

Tuesday 16 May a.m.

- 08:00** **Aligning Genebank concept with ICIS Objects** Ruaraidh Sackville Hamilton
- 08:40** **IRRI Genetic Resources Information Management System** Thomas Metz
- 09:20** **Coffee Break**
- 09:40** **IRRI GRIMS GUI** Thomas Metz
- 10:20** **CIMMYT crop-generic Genebank, phase 1** Jesper Norgaard Welen
- 11:00** **Discussion / Q&A / Wrap-up**

***Aligning Genebank concepts
with ICIS Objects***

or:

***Aligning ICIS Objects
with Genebank concepts?***

Ruaraidh Sackville Hamilton



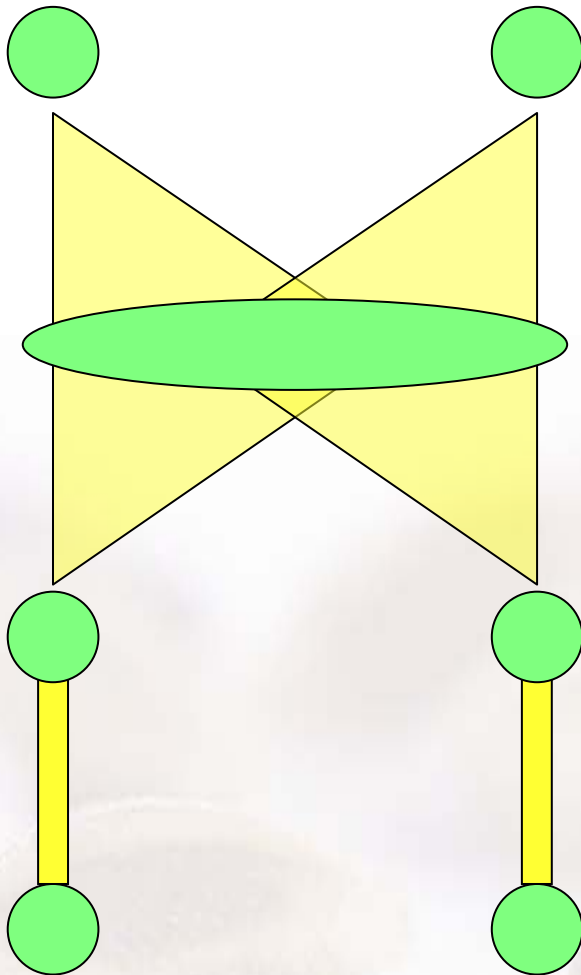
Background

- **Last year of 3 year project to migrate IRGCIS to IRIS**
- **IRGCIS**
 - Functional genebank management system
 - Based on 40 years of internationally agreed genebank standards
 - Guiding genebank staff through daily operations
 - Needing further development for new needs
 - Needing to integrate with other IRRI germplasm data
- **Migration requires ICIS development**

Outline

- **Germplasm creation methods**
 - ICIS vs genebanks
- **Germplasm management & identification**
 - What are the entities?
 - How do they map to GIDs and LotIDs?
- **Passport data in ICIS**
 - Mapping Multi-Crop Passport Descriptors onto ICIS
- **DMS**
 - characterization data
 - germination data
- **Definitions and Procedures**

GMS Germplasm creation methods



Generative creation method:

Increasing diversity
by combining germplasm

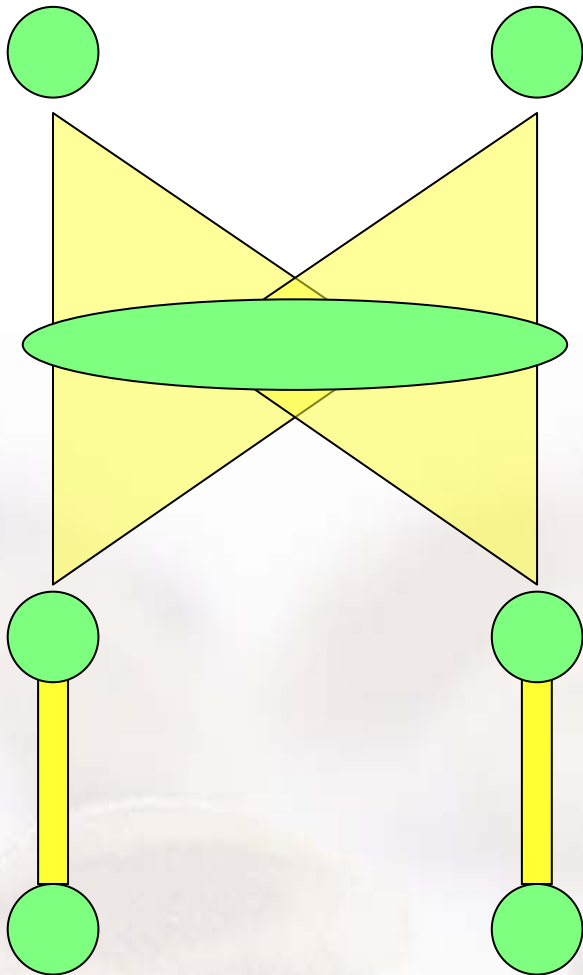
Derivative creation method:

Reducing diversity
by selecting variants from a population

Maintenance creation method:

Maintaining genetic composition unchanged
(within limits: drift and uncontrolled selection)

Germplasm management in genebanks



Generative creation method:

Combining duplicate accessions
(should not increase diversity)

Derivative creation method:

Splitting a variable accession into several
(one option for managing variability)

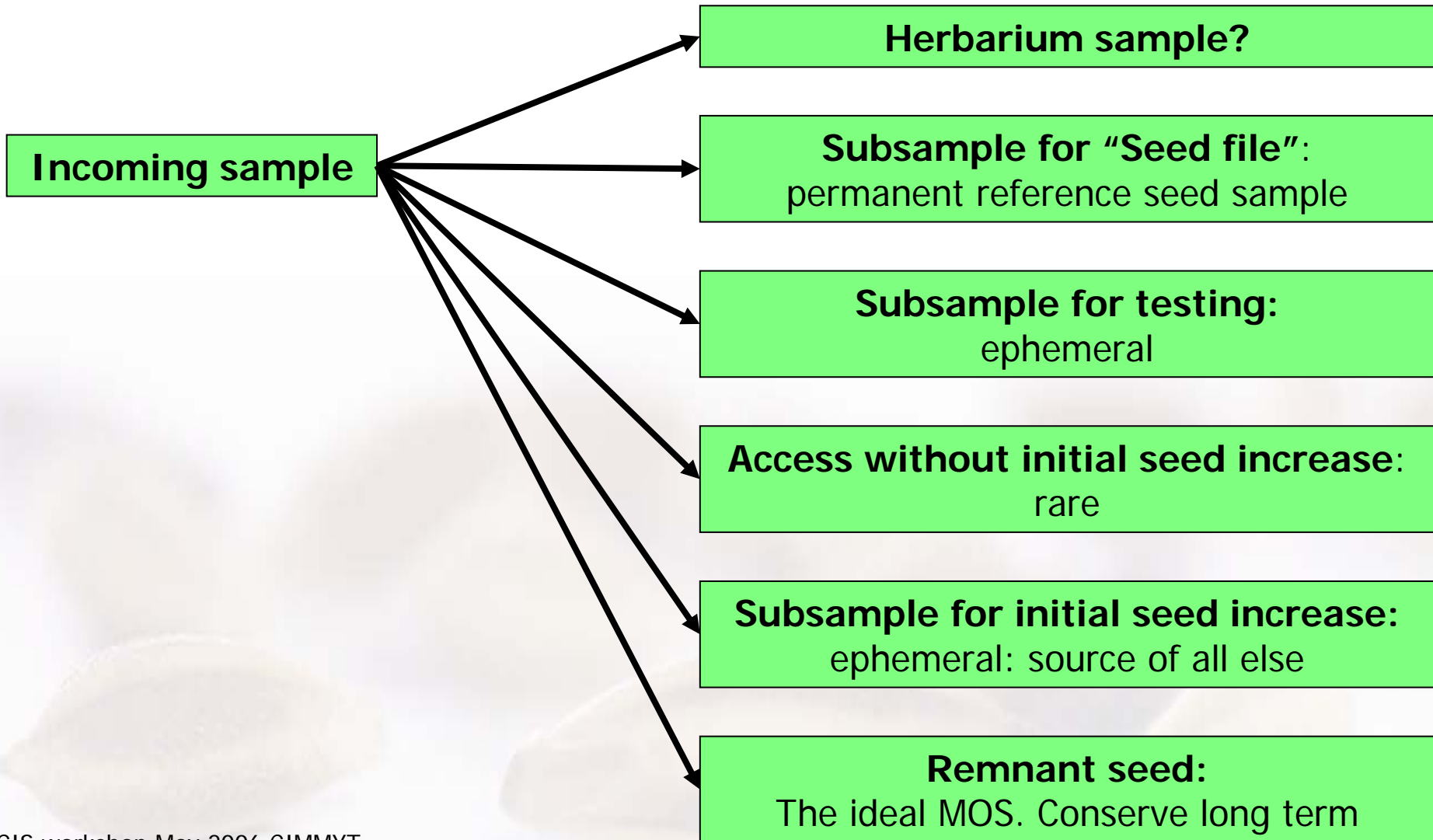
Maintenance creation method:

Maintaining constant composition
(most seed multiplication / rejuvenation)

