecular database for the future? GEVES opinion

GEVES
Expertise & Performance

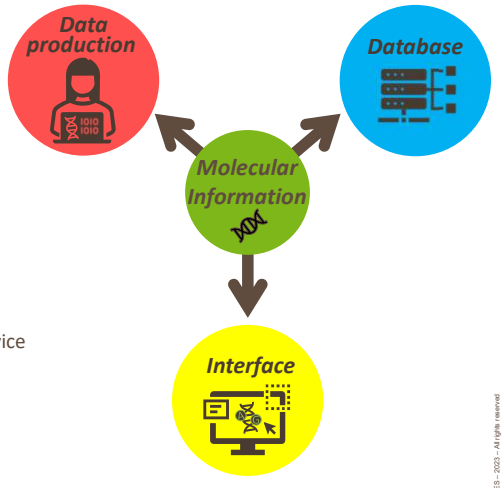
1

1

First, BETTER DEFINE

To keep in mind for the discussion

- There is a need to distinguish:
  - The **database** itself
  - The **tool and interface** to connect/interrogate the database
  - The **data production** **by the lab :** service provider or Examination Office (EO)

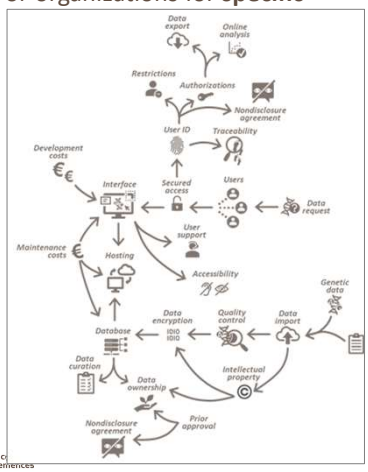


2

2

Definition of a shared molecular database?

- A **shared molecular database** is a centralized repository of **molecular information** that is accessible to **multiple users** or organizations for **specific goal(s)**.
- What do we mean by **molecular information**?
- **Multiple users**? => Several issues
- What do we want to achieve by **sharing data**?



GEVES | Groupe d'Étude et de c… des Variétés Et des Semences

3

Molecular information:
What do we mean by molecular information?

– Genetic distances only?

– Genetic profiles?

	Locus 1	Locus 2	Locus 3
Variety 1	A/A	B/B	A/B
Variety 2	A/B	B/B	A/B
Variety 3	A/A	?/?	B/B
…			

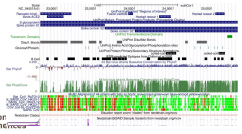
Encoded

	GWM345	Fgh34	Empas354
Variety 1	A/A	B/B	A/B
Variety 2	A/B	B/B	A/B
Variety 3	A/A	?/?	B/B
…			

Not encoded

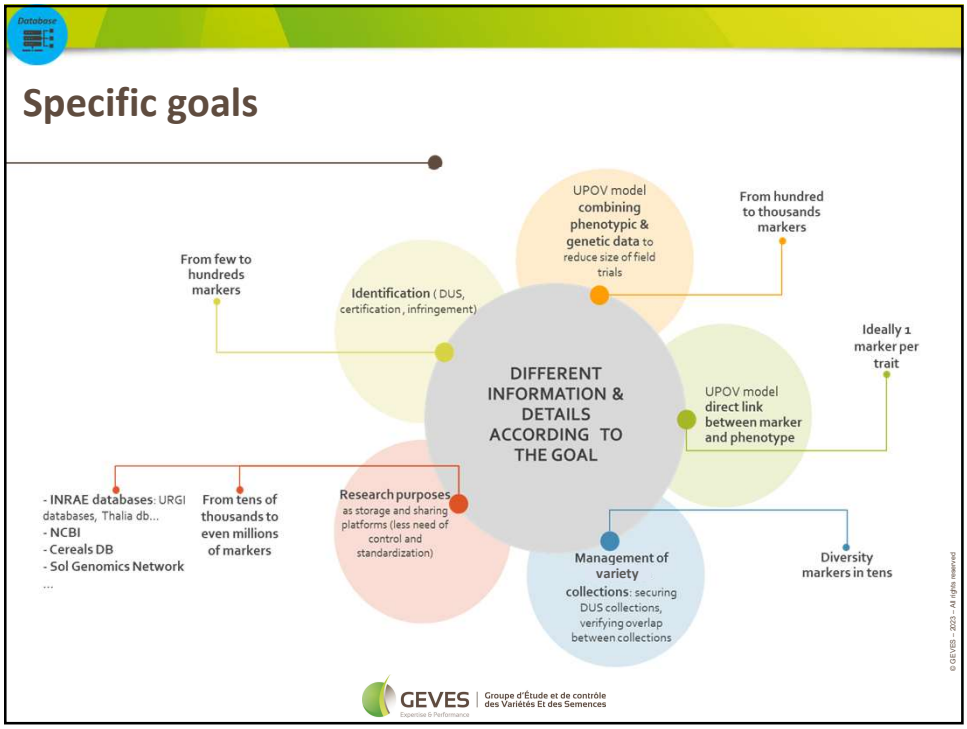
– Whole genome sequences with annotation?

