

# Permanent Unique Identifiers for germplasm

Susan McCouch  
and  
Ruairaidh Sackville Hamilton



# The need for PUIs

- Genebank managers need to know
  - What has been done with their accessions
  - What duplication there is among collections
  - > **20 years of unsuccessful attempts**
- Collaborators in DivSeek need assurance
  - That they are truly working on the same genetic material
  - Learning by bitter experience**
- The Treaty`s GLIS needs to
  - Document holdings and transfers of all types of PGRFA
  - Principles embedded in the Treaty & SMTA**



# What is a Permanent Unique Identifier?

## Minimum definition

a text string that **unambiguously** and **permanently** identifies a single object of interest

*Marco Marsella*

- Purpose of identifier?
- What is the object to be identified?
- Text string format?
  - Identifier, name or description?
- Scope?
  - Unambiguous among what set of objects?



# Purpose of identifier

## Record identifier in database

- Primary key
- Unique within the database
- Internal, not for publication or human use

## Identifier to label packet

- Chosen by curator
- Public
- Unique within curator`s system
- May be a code or a name, descriptive or not

## Identifier for global online access

- Globally unique
- In one of the standard formats for www access
- Labelling seed packets is not primary purpose





# What is a Permanent Unique Identifier?

## Minimum definition

a text string that **unambiguously** and **permanently** identifies a single object of interest

## Key characteristics of a good PUI

- uniqueness
- permanence
- opaqueness / anonymity
- actionability / resolvability
- discoverability



# What do we need to identify?

*It depends on context*

- Crop: **rice**
- Traditional variety (no formal control of identity): **Malagkit**
- Modern variety (controlled identity): **Swarna**
- Accession: **IRGC 326, TOG 123**
- Seed lot of an accession: **IRGC 326:2012DS**
- Harvest from a single seed: **IR 1330-5**
- DNA extracted from a tissue sample: **4987289**
- Fixed line from a single seed: **IR 1330-5-3-3**
- Mixed: **IR 1330-5-3-3//IR 24\*4/O. nivara**

*Unambiguous in local context:  
often not outside local context*

**IRRI**

INTERNATIONAL RICE RESEARCH INSTITUTE



# What do we need to identify?

Suppose seed sample B is created from A



In which of these cases does B need a different identifier?

- B is a subsample of A
  - Taken for storage in a different place
  - Taken for a viability test
  - Given to a different organization for outsourced data collection
  - Given to a different organization for their own maintenance / research
- B is a new generation of seed
  - Created by seed multiplication to keep the same genetic composition
  - Created by growing a single random seed of A
  - Created by selecting a specific variant found in A



# Methods of creating progeny

## Many methods: three classes

- “**Generative**” methods generate new diversity
  - Crossing / hybridization
  - Induced mutation
  - GM methods
- “**Derivative**” methods derive progeny that are subsets of diversity in their parents
  - Selections from segregating populations
  - Separating components of a mixture
- “**Maintenance**” methods create progeny intended to be the same as their parents
  - Seed multiplication
  - Sub-sampling, e.g. For material transfers



