DYNAVERSITY

DYNAmic seed networks for managing European diVERSITY

Grant agreement n°: 773814

H2020 – Coordination and support action

D 3.5 - D13
Requirements for on farm database
"Technical manual summarizing the requirements for on farm databases"

Due date: M38 (31 January 2021)

Actual submission date: M36

Project start date: November 1st, 2017 **Duration:** 36 months

Work package concerned: WP3

Concerned work package leader: INRA

Dissemination level:

X PU: Public (must be available on the website)

□ CO: Confidential, only for members of the consortium (including the Commission Services)

□ CI: Classified, as referred to in Commission Decision 2001/844/EC

Prepared by: Gea Galluzzi (ARCADIA), Riccardo Bocci (RSR) and Matthias Lorimer (EC-LLD)

With contributions from: DYNAVERSITY Partners and EC-LLD Members

Abstract

WP3 "*Increasing diversity in food systems by breeding*" is led by INRA and one of the goals of the task is to "*sustain collective action and networking on PGRFA, promoting Community Seed Banks and databases*".

In this framework, in order to support collective action, it was decided to give a simple and immediate tool and general knowledge to start a CSBs by a series of 3 handbooks.

The title of the deliverable 3.5 (second manual of the series) is "*Technical manual summarizing the requirements for on farm databases*" although in the manual series it is named "*Documentation systems: a tool for data and information management*"

This document follows the template provided by the European Commission in the Participant Portal.

This deliverable is based on and complying with the following reference documents:

- The GA, Annex I and Annex II (downloadable from the participant portal); and
- The Consortium Agreement (CA).

Table of contents

Introduction

This document, *D 3.5 - "Technical manual summarizing the requirements for on farm databases"* is a deliverable of the DYNAVERSITY project, which is funded by the European Union's Horizon 2020 Programme under Grant Agreement #773814.

DYNAVERSITY aims to increase capacities for in-situ conservation of plant genetic resources by mapping and bringing together all stakeholders involved in the dynamic management of plant genetic resources. The project intends to develop new management and governance models, establish new forms of seed networking and exchange and promote socio-environmental practices.

Technical manual series on Community Seed Banks

Manual 2
Community Seed Banks
Documentation systems: a tool for data and information management

Acknowledgments
This booklet is part of a series of three manuals developed within the European Union Horizon 2020 project “DYNAmic seed networks for managing European diversity” (DYNAVERSITY). The creation of the booklets was possible thanks to the collaboration of project partners and members of the European Coordination Let’s Liberate Diversity! (ECLLD). Special thanks also go to Alexandra Baumgartner and Italo Rondinella for the photographs they kindly authorised us to use.

European Coordination Let’s Liberate Diversity!

“Our diversity is our strength”

ECLLD draws its origins and inspiration from the annual gatherings of the European movement on agricultural biodiversity known as the *Let’s Liberate Diversity! Forums*. Since 2005 the LLD gatherings have become a tradition and they have been organised in many different European countries!
Our vision is to encourage, develop and promote the dynamic management of cultivated biodiversity on farms and in gardens, and our goal is to bring back diversity in our food systems in a socially and economically sustainable way. Such diversification can be achieved by better linking the work of different food system actors (*e.g. farmers, gardeners, citizens, researchers, processors, technicians, small-scale seed companies*), supporting and promoting their knowledge and actions around cultivated biodiversity. In order to achieve this vision, EC-LLD’s objective is to be an open and fruitful space for knowledge exchange and sharing of experiences among its members and the civil society, favouring the dissemination of alternative solutions to those of mainstream agriculture and food systems. Over time, EC-LLD has developed two kinds of events to achieve this objective:
- “*Let’s Liberate Diversity!*”, targeted to the general public and all citizens, it promotes exchange of best practices, experiences and seeds across countries and raises awareness. It is usually organised in a city to enable a broader participation.
- “*Let’s Cultivate Diversity!*”, targeted to farmers, food manufacturers and other food system practitioners with the aim of sharing their knowledge around crop and diversity. It is usually organised on a farm.

Germplasm documentation systems, why are they important?

Conserving germplasm in community seed banks is important for making certain plant varieties and mixtures that are not easily available on the commercial market accessible to a community of interested users, as explored in Manual 1 of this series. However, the value of any collection, including those which are maintained in a community setting, is strongly related to the quantity, type and quality of the information associated with it and to its accessibility to users, who are in this way enabled to make informed choices, based on their specific needs and preferences.
Having a supply of up to date, accurate and reliable information on the conserved accessions is of vital importance also for the efficient operation of the CSB itself, such as: setting priorities, planning activities, and managing resources. An example is germplasm regeneration, which we talked about in Manual 1: since not all accessions can be regenerated each season, priorities have to be set, through questions such as:

Which samples must, at all costs, be regenerated?
Which regenerations are less urgent?
What are the consequences of not regenerating certain samples?