– Metadata : **genetic markers used**, including **sequences**, genomic? location?

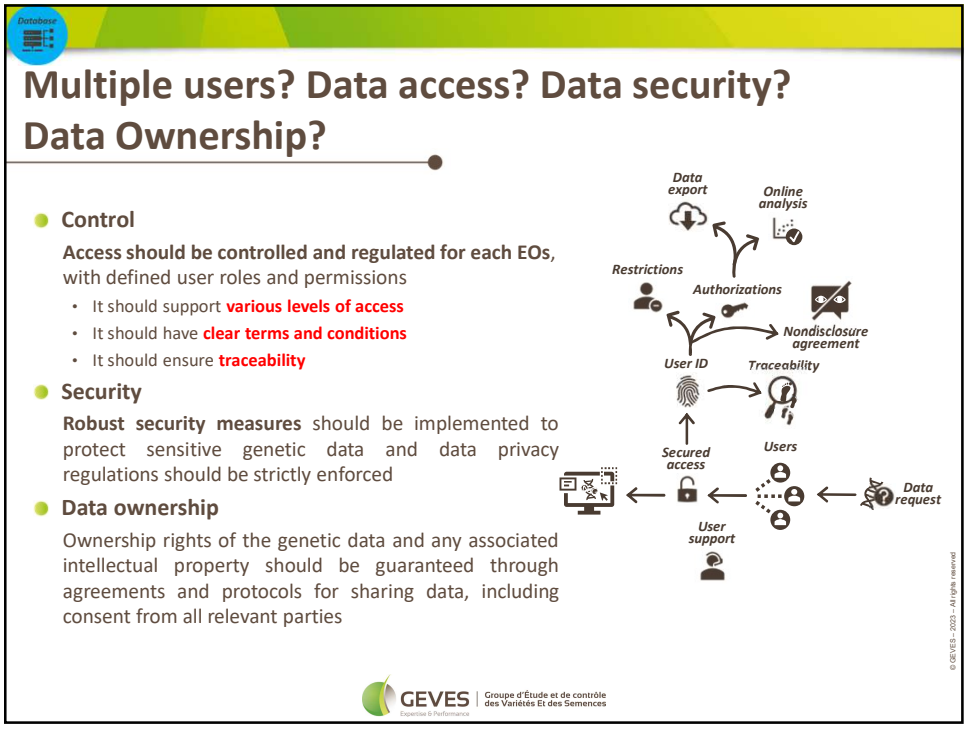


GEVES | Groupe d'Étude et de con… des Variétés Et des Semences

4


Database
Specific goals
DIFFERENT INFORMATION & DETAILS ACCORDING TO THE GOAL
Identification (DUS, certification , infringement)
From few to hundreds markers
UPOV model combining phenotypic & genetic data to reduce size of field trials
From hundred to thousands markers
UPOV model direct link between marker and phenotype
Ideally 1 marker per trait
Research purposes as storage and sharing platforms (less need of control and standardization)
- INRAE databases: URGI databases, Thalia db...
- NCBI
- Cereals DB
- Sol Genomics Network
...
From tens of thousands to even millions of markers
Management of variety collections: securing DUS collections, verifying overlap between collections
Diversity markers in tens
GEVES | Groupe d'Étude et de contrôle des Variétés Et des Semences

5


Database
Multiple users? Data access? Data security? Data Ownership?
Control
Access should be controlled and regulated for each EOs, with defined user roles and permissions
It should support various levels of access
It should have clear terms and conditions
It should ensure traceability
Security
Robust security measures should be implemented to protect sensitive genetic data and data privacy regulations should be strictly enforced
Data ownership
Ownership rights of the genetic data and any associated intellectual property should be guaranteed through agreements and protocols for sharing data, including consent from all relevant parties
Data export
Online analysis
Restrictions
Authorizations
Nondisclosure agreement
User ID
Traceability
Secured access
Users
Data request
User support
GEVES | Groupe d'Étude et de contrôle des Variétés Et des Semences