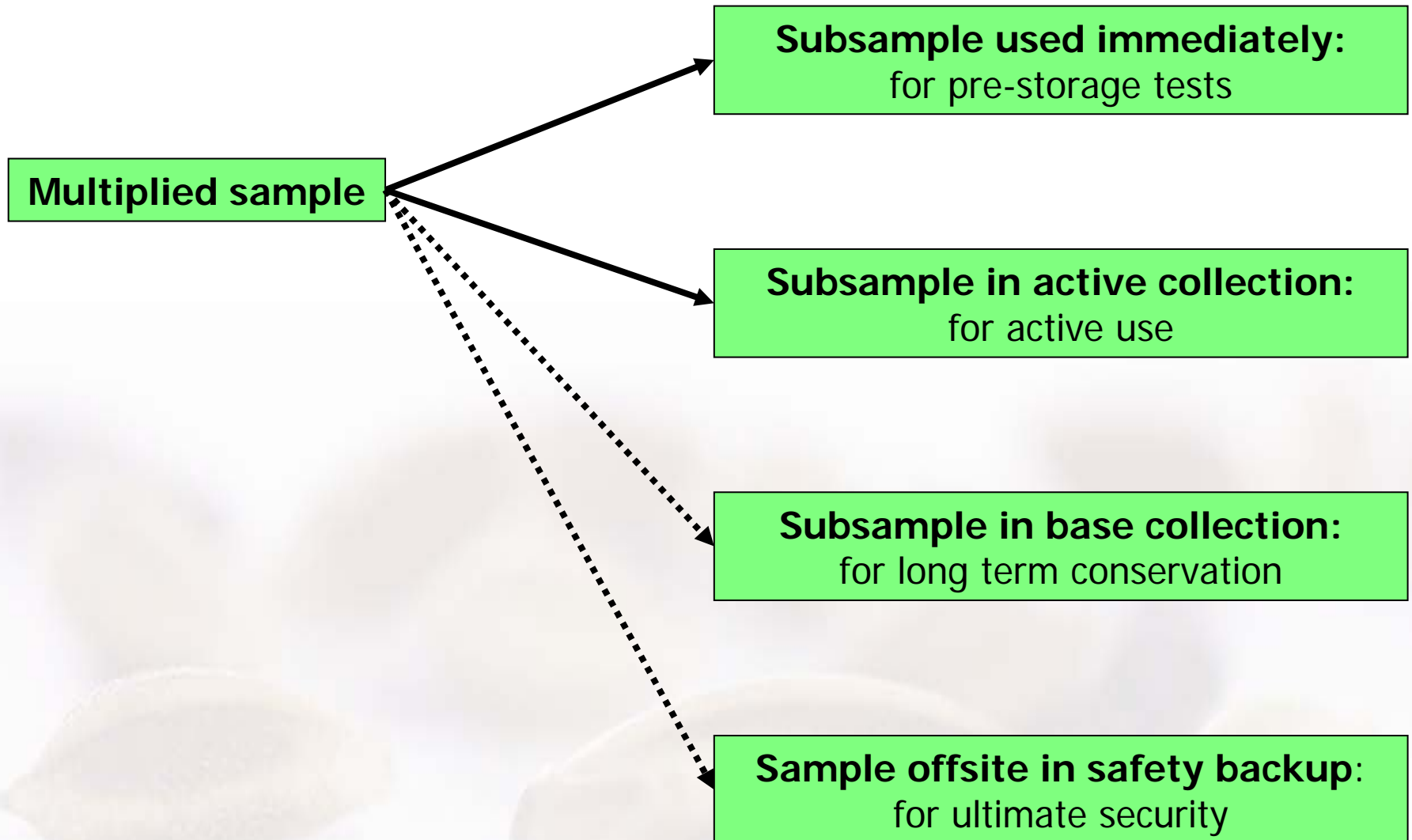
Genebank Accession

- **A sample of germplasm**
 - accessed into a germplasm collection
 - conserved, characterized and distributed as an entity, a family of samples
 - maintained with genetic composition as close as possible to the original sample, despite repeated generations of seed multiplication
- **MOS (Most Original Sample):**
 - The extant sample that is the closest to the original sample

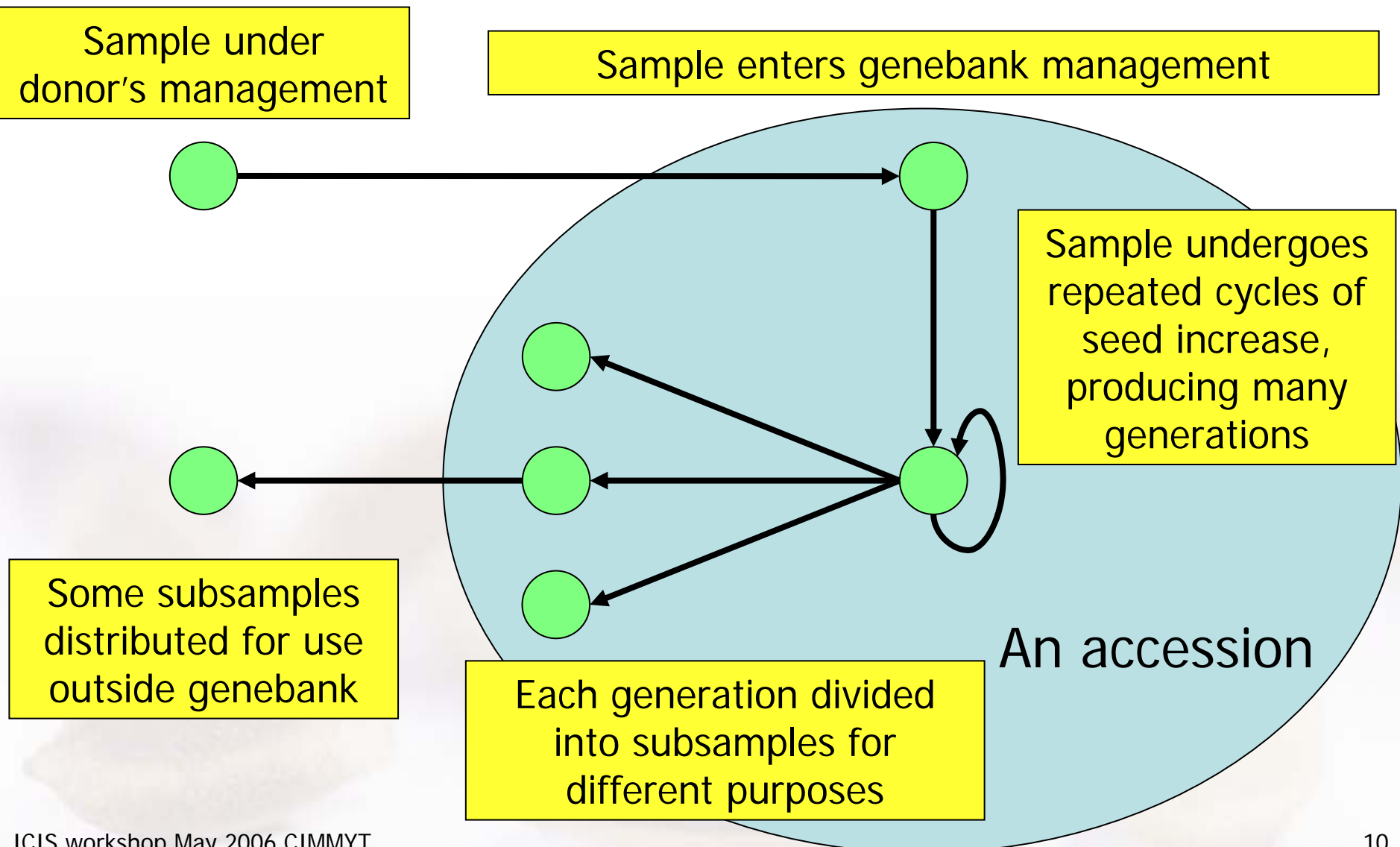
Accessing an incoming sample into the collection