Information about seed stock levels, the viability of seeds and how frequently particular accessions are requested and distributed help answering these questions and making the most appropriate decisions about regeneration.
Depending on the size, nature and scope of different CSBs, a documentation system may be used simply for information storage and retrieval, or for additional processes such as data maintenance (updating existing data), data processing and analysis, and data exchange.

Documentation systems in European CSBs
In a survey conducted among European CSBs within the Diversifood project, the use of databases for the data management of the seed banks was reported by all the seed savers´ organisations in Northern & Central Europe, and by more than half of the initiatives in other areas of Europe, with the only exception of some initiatives in Portugal and Greece, where no database was reported.

How to set up a documentation system for CSBs - 1

The construction of a documentation system requires detailed analysis and planning before the design of any manual or computer-based forms and databases. Six stages can be identified as follows:

1. Have clear background information about the CSBs' set up. This will help define documentation objectives and facilitate resource management, giving essential information about the CSB from which you can later develop documentation objectives. It will also help to make decisions on how best to use available resources.

2. Define priority areas for the documentation effort. You should write down the documentation objectives and list them in order of importance (you'll probably set more documentation objectives than you can cope with at the same time – that's why you have to make priorities). It's important that you identify the priority areas right from the start to avoid problems later on. It's important to identify whether documentation of certain data is essential.

3. Analyse CSB procedures. Having defined your documentation priorities you can start the detailed analysis of the routine CSB procedures, to identify documentation requirements of each and the relations between them. This should explore the different types of data which are generated or used in each procedure, which will in turn allow deciding how best to handle the data e.g. the suitability of computerized and/or manual forms. One can build a flow chart showing the relationship between CSB procedures and information flow, aiding the decision on how to best handle the data. It's useful if you can build up a list of questions that are asked of the documentation system by the users. These questions will help later on in deciding how best to arrange the data for flexible and effective information retrieval and help define documentation procedures.

How to set up a documentation system for CSBs - 2

4: Identify meaningful sets of descriptors. Much of the data you'll be recording will be connected to individual accessions. To facilitate the operation and maintenance of the documentation system, you need to organise the descriptors into practical sets. You can think of these sets as separate books, folders or forms in a manual system or separate files in a computer system (e.g. "characterisation of T. *monococcum*", "viability testing"). These sets are practical in terms of how data are recorded and used and practical in terms of information retrieval.
5: Develop data formats and manual forms and/or computer screen entry forms which facilitate data recording and flexible information retrieval and that will be used at each stage of the documentation process. These forms should favour straightforward data entry, which minimize the risk of errors creeping into the system. If these forms are designed well, the accuracy of data entry is ensured.
6: Develop documentation procedures and implement the new system. A documentation system can be well designed and well-constructed but not used unless there are clearly defined documentation procedures and training is given in the use of the system.

General desirable features of a documentation system

Data integrity. Information retrieved from a documentation system must be accurate, reliable and up to date for it to be of value.
Fast information retrieval. If your system is well designed, retrieving information will be a simple and straightforward process. If it isn't well designed you may spend hours or won't be able to supply the information at all.
User-friendly operation. Data does not appear in a documentation system as if by magic – a CSB member or staff will have entered the data, devoting a fair amount of time filling digital or paper forms. Anything that reduces this workload is really helpful, making the documentation system user-friendly, minimising errors and requiring minimum training.
Flexible operation. The documentation system should not be rigid, but rather be flexible in order to be able to cope with different requests for information and to accommodate changes in CSB activities and routine procedures.
Data organisation. Data are not stored in a documentation system randomly, but rather organised into groups which are practical for data recording, storage, maintenance and information retrieval. Users' needs must be taken into account when organising data into *practical* groups. Part of the skill in designing a documentation system is to define groups which are practical for all operations that are likely to be performed. These practical groups are usually associated with the CSB's routine procedures.

Documentation procedures and priorities

CSB is managed through the execution of a (more or less complex) series of procedures, i.e. sequences of actions which collectively allow to accomplish a desired task. Priority procedures which will be reflected in the CSB's documentation system may include operational procedures and/or scientific procedures.