6

Several considerations depending on the use of the molecular DB

- **Harmonization**

  In case several offices (or service providers) produce molecular data, there should be a set of **control mechanisms** to ensure data quality and comparability:
  - **Clear technical specifications** for producing molecular profiles across different platforms or using different technologies
  - Agreed-upon **standards and formats** for coding and sharing DNA profiles
  - **Consensual set of reference varieties** (biological and technical replicates) for well-characterized or standard plant varieties

QUESTIONS TO AGREE IN ADVANCE/ PREREQUISITE

- **Transparency, quality system & traceability**

  Information on the methodology used for DNA profiling, quality control measures and data validation procedures should be made readily available

- **Consensus**

  Where multiple contributors have produced DNA profiles for the same plant variety, should each office upload its reference profile, or should there be mechanisms to combine data in order to establish a **single, agreed-upon consensus profile** for each variety?

GEVES | Groupe d'Étude et de contrôle des Variétés Et des Semences

7

7

Curating & maintaining the database

- **Who is in charge**?
  - Should a single (host) office or all offices (users) be in charge of the contents of the database ?

- Sharing the costs ? Towards a **fair economic model**
  - Cost for the database development
  - Access to the database subject to a fee to cover all the costs related to:
    - Hosting the database
    - Maintaining the database
    - Improving the database

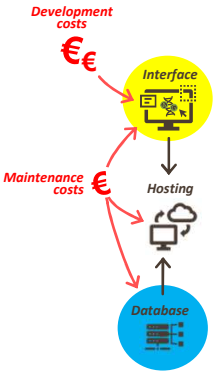
- **Curating**
  - Procedures are required to ensure that the database is up to date

8

8

User-friendly interface

- Developing the interface
  - The database should have an **intuitive** and **user-friendly** interface that allows :
    - Contributors: **easily** and **reliably** input data,
    - Users: easy to **search, access, analyse** and **export** molecular information
- Cost : **Fair economic model**
  - Develoment
  - Hosting the online tool/interface
  - Maintaining and improving tool/interface



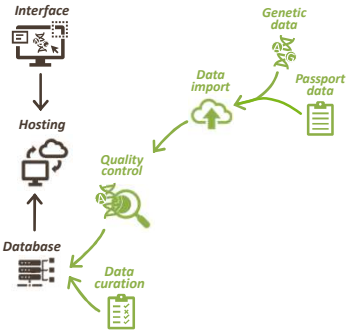
GEVES | Groupe d'Étude et de contrôle des Variétés Et des Semences

9

9

Data production

- One or several labs?
- In house or out-sourced ?
- Which genotyping technology ?



10

10

In house or outsourced?

- Sometimes it might be necessary to **outsource the molecular work to a third-party**
  - ✗ *because there is **no in-house facilities***
  - ✗ *because the **technology is not available***
  - ✗ *because the **workload** is too high*
  - ✗ *because **delays** cannot be met*

- If outsourced :
  - Prerequisites and criteria for the choice of the service provider must be defined and agreed
  - Legal aspects need also to be considered ahead (contract template to be made) including transparency of methods and information , etc

11

11

UPOV_INF_17 GUIDELINES FOR DNA-PROFILING: MOLECULAR MARKER SELECTION AND DATABASE CONSTRUCTION ("BMT GUIDELINES")

- consider the approach on a crop-by-crop basis;
- agree on an acceptable marker type and source;
- agree on acceptable detection platforms/equipment;
- agree on laboratories to be included in the test;
- agree on quality issues;
- verify the source of the plant material used;
- agree which markers are to be used in a preliminary collaborative evaluation phase, involving more than one laboratory and different detection equipment;
- conduct an evaluation;
- develop and agree a protocol for scoring the molecular data;
- agree on the plant material/reference set to be analyzed, and the source(s);
- analyze the agreed variety collection, in different laboratories/different detection equipment, using duplicate samples, and exchanging samples/DNA extracts if problems occur;
- use references (varieties, DNA samples and alleles, as appropriate) in all analyses;
- .....