Maintaining seed stocks in genebanks



Managing accessions



3-level hierarchy of entities

- **Accession**
- **Generation of accession**
- **Sample of generation**
- **Conventional genebank databases have 3 corresponding sets of tables**

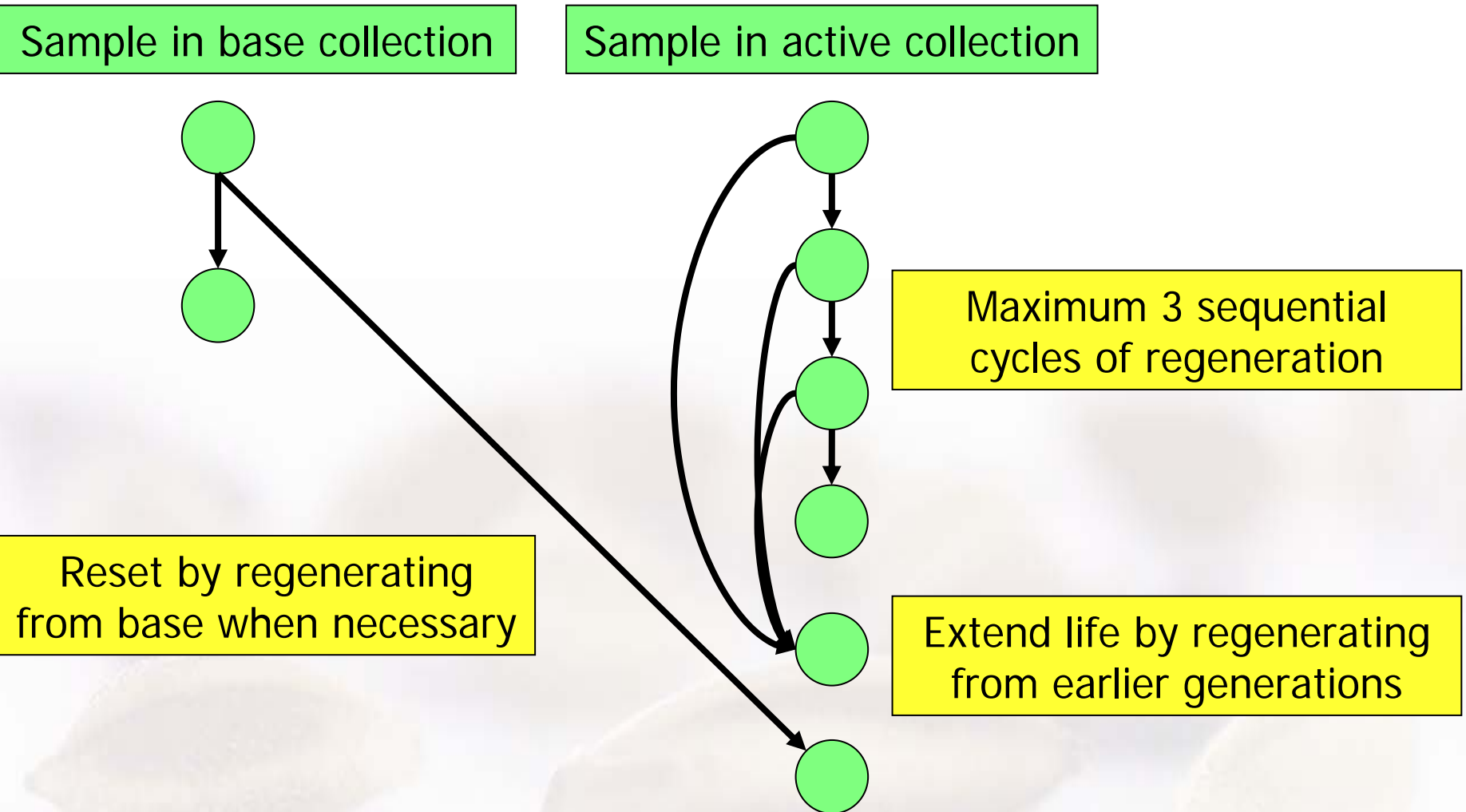
Accession entity

- **Unique ID within genebank**
 - Genebank ID prefix → globally unique
- **Carries passport data / IP data**
 - Retained over generations
- **Preferred ID for publication & distribution**
 - Accession-specific rules determine how we conserve & distribute
 - Accessions of same variety name differ genetically
- **Queries and requests for germplasm usually based on accession**