Examples of operational procedures	**Examples of scientific procedures**
Sample registration (passport and batch data), seed cleaning, seed moisture content determination, seed drying, seed viability testing, seed packaging, seed storage, seed monitoring and regeneration/multiplication, seed distribution	Germplasm characterisation and evaluation, traditional knowledge data, breeding and selection data

Data produced from *operational procedures* (which are often part of a sequence) are of vital importance for the management and continued viability of the collection, so a *high priority* is usually given to their documentation, even in very small CSBs. Not all CSBs will have the capacity to conduct scientific procedures, but when they do, the data generated is likely to be of interest to CSB users or external parties, which makes it important to devise ways of communicating such information using the documentation system itself (data ownership and sharing issues will be discussed later). The capacity to actively experiment with genetic diversity and to flexibly communicate the resulting data within (or even outside) the community are key elements of fulfilling the CSBs' role as catalysts of a truly dynamic and inclusive management model for agricultural biodiversity, rather than acting as simple repositories of static genetic diversity.
When designing a procedure, careful thought must be given to what is being achieved and how best to perform the task with the human and physical resources available. It is also important to take care of the possibility to easily update and modify changing data. Here are some examples of procedures, the activities they imply and the resulting information they entail.

Using descriptors - 1

Within the procedures of a CSB, you will be making many observations on different plant characteristics (weight of a batch of seeds, height of a plant, colour of flowers, etc.). These characteristics can be referred to as descriptors. Descriptors can refer to a quantitative measure (i.e. dealing with numbers such as 56gr, 42mm, 82%) or a qualitative characteristic (e.g. brown, hairy, horizontal). Each descriptor has a number of states which the descriptor can take, either categorical (flower colour) or continuous (seed weight or plant height).
Some examples are given below:

DESCRIPTOR	DESCRIPTOR STATES
Flower colour	white, cream, yellow, orange, green, dark green, red, dark red
Collection source	wild habitat, farm land, farm store, backbyard, village market, commercial market, institute, other
Altitude of collection site	continously variable
Monthly rainfall of collection site	continously variable
Seed germinating in ripe fruit	absent, present

The use of numeric codes greatly facilitates the simple and accurate scoring of descriptors. For instance, it would be time consuming if you had to type or write “village market” as a collection source for a whole series of accessions: it would take up a lot of space on the forms you are using or in the computer database and spelling mistakes might creep in. It is useful to assign a numeric code to each descriptor state: it’s a lot easier to type or write “5” (and to later search and filter for this category) than to write “village market”. There are occasions where non-numeric, letter codes are used (an example is the 3 letter codes for country names recommended by the Statistical Office of the United Nations).
Under this system, all gene bank procedures will have their associated set of descriptors (see table below), ensuring uniformity and reducing error, while also allowing data sharing.

ACCESSION DESCRIPTORS	SEED CLEANING DESCRIPTORS	SEED VIABILITY TESTING DESCRIPTORS	SEED DISTRIBUTION DESCRIPTORS
Accession number Scientific name Variety name Donor name Donor identification number	Accession number Batch reference Date seed cleaned Reference to method Estimate of total seed	Accession number Batch reference Collection type (the collection from which the sample was derived, e.g. base, active)	Accession number Batch reference Date of supply Amount of seed sent Reference to recipient address

Other numbers associated with the accession Acquisition date Date of last regeneration or multiplication	Proportion of empty seed Treatment of seed Operator	Reference to method Date of viability test Viability (%) Operator	Phytosanitary certificate number Export permit number Recipient's import permit number Mail registration number

While CSBs may choose to devise and adopt their own set of descriptors, many will benefit from using standard descriptors devised for institutional gene banks or other genetic resource related fields, such as those developed by Bioversity International (https://www.bioversityinternational.org/e-library/publications/descriptors/) wich are used by a number of seed networks and CSBs in Europe already.

Using descriptors – 2 (heterogeneous and special data)

When some samples in a collection are not genetically uniform but contain a certain amount of variation, how can it be adequately reflected in the documentation system? How can one score the descriptors related to this material? This is an important question for many of the varieties often maintained in a CSB, either landraces or mixtures and evolutionary populations, which are by nature much more variable than commercial or modern varieties. In this particular case, but not only, it is very important to keep track of the raw experimental data (i.e. the original observations). If this is not possible or practical, there are a number of other approaches, such as:

Record as "variable". This is probably the easiest way out because it ignores the problem entirely.

Record the mean or most frequently occurring state. This is another easy way out but loses all the information on the extent of variation.

Record the mean and standard deviation. This is probably the best approach for continuous scale data: it gives an idea of the midpoint of the range of values for a particular descriptor and the extent of variation shown in the sample. It is useful for normally distributed data as useful predictions can be made

Record the frequency of each descriptor state. In the case of flower colour in an accession, the record could be something like: 25% yellow, 10% cream, 50% orange, 15% red and 0% purple. Genetic data often take this form too.