# What do we need to identify?

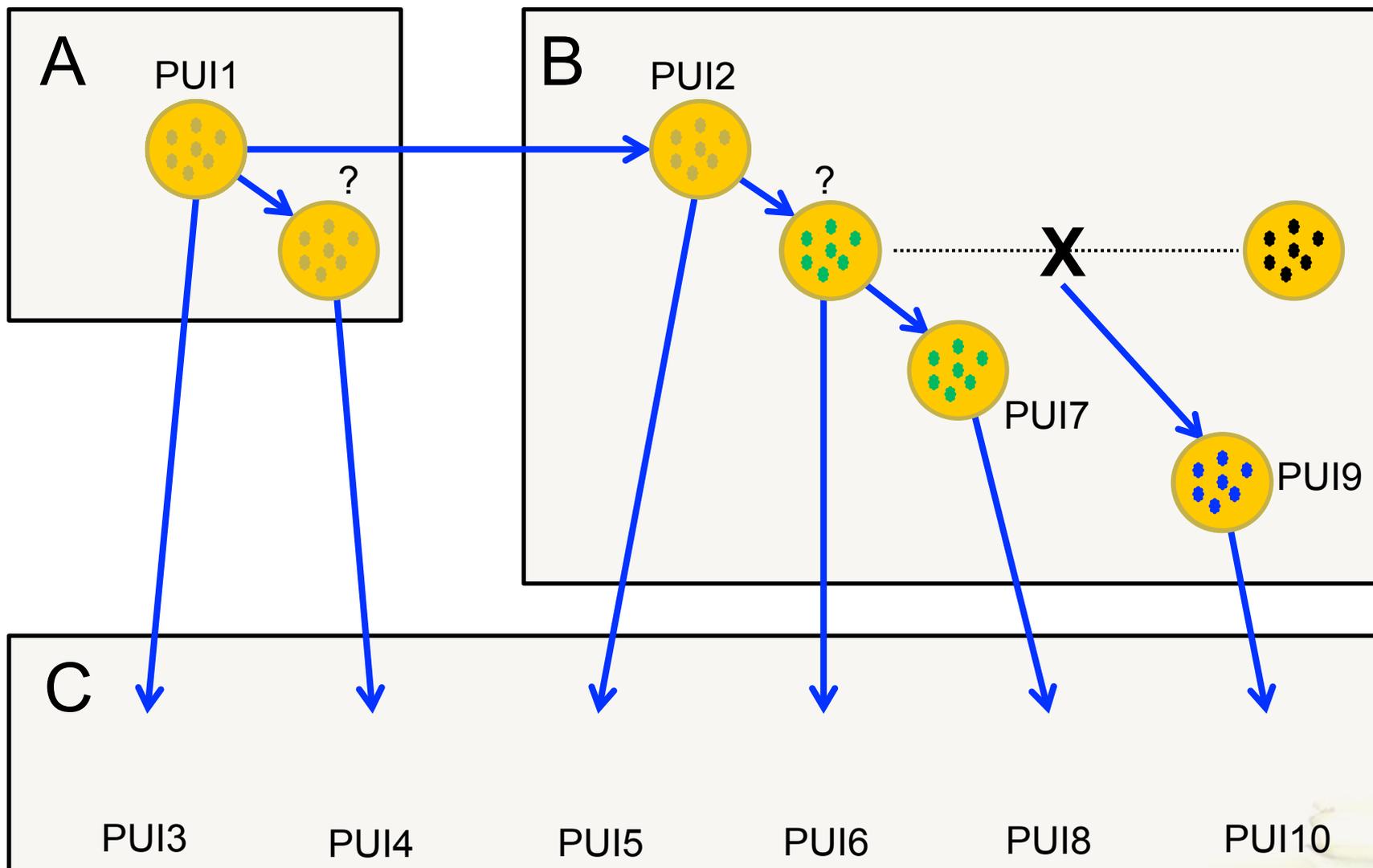
Suppose B is a subsample of A given to a different organization for its own research



- Genebanks:
  - Want reliable accountability & attribution
  - B might be or become different,
    - especially if B is not managed using genebank standards
- DivSeek:
  - Need reliable accountability & attribution
  - Need traceability in case something goes wrong
- GLIS:
  - Need reliable accountability and attribution
  - B is legally a different entity
  - Treaty is sample-based, not genotype-based



# Treaty vs DivSeek perspectives



# ICIS germplasm table:

handling parent-offspring relationships  
with  $\geq 1$  records for each genetic entity

Global germplasm identifier (GID) of sample

Number of immediate parents

## **GID of immediate parental sample**

Method of derivation from parent

Date of derivation from parent

Place of derivation from parent

## **GID of original sample**

ID of data contributor's database

Data contributor's local germplasm ID

Reference to data source

**IRRI**

INTERNATIONAL RICE RESEARCH INSTITUTE



# Scope

- Genebanks
  - Only PGRFA that are accessions
- DivSeek
  - All types of PGRFA held *ex situ*
    - Genebank accessions, purified stocks, mapping and other specialised research populations, elite and other prebreeding lines, released cultivars ...
  - Subset = PGRFA useful for genetic diversity analysis
- Treaty
  - All PGRFA (*ex situ* and *in situ*)
- Treaty`s Multilateral System
  - All types of PGRFA
  - Subset = PGRFA available for sharing under MLS