Generation entity

- **Must not mix**
 - Might differ genetically
 - because of drift, shift & gene flow
- ➔ **1 generation of an accession = 1 GID**
- **Characterisation & distribution data attached to generation**
 - Not to accession because of drift, shift & gene flow
 - (But characterisation data searched by accession)
- **Must track**
 - To monitor & reduce drift, shift & gene flow

Management to prevent accumulation of genetic changes



Seed packet entity

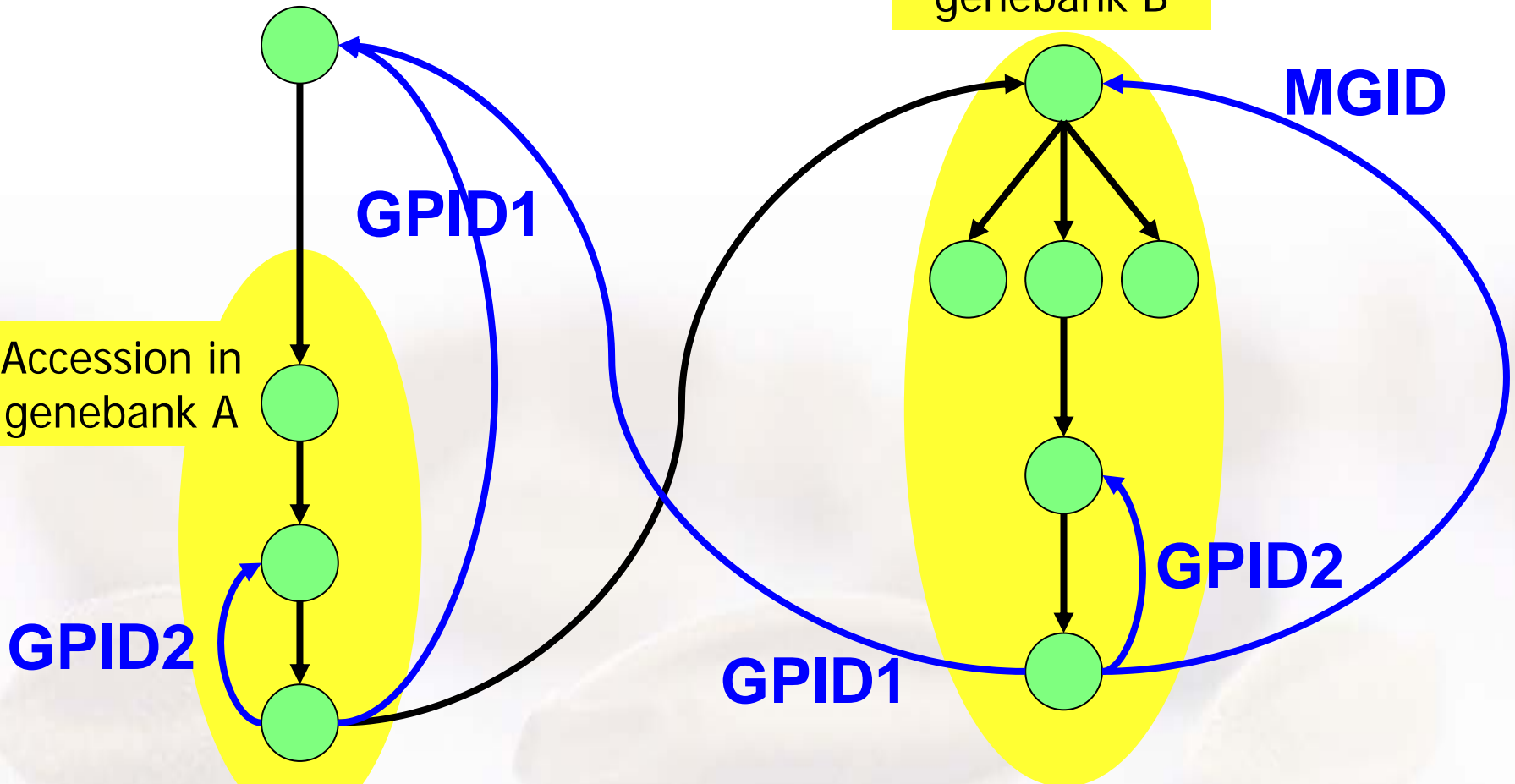
- **Needed for inventory management**
- **Carries data on viability / health / stocks**
- **1 seed packet = 1 LotID**
 - **GID:LotID = 1:n**

Mapping accessions & generations onto ICIS GIDs

Collected sample

Accession in
genebank B

Accession in
genebank A



Passport data in ICIS

- **FAO international standard = Multi-Crop Passport Descriptors MCPD**
 - Crop-specific descriptors & states added as necessary
- **28 descriptors**
 - Identifiers 5
 - Institutes 4
 - Taxonomy 6
 - Origin 8
 - Status 5
- **In ICIS as germplasm attributes unless they fit elsewhere in ICIS**

MCPD germplasm identifiers: All are Names

| Name | MCPD type | ICIS type |
|-----------------------------|------------------|---------------------------------|
| Accession ID | ACCENUMB | ACCNO |
| Collecting number | COLLNUMB | COLLNO |
| Accession name | ACCENAME | RELNM, CVNAM, DRVNM, etc |
| Donor accession ID | DONORNUMB | FACCN |
| Other identification | OTHERNUMB | |

MCPD institutes

| Name | MCPD type | ICIS type |
|----------------------|-----------|-------------------------------------|
| Holding institute | INSTCODE | x (!) |
| Collecting institute | COLLCODE | Within attribute: MISSION_ |
| Breeding institute | BREDCODE | x |
| Donor institute | DONORCODE | ? GID GLOCN + Attribute SS_STATN |

MCPD Taxonomy

- **6 fields:**
 - Genus, Species, Species authority, Subtaxa, Subtaxa authority, Common crop name
- **All as one Germplasm Attribute *TaxNo***
- **Separate Taxon table indexed by TaxNo?**
 - Multiple fields → not suitable for Scale / trait definition?