Record the range of variation. For example plant height for a particular accession would be stored as "0.75-1.2m"; this would say something about the total extent of variation but nothing about where most values lie.

Score using a binary scale. This will indicate if an accession is heterogeneous but it will not be able to say how heterogeneous the sample is.

Other data types which are specific to a CSB context are those related to traditional knowledge and to quality characteristics of any given accession (e.g. organoleptic or nutritional traits). However, standard descriptors for the gathering, storage, retrieval and exchange of farmers' knowledge of plants exist, such as those developed by Bioversity International; such list is a first attempt to combine a documentation system traditionally used in controlled environments (gene banks, breeding institutes) with an approach that involves people and their knowledge 'in the field'.

However, another possibility adopted by some European CSBs is to develop their own list of descriptors for the features related to farmers' knowledge and/or qualitative traits which are more relevant to that specific CSB or context. It has to be kept in mind that this limits what type of data can be shared across CSBs and communities.

Data entry forms

Many documentation procedures of a CSB rely on the use of manual forms for recording raw data or scoring descriptors even if a computerized documentation system is (or will be) used. Indeed, sometimes direct data entry is impractical (you are in a boiling hot and very sunny field with no access to electricity), or some data analysis is required before documentation, or some operational difficulties exist (if you are recording data from a chain of procedures you may have to wait for the previous steps to be completed).
Manual forms hence become the basis for the files or tables in the database: the sections of table which always hold the same descriptor are called fields, while records hold different descriptors which relate to a single element. See the following example (referring to seed drying descriptors).

			FIELDS		
RECORDS	**Accension number**	**Batch reference**	**Date of final moisture determinati on**	**1000 seed weight (g)**	**Final moisture content**
	1363	15/10/18	17/09/19	527	7
	1427	08/11/2018	17/09/19	692	6

When you are recording a lot of data for a particular accession, a page layout is more appropriate, since it provides more freedom in where you can place descriptors on the page. With page layouts you can use a mixture of columns or text boxes, multiple choice questions and/or comments to assist the user (see example below).

Computerised databases - 1

In this section, we will focus on the steps to create a database that's functional yet simple to handle. Computerized tools are far more flexible and by now universally available and accessible to any user in a European context, and indeed many of existing European CSBs already use some form of computerized documentation system.

A computer database can be defined as a well-organised set of interrelated data held in one or more files or tables (see an example on the previous page) which are capable of being managed by the same software. Files managed by different softwares are completely separate databases. Most database management softwares allow you to perform the following basic activities:

Enter new data

Modify or delete data

Search and retrieve data for reports

Sort data

Import and export data

Modify the structure of a file in response to changing information needs

The performance of a database will depend on how it is designed, as well as on the capabilities and features of the management software used. Financial aspects (costs of the software and usability for non-technical people) should also be kept in mind, particularly in the context of the resources available to many smaller CSBs. Two main categories of database management software exist: *flat file managers* and *relational database managers.*

Flat file managers are the more simple type of database managers. They consist of a simple, two-dimensional table structure made up of rows and columns, similar to what you might create in Microsoft Excel. Typically they are saved as simple plain text files, and use distinct separation characters called delimiters to define where each column starts and stops. Each database consists of a single file. The basic activities listed above can be performed on each file in a flexible way, but usually on one file at a time, which may be a limitation when you need to work with data from different files at the same time. However, with many currently available flat file managers, it is possible to set up the system so that different files make reference to each other and can be worked with at the same time.

Relational database managers are designed specifically to work on more than one file at the same time. This is achieved by linking two or more separate files through a field which is shared between files (i.e. producing a *relationship* between the files). The shared field is only stored once and the linked files can then be worked on simultaneously. The theory of relational database managers is fairly complex but the above simplified description should be sufficient for our current needs.

Computerised databases – 2

One can use multiple fields to establish links. Through linked files, you may use one file to modify data in another as well as to retrieve information stored in different files, which can then be manipulated into the form you need. One must think carefully about which fields will be used to link files so to avoid duplicating data unnecessarily.

With a computerized system (e.g. FilemakerPro, Mysql), when new records are allocated an accession number, they are simply added to the end of the file, meaning that records can theoretically appear in any order. This random ordering is not a problem for a computerized system as it would be for a manual system: before information is retrieved, the computer can be instructed to sort the records into a specific order, based on a single or multiple fields of choice, for example by accession number, or by batch reference (useful to know which accessions need regeneration) or by both accession number and other fields (e.g. date of last viability test). Records are commonly sorted in alphabetical, numeric (ascending or descending) or date order but different database software may have additional sorting capabilities.