## **Digital Object Identifiers: the PUIs for GLIS**

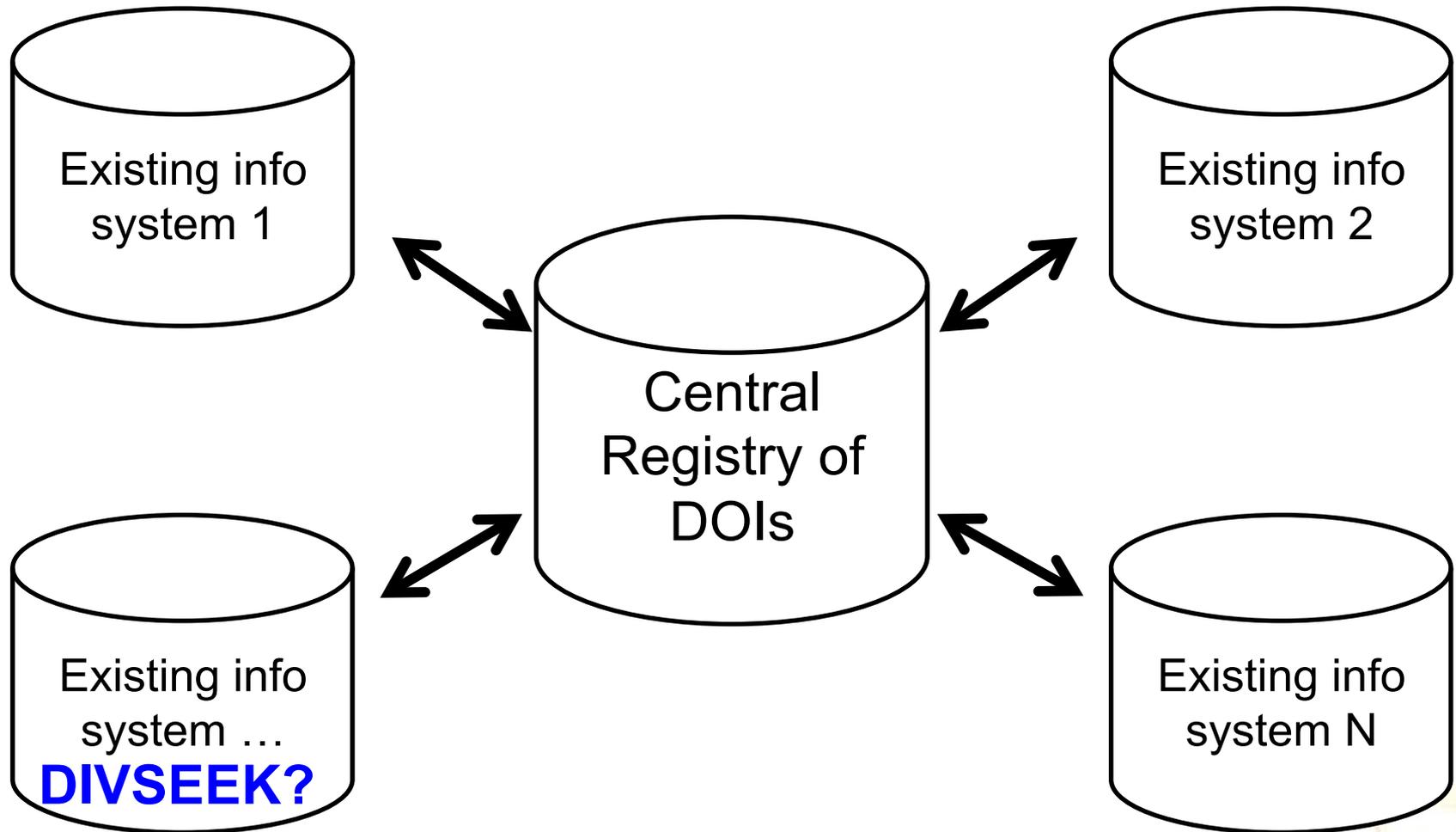
Digital Object Identifiers (DOIs) have been selected as the PUI type for GLIS because:

- they are a ISO standard (ISO 26324)
- they are managed by a central authority (International DOI Foundation)
- they are widely used in the scientific community
- by design, they accommodate existing identifiers
- they have a flexible and extensible metadata structure
- they support advanced features such as Content Negotiation and Multiple Resolution

*Source: Marco Marsella*

# GLIS concept

minimal data centralised: link to existing systems



# Data associated with a DOI

- **Essential** (copied to central registry)
  - Who holds the material
  - How the holder labels the material
  - Minimal description of the type of material
    - Crop or genus
- **Highly recommended** (centralised or links?)
  - Provenance of the material
    - Its origin, how it was created or obtained
  - Further description of the type of material
    - Species, type of PGRFA ...
- **Desirable** (through links to existing systems)
  - Any additional available passport data (e.g. crop-specific ecological data), genotypic data, phenotypic data



# First steps: Indonesian BSF project

## Two use cases

1. Collection holder declares a PGRFA sample available under the MLS
  - Create a DOI for the sample
  - Associate DOI with other available data
    - In central registry or in system used by holder
2. Provider transfers a sample to a Recipient
  - Create DOI for provider`s sample if it doesn`t already exist
  - Create DOI for recipient`s sample
  - Create associated passport data for recipient`s sample
    - Including pointer to provider`s sample as source



THANK YOU!

**IRRI**

INTERNATIONAL RICE RESEARCH INSTITUTE