MCPD Origin

- **4 fields for location of collecting site:**
 - Latitude, Longitude, Altitude, Description
 - All as one Germplasm Attribute CollLocID

| Field | MCPD name | ICIS data |
|-----------------------------|------------------|---------------------------|
| Country of origin | ORIGCTY | Attribute ORI_COUN |
| Source of collection | COLLSRC | Attribute COLL_SOU |
| Date of collection | COLLDATE | Attribute COLL_DAT |
| Date of acquisition | ACQDATE | GID GDATE |

MCPD germplasm status

| Name | MCPD type | ICIS type |
|--------------------------------------|------------------|----------------------------------|
| Biological status | SAMPSTAT | Attribute SAMPLE_STAT |
| Ancestral data / pedigree | ANCEST | Dynamic from GMS |
| Location of safety backup | DUPLSITE | IMS |
| Type of germplasm storage | STORAGE | IMS |
| Remarks | REMARKS | Attributes remarks |

DMS for characterisation

- **~ 80 morpho-agronomic traits**
 - Highly heritable
 - Same traits every year
- **~ 40 years * ~ 2 studies / year**
- **Single replicate, single location**
- **Mostly single year per accession**
- **Replicate checks**

- **All years combined for searches**

Queries to select accessions for use

- **Querying must be efficient**
 - 2005: 39,000 samples distributed
 - case-by-case selection of samples for users' needs
- **Input to each query:**
 - ~ 100,000 accessions * ~ 80 passport descriptors
 - ~ 300,000 GIDs * ~ 80 morpho-agronomic traits
 - Collapse trait data on 300k GIDs to 100k accessions
 - Flexible queries on 100k accessions * 160 chars
- **Perceived difficult with DMS**
 - Searching across studies one extra complexity too many

DMS for germination tests

- Regular germination tests for every accession
- Queries to select accessions for multiplication
- Compare initial vs latest germination
- Ideally base on survivorship curves:



- **Difficult searching across studies**

Definitions (1) Germplasm date

= Germplasm *birth* date

- **ICIS: date seed harvested**

- Accession: acquisition date
- Collected sample: collection date
- GID representing a neighbourhood:
birth date of founding sample
- Birth date of released variety = release date?

Definitions (2) Germplasm Location

= Location of germplasm *at birth*

- **ICIS: place seed harvested**

- Accession: location (country? institute?) of donor (!)
- Collected sample: collection location
- GID representing a neighbourhood:
location of founding sample
- Birth date of RELNM = release date?

Definitions (3) Name location & date

- **ICIS: location and date name assigned**
 - By whom to what?
 - IRIS data inconsistent
- **By original name creator to original GID**
 - Name inherited from source
 - Name location and date inherited from source
 - NDATE predates GDATE
 - New name assigned to this GID
 - NDATE = name date assigned – on or after GDATE

How similar are “duplicate” samples?

What is the COP of:

- **Two samples of Khao Kam (rice landrace)**
 - a. Derived by splitting a sample collected in one field and maintained for n generations in different genebanks;
or
 - b. Derived from samples in different fields
- **Two samples of IR36**
 - a. Both obtained from the INGER sample
 - b. Both obtained from one IRRI genebank accession
 - c. Collected from 2 different farmers' fields

ICIS does not distinguish

IR36 in INGER

| Alternative Names | | Attributes | | | |
|-------------------|-------------------|------------|-------|---------|-------|
| Type | Name | Type | Value | | |
| RELNM | IR 36 | INGER | 00266 | PHL 005 | 1IRGC |
| DRVNM | IR 2071-625-1-252 | IRGD | P | PHL | |
| ITEST | IRTP 266 | | | | |