Before building the main database file, it is useful to produce a "data dictionary", which describes the records' structure and the characteristics of each field, with the following information:

Full descriptor name – e.g. "Accession number"

Field name (usually an abbreviation) – e.g. "acc_no" (make it unique, simple, descriptive!)

Field type – e.g. character, numeric, logical, date, etc.

Field description – an explanation of how the field should be used, incl. type of data and its form

Data validation rules – any rules for data validation which apply to the field

Index – is the field indexed, i.e. used for linking with other fields in the database?

Field width (where appropriate): e.g. species = 24 characters; subspecies = 26 characters, etc.

The following example refers to the accession number descriptor:

Exchanging, exporting and analysing data

The ability to exchange data between the documentation system and databases on separate computers or sites or between different applications is a very important consideration, since data exchange considerably increases the value of any documentation system. Different CSBs may have different needs. While smaller or starting CSBs may just need to keep track of seed movements in and out, other CSBs with an extended range of activities and engaged in active experimentation and innovation around seeds, may wish to perform more complex operations on the data they host and generate. Quality and security assurance checks are essential in both cases, but more so when the operations become more elaborate. In general, one may wish to:

Export/import data to/from other database systems

Export data to a word processor or desktop publishing software for the preparation of reports or catalogues

Export data to a spreadsheet or to statistical software in order to perform complex calculations

Import data from spreadsheet or statistical software When allowing for data exchange between different databases, it is important that the same descriptors with the same descriptor states are used in the different databases (it is not impossible to exchange data when descriptors and descriptor states are different, but the process is more complicated). For example, if a descriptor has been recorded in separate files using different scales or a different set of codes, before you can exchange data it needs to be translated to the system of scales or codes used by the recipient file.

One of the advantages of a computerized database is the possibility to flexibly export data for further elaboration, analysis and reporting. This allows a CSB to share the achievements of its biodiversity management system within its own constituency or even beyond it, generating greater awareness on the importance of conserving, documenting and sustainably using seeds. The most commonly used tools for performing a whole variety of statistical and mathematical calculations are spreadsheets, which allow more complex calculations than are possible with a database alone. For example, one could export seed viability and storage data from a CSB's database to a spreadsheet, and calculate the effects of varying the storage parameters on seed viability. Spreadsheets can also be used to print reports or generate graphs using the data they hold. They are also the most common input formats for other statistical software and/or programming tools (including those free of charge, such as R) which allow further elaboration of the data exported from a database..

Keep track of raw data! Notebooks are often used to record the data on a day to day basis from CSB operations, which are then transferred to the documentation system. These notebooks should be kept, not thrown away, and filed according to subject and date so that the raw data can be located later on. The time period that each notebook or file represents should be clearly marked on the cover of the notebook/file (e.g. March 2019 – September 2020). It is also important to make a reference to the source of the raw data in your system. For many operations (regeneration, characteristics, evaluations) you can store the date of the tests in your documentation system so that you can locate the raw data (and its hosting notebook) at a later date.

Building and managing the system

Implementing the documentation system involves drawing together the work described until now, to form a fully operational physical system that meets the specification developed in your analysis. System implementation is a big project, particularly if the system will be used by a number of people or if it will be operating on more than one site. Some CSBs are run in a decentralized manner by small groups of staff/volunteers operating in different areas of a given country, so this issue is rather relevant. It is usually essential to have someone accept responsibility for the system's supervision and management (i.e. a *system manager*) who is available to respond to comments and suggestions from internal and external users, and to deal with any problems.

The following is a simple list of operations that need to be undertaken to complete the implementation of the documentation system:

Develop a screen form for each procedure (makes data entry more user friendly)

Group relevant procedures into logical groups which become "menus"

Develop reports for information required regularly (define criteria for searches and information retrieval)

Explore and implement the data management features of the chosen software (e.g. protect certain fields, identify duplicate records)

Implement data quality and security checkpoints and procedures

Produce a detailed documentation for your system (i.e. a MANUAL!)

Naturally, before launching the system, all those who will be using it actively within the CSB should be appropriately trained in its functionalities! Ideally, training should be an ongoing process. As staff/volunteers leave and new ones join, training will be required. As new computer equipment is added to your system and software updates are implemented, users will require training in the use of this new equipment and software.

Looking forward

CSBs may expand over the years. As they do, internal policies change, new accessions are added, new areas of work undertaken and others phased out, new collaborations are entered into. All these changes will have an effect on the information requirements and the operation of the documentation system. For example, as the areas of work change, new documentation procedures will need to be developed and existing procedures modified. It may be necessary to develop new reports for the changing information needs, new screen forms or menus, to include new procedures or exclude redundant ones.

It would be important to keep the possibility of these changes in mind as much as possible when designing the system, so that the naturally evolving needs of the system can be accommodated as easily as possible in the future.