12

12

UPOV_INF_17 GUIDELINES FOR DNA-PROFILING: MOLECULAR MARKER SELECTION AND DATABASE CONSTRUCTION ("BMT GUIDELINES")

- consider the approach on a crop-by-crop basis;
- agree on an acceptable marker type and source;
- agree on acceptable detection platforms/equipment;
- agree on laboratories to be included in the test;
- agree on quality issues;
- verify the source of the plant material used;
- agree which markers are to be used in a preliminary collaborative evaluation phase, involving more [illegible] one laboratory and different detection equipment;
- conduct an evaluation;
- develop and agree a protocol for scoring the molecular data;
- agree on the plant material/reference set to be analyzed, and the source(s);
- analyze the agreed variety collection, in different laboratories/different detection equipment, using duplicate samples, and exchanging samples/DNA extracts if problems occur;
- use references (varieties, DNA samples and alleles, as appropriate) in all analyses;
- .....

13

13

As for any shared DB…

- Main issues to take into account **before** creating a shared Database
  - Purposes : impact on structure , choice of genotyping strategy
  - Usefullness : **avoid redondancy** with other DB, seek completeness, regular updates
  - Users : impact on access rules, need for **consent** from variety owner?
  - Contributors : harmonization needed if several contributors (see UPOV/INF17), rules for updating, **rules for the choice if outsourced**
  - Ownership of data
  - Ownership of tool/interface
  - Coordination
  - Maintenance, curation
  - Economic model : data production, access fee, maintainance, evolution

14

14

In particular for DUS molecular Database…

- Network of Offices
  - to work together,
  - to encourage cooperation
  - to avoid **redondancy** and seek efficiency
- To be cost efficient, genotyping a complete DUS variety collection should be **useful for other purposes too**, such as identity for certification, identity for infringement
- Useful for reliable DUS procedures including molecular information respecting UPOV requirements and models : **transparency** on how molecular information is used, harmonized between EOs to prevent different decisions on basis of same data

15

15

CONCLUSION

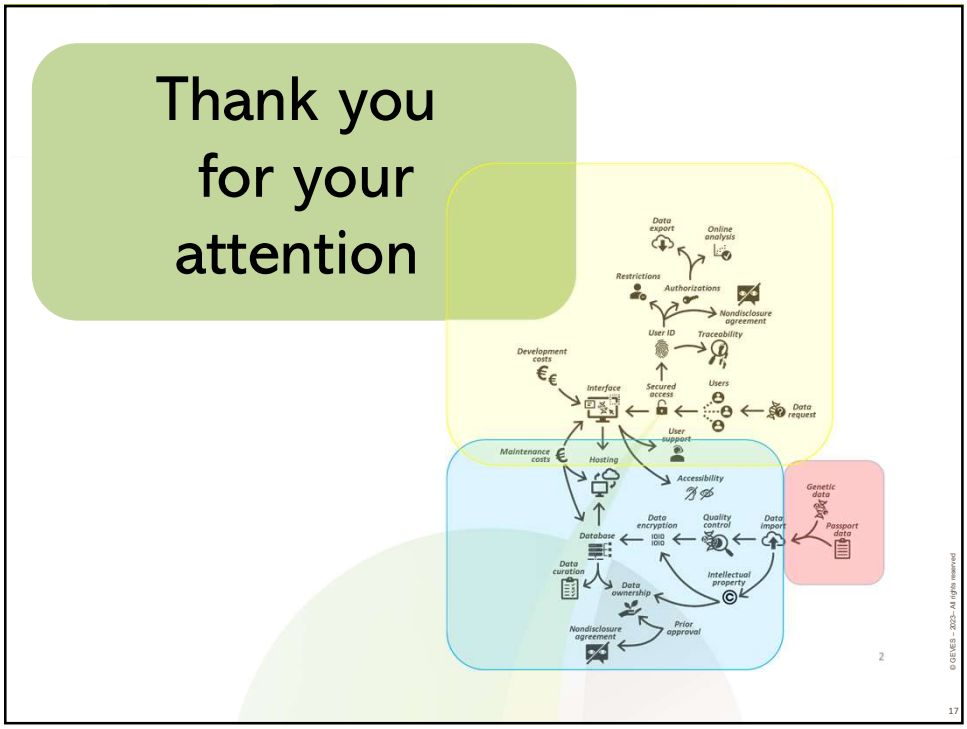
In particular for DUS molecular DB

- ***Take home message for an ideal shared molecular database:***
- **Transparency of methods** for EOs and breeders/applicants
- **Access to molecular information for EOs**
  - Full DNA profiles
  - Full information on the genetic markers used
  - Metadata associated with each DNA profile
  - Information needed about the plant varieties

- **Fair economic model**

16

16



17

[End of Annex and of document]