List/s where this Germplasm is present

| | | | |
|--------------------------------------|--------------------------------------|---|--------------------------------------|
| <input type="checkbox"/> 1975 (1448) | <input type="checkbox"/> 1978 (1582) | <input type="checkbox"/> 1981 (1618) | <input type="checkbox"/> 1983 (1528) |
| <input type="checkbox"/> 1975 (1479) | <input type="checkbox"/> 1978 (1615) | <input type="checkbox"/> 1981 (1675) | <input type="checkbox"/> 1983 (1561) |
| <input type="checkbox"/> 1975 (1553) | <input type="checkbox"/> 1978 (1659) | <input type="checkbox"/> 1981 (1765) | <input type="checkbox"/> 1983 (1620) |
| <input type="checkbox"/> 1975 (1612) | <input type="checkbox"/> 1978 (1698) | <input type="checkbox"/> 1981 (1806) | <input type="checkbox"/> 1983 (1654) |
| <input type="checkbox"/> 1975 (1662) | <input type="checkbox"/> 1978 (1762) | <input type="checkbox"/> 1981 (1812) | <input type="checkbox"/> 1983 (1677) |
| <input type="checkbox"/> 1975 (1695) | <input type="checkbox"/> 1978 (1859) | <input type="checkbox"/> 1981 (1950) | <input type="checkbox"/> 1983 (1742) |
| <input type="checkbox"/> 1976 (1453) | <input type="checkbox"/> 1979 (1456) | <input type="checkbox"/> 1981 (1972) | <input type="checkbox"/> 1983 (1767) |
| <input type="checkbox"/> 1976 (1480) | <input type="checkbox"/> 1979 (1511) | <input type="checkbox"/> 1981 (1975) | <input type="checkbox"/> 1983 (1840) |
| <input type="checkbox"/> 1976 (1508) | <input type="checkbox"/> 1979 (1524) | <input type="checkbox"/> 1981 (1981) | <input type="checkbox"/> 1983 (1954) |
| <input type="checkbox"/> 1976 (1554) | <input type="checkbox"/> 1979 (1557) | <input type="checkbox"/> 1981-101 (266) | <input type="checkbox"/> 1983 (1962) |
| <input type="checkbox"/> 1976 (1580) | <input type="checkbox"/> 1979 (1583) | <input type="checkbox"/> 1982 (1486) | <input type="checkbox"/> 1983 (1987) |
| <input type="checkbox"/> 1976 (1613) | <input type="checkbox"/> 1979 (1616) | <input type="checkbox"/> 1982 (1527) | <input type="checkbox"/> 1983 (1990) |
| <input type="checkbox"/> 1976 (1661) | <input type="checkbox"/> 1979 (1658) | <input type="checkbox"/> 1982 (1560) | <input type="checkbox"/> 1984 (1488) |
| <input type="checkbox"/> 1976 (1696) | <input type="checkbox"/> 1979 (1699) | <input type="checkbox"/> 1982 (1676) | <input type="checkbox"/> 1984 (1562) |
| <input type="checkbox"/> 1976 (1760) | <input type="checkbox"/> 1979 (1763) | <input type="checkbox"/> 1982 (1702) | <input type="checkbox"/> 1984 (1678) |
| <input type="checkbox"/> 1977 (1454) | <input type="checkbox"/> 1979 (1860) | <input type="checkbox"/> 1982 (1766) | <input type="checkbox"/> 1984 (1743) |
| <input type="checkbox"/> 1977 (1481) | <input type="checkbox"/> 1980 (1484) | <input type="checkbox"/> 1982 (1807) | <input type="checkbox"/> 1984 (1768) |
| <input type="checkbox"/> 1977 (1555) | <input type="checkbox"/> 1980 (1525) | <input type="checkbox"/> 1982 (1813) | <input type="checkbox"/> 1984 (1841) |
| <input type="checkbox"/> 1977 (1581) | <input type="checkbox"/> 1980 (1558) | <input type="checkbox"/> 1982 (1953) | <input type="checkbox"/> 1984 (1955) |
| <input type="checkbox"/> 1977 (1614) | <input type="checkbox"/> 1980 (1617) | <input type="checkbox"/> 1982 (1973) | <input type="checkbox"/> 1984 (1963) |



Dark leaves, normal grain

**Both
called
Khao kam**

**+ any combination of dark leaves,
dark grain, dark seed coat**



*Normal leaves,
dark grain*

The need for consistent standards

- **ICIS cannot identify GID as**
 - One seed sample
 - One entire management neighbourhood
 - The “global neighbourhood” of a name
- **Landraces**
 - Basmati, Khao Kam & Azucena in IRIS
 - Each has a GID created to be a common GPID1
 - IRTP4209 is GPID1 of all Azucena
 - = combined generative neighbourhoods of introgressing populations of ancient crosses
 - Better never to join landraces into 1 Root GID?

Collecting samples

GERMPLASM

Variety released

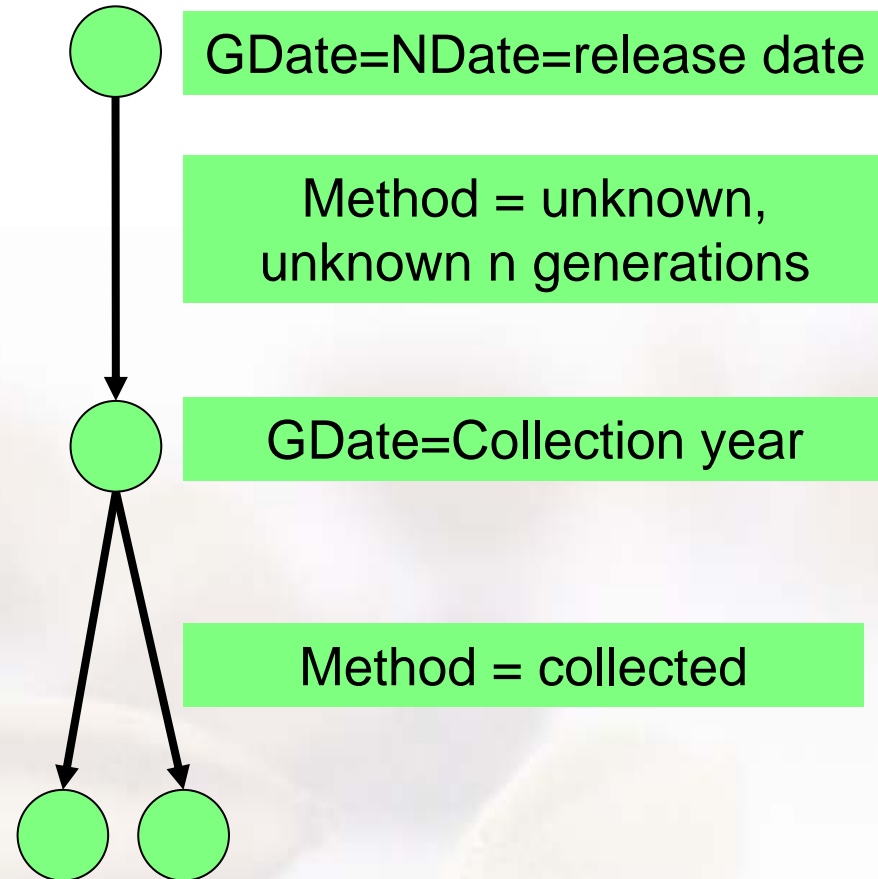
Multiplied by seed company –
unknown no. of generations.