Software, licences and data management rules

The DIVERSIFOOD project (Horizon2020, 2014-2018), surveyed some of the European seed networks about the types of databases they used. The following list describes the different software and license used by five European seed networks for constructing their CSB databases (details of each are not given here, for sake of simplicity and space, but can be easily found on the web or by consulting computer/programming experts):

Réseau Semences Paysannes (France) uses two databases with different objectives: *Spicilege* which is an open access database online to display characteristics regarding varieties cultivated by farmers or gardeners; *SHiNeMaS* is dedicated to internal CSB use for data management regarding participatory plant breeding programmes and reproduction, selection, diffusion, mixture of seed lots. *ShiNeMaS* uses postgresql, python and django software, under the Affero General Public license.

Rete Semi Rurali (Italy)'s database is constructed under Mysql and handled under an open source licence.

Arche-Noah (Austria) and Red Andaluza de Semillas (Spain) use Microsoft Access, an easy-to-use tool to quickly create customizable database applications

Pro Specie Rara (Switzerland) uses a FileMaker Pro12 database under a dedicated license for Non-Profit/Education purposes.

Any CSB should decide how to manage access to their data, ranging from leaving it fully open to anyone interested to restricting it to a more or less narrow group of users. Data security is an important consideration here: an effective way of ensuring data security is to restrict the number of users who have access to the system, most commonly by using passwords. Indeed, among European CSBs, most organisations place some form of restrictions on who can access their databases, providing passwords and different authorisation levels to different user groups. At the same time, most European CSBs actively and openly disseminate reports, summaries and research results deriving from their work with the genetic resources conserved in their collections.

The inclusive database system of Pro Specie Rara

Pro Specie Rara is a network of about 3500 activists, with seed conservation carried out by 650 private and voluntary seed savers trained through a four days seed propagation course and follow-up yearly training. PSR has integrated most of its material in the Multilateral System of the International Treaty on Plant Genetic Resources (see Manual 3 for details) and the genbank stores duplicates as a backup in a black box system.

The data management system of Swiss-based seed network and CSB Pro Specie Rara (PSR) is a very inclusive example, involving all the stakeholders engaged in activities connected to the use of plant genetic resources, from gardeners to farmers, breeders and retailers. It provides a tool for the description of the seeds, as well as for the monitoring of their conservation and diffusion. Its main components (based on FileMaker) allow describing the seed lots in the collection through increasingly detailed levels of information: species, variety, accession, conservation unit (i.e. the number of accessions cultivated on a specific site or at a specific address). Each new seed lot entering the system is linked to a specific conservation unit and registered in the database. Each outgoing seed lot becomes a new conservation unit when sent to a new seed saver.

While the core of the PSR database follows a linear and hierarchical structure, there are other modules of documentation interlinked with it. Some examples beside the mentioned contacts/sites are projects, field collections, cultural regions, literature, photos, descriptors, recipes, management of mixtures and evolutionary populations. The DB is connected to the PSR website, as well as to the national database (ndb), to which passport data about the PSR collection are uploaded regularly. Besides the international coding system, the DOI[1] system has been partly implemented. Accession and variety status, and their availability, can be checked real time. Dynamic population management activities like mixing genepools and fostering adaptation processes on different sites have recently been integrated in the monitoring system. For the time being, PSR accepts populations or mixtures into their collection by creating a new accession, and attaching all the documentation related to the breeding process that leads to its creation to the passport data. However, it will be particularly important to devise ways to link this information to data management and statistical analysis tools.

Glossary

Accession: a distinct, uniquely identifiable sample of seeds representing a cultivar, breeding line or a population, which is maintained in storage for conservation and use.

Agricultural biodiversity or Agrobiodiversity: the variety and variability of animals, plants and microorganisms that are used directly or indirectly for food and agriculture, forestry and fisheries. It comprises the diversity of genetic resources (varieties, breeds) and species used for food, fodder, fibre, fuel and pharmaceuticals. It also includes the diversity of non-harvested species that support production (soil microorganisms, predators, pollinators), and those in the wider environment that support agro-ecosystems (agricultural, pastoral, forest and aquatic).

Genetic diversity: the genetic variability among or within a sample of individuals of a variety, population or species.

Seed system: an ensemble of individuals, networks, organizations, practices and rules that provide seeds for plant production.

Food system: collaborative network that integrates all components from food production through food consumption based on ecological, social and economic factors and values of a region or sub-region.

Variety: a plant or group of plants selected for desirable characteristics and maintained in cultivation. It may be traditional (i.e. a landrace) and maintained by farmers, or modern and developed through deliberate scientific breeding programmes (i.e. a commercial variety). **Landraces** harbour a degree of genetic variability with a certain genetic integrity that has evolved in cultivation, usually in a traditional agricultural system over long periods, and has adapted to a specific local environment or purpose. They are usually not registered in formal variety lists or registers for commercialisation. **Commercial varieties** are characterised by greater genetic uniformity and are registered in formal (official) variety lists although some which were developed in the past may have since been delisted and made redundant (and can be known as "old" or "historical" varieties).

Population (or '**crop population**'): the term is used here to generally refer to a (large) number of plants in one location (field), in which individual plants are not genetically identical to each other. Two special cases of populations are Composite Cross Populations (CCPs) and varietal mixtures, differing in the way in which they were created, i.e., by crossing in the case of CCPs, and by physical mixing seed of existing varieties in the case of varietal mixtures. Depending on the genetic variation available and the strength and direction of environmental variables, the frequencies of different genotypes in the population will change from season to season, thus CCPs and varietal mixtures are evolving populations.

Plant breeding: the science of changing the traits of plants in order to produce desired characteristics. Plant breeders strive to create a specific outcome of plants and potentially new plant varieties. **Participatory plant breeding** is a form of plant breeding in which farmers, as well as other partners (extension staff, seed producers, traders, NGOs) participate in the development of a new variety. The objective is to produce varieties adapted not only to the physical but also to the socio-economic environment in which they are utilized. In **evolutionary plant breeding**, crop populations with a high level of genetic diversity are subjected to the forces of natural selection: year after year, those plants favored under prevailing growing conditions are expected to contribute more seed to the next generation than plants with lower fitness, thus, evolving crop populations have the capability of adapting to the conditions under which they are grown.

Organic farming: an agricultural method that aims to produce food using natural substances and processes, limiting the environmental impact of food production and encouraging responsible use of energy and natural resources, maintenance of biodiversity and fertility and preservation of regional ecological balances.

Genebank: a type of biorepository which preserves genetic material. For plants, this is done by stocking the seeds (e.g. in a seedbank), or through in vitro storage, or freezing cuttings from the plant.

***In situ* conservation**: the conservation of ecosystems and natural habitats and the maintenance and recovery of viable populations of species in their natural surroundings. In the case of domesticated or cultivated species, it refers to conservation in the surroundings where they have developed their distinctive properties. **On farm conservation** is a dynamic form of crop and animal genetic diversity management in farmers' fields, which allows the processes of evolution under natural and human selection to continue.

***Ex situ* conservation**: the conservation of components of biological diversity outside of their natural habitats.

Recommended readings _DIVERSIFOOD (2018) Community Seed Banks in Europe. Report from a DIVERSIFOOD stakeholder workshop in Rome on September 21st, 2017. http://www.diversifood.eu/community-seed-banks-in-europe/ Vernooy, R.; Sthapit, B.; Bessette, G. (2017). Community seed banks: concept and practice. Facilitator handbook and three associated manuals. https://www.bioversityinternational.org/e-library/publications/detail/community-seed-banks-concept-and-practice/ Vernooy, R.; Shrestha, P.; Sthapit, B. (eds)(2015). Community seed banks: origins, evolution and prospects. Earthscan/Routledge, USA and Canada. https://www.bioversityinternational.org/e-library/publications/detail/community-seed-banks-origins-evolution-and-prospects/ Resources on descriptors can be found at: https://www.bioversityinternational.org/e-library/publications/descriptors/ and https://www.upov.int/portal/index.html.en

Technical manual series on Community Seed Banks

Manual 2

Community Seed Banks

Documentation systems: a tool for data and information management

DYNAVERSITY

Acknowledgments

This booklet is part of a series of three manuals developed within the European Union Horizon 2020 project "DYNAmic seed networks for managing European diversity" (DYNAVERSITY). The creation of the booklets was possible thanks to the collaboration of project partners and members of the European Coordination Let's Liberate Diversity! (ECLLD). Special thanks also go to Alexandra Baumgartner and Italo Rondinella for the photographs they kindly authorised us to use.

European Coordination Let's Liberate Diversity!

LET'S LIBERATE **DIVERSITY!**

"Our diversity is our strength"

ECLLD draws its origins and inspiration from the annual gatherings of the European movement on agricultural biodiversity known as the *Let's Liberate Diversity! Forums*. Since 2005 the LLD gatherings have become a tradition and they have been organised in many different European countries!

Our vision is to encourage, develop and promote the dynamic management of cultivated biodiversity on farms and in gardens, and our goal is to bring back diversity in our food systems in a socially and economically sustainable way. Such diversification can be achieved by better linking the work of different food system actors (*e.g. farmers, gardeners, citizens, researchers, processors, technicians, small-scale seed companies*), supporting and promoting their knowledge and actions around cultivated biodiversity. In order to achieve this vision, EC-LLD's objective is to be an open and fruitful space for knowledge exchange and sharing of experiences among its members and the civil society, favouring the dissemination of alternative solutions to those of mainstream agriculture and food systems. Over time, EC-LLD has developed two kinds of events to achieve this objective:

• "*Let's Liberate Diversity!*", targeted to the general public and all citizens, it promotes exchange of best practices, experiences and seeds across countries and raises awareness. It is usually organised in a city to enable a broader participation.

• "*Let's Cultivate Diversity!*", targeted to farmers, food manufacturers and other food system practitioners with the aim of sharing their knowledge around crop and diversity. It is usually organised on a farm.

DYNAVERSITY

This Manual is the result of the collective work of DYNAVERSITY partners, coordinated by Gea Galluzzi (ARCADIA), with the support of Matthias Lorimer (European Coordination Let's Liberate Diversity) and Riccardo Bocci (Rete Semi Rurali)

Table of contents


working together to conserve irish biodiversity
Irish Seed Savers Association
CROP: BEETROOT
VARIETY: Egyption Flat Top
HARVEST DATE: 08/18
GROWER: ISSA
ORG/NON-ORG:

Germplasm documentation systems, why are they important?

Conserving germplasm in community seed banks is important for making certain plant varieties and mixtures that are not easily available on the commercial market accessible to a community of interested users, as explored in Manual 1 of this series. However, the value of any collection, including those which are maintained in a community setting, is strongly related to the quantity, type and quality of the information associated with it and to its accessibility to users, who are in this way enabled to make informed choices, based on their specific needs and preferences.
Having a supply of up to date, accurate and reliable information on the conserved accessions is of vital importance also for the efficient operation of the CSB itself, such as: setting priorities, planning activities, and managing resources. An example is germplasm regeneration, which we talked about in Manual 1: since not all accessions can be regenerated each season, priorities have to be set, through questions such as:

- Which samples must, at all costs, be regenerated?
- Which regenerations are less urgent?
- What are the consequences of not regenerating certain samples?

Information about seed stock levels, the viability of seeds and how frequently particular accessions are requested and distributed help answering these questions and making the most appropriate decisions about regeneration.
Depending on the size, nature and scope of different CSBs, a documentation system may be used simply for information storage and retrieval, or for additional processes such as data maintenance (updating existing data), data processing and analysis, and data exchange.

Documentation systems in European CSBs
In a survey conducted among European CSBs within the Diversifood project, the use of databases for the data management of the seed banks was reported by all the seed savers´ organisations in Northern & Central Europe, and by more than half of the initiatives in other areas of Europe, with the only exception of some initiatives in Portugal and Greece, where no database was reported.

INFORMATION → **DECISION** → **ACTION**

Note: This manual is designed to be as general and comprehensive as possible, but focuses on documentation systems for seed collections, which are the most common in CSBs.


KNJIŽNICA SJEMENA
• SJEME KOJE STE POSUDILI TREBATE
RAZMNOŽITI I KAD DOZRIJE, VRATITI BAR
ONOLIKO KOLIKO STE DOBILI
• SJEME NIJE NIČIJE VLASNIŠTVO, ALI SVI IMAJU

How to set up a documentation system for CSBs - 1

The construction of a documentation system requires detailed analysis and planning before the design of any manual or computer-based forms and databases. Six stages can be identified as follows:

1. Have clear background information about the CSBs' set up. This will help define documentation objectives and facilitate resource management, giving essential information about the CSB from which you can later develop documentation objectives. It will also help to make decisions on how best to use available resources.

2. Define priority areas for the documentation effort. You should write down the documentation objectives and list them in order of importance (you'll probably set more documentation objectives than you can cope with at the same time – that's why you have to make priorities). It's important that you identify the priority areas right from the start to avoid problems later on. It's important to identify whether documentation of certain data is essential.

3. Analyse CSB procedures. Having defined your documentation priorities you can start the detailed analysis of the routine CSB procedures, to identify documentation requirements of each and the relations between them. This should explore the different types of data which are generated or used in each procedure, which will in turn allow deciding how best to handle the data e.g. the suitability of computerized and/or manual forms. One can build a flow chart showing the relationship between CSB procedures and information flow, aiding the decision on how to best handle the data. It's useful if you can build up a list of questions that are asked of the documentation system by the users. These questions will help later on in deciding how best to arrange the data for flexible and effective information retrieval and help define documentation procedures.

How to set up a documentation system for CSBs