Grown by farmer – maybe farm-saved
unknown no. of generations

Sample collected from farmer's field

Accessions in genebanks

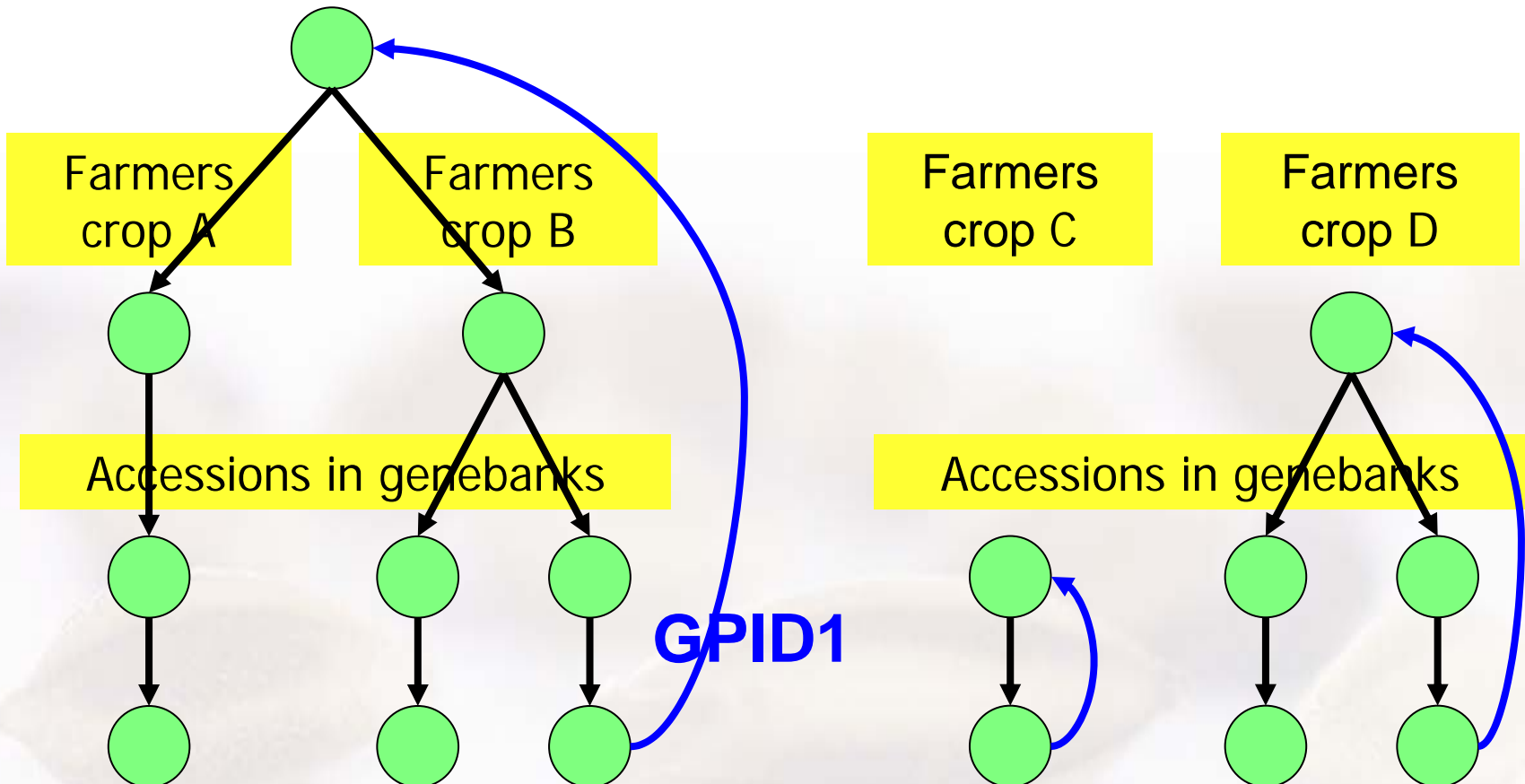
GIDs



GIDs for collected samples

Released variety re-collected from farmer's field
ASSIGN GID TO VARIETY

Land race: < 10,000+ years of unknown selection & introgression
DON'T ASSIGN GID



Conclusions

- **Migrating a conventional genebank database to ICIS a significant challenge**
 - DMS
 - Definitions and procedures
 - Still some issues to be resolved
- **GMS a major functional advance over other genebank management systems**
 - for tracking accessions
 - for tracking generations of an accession
- **On target for success this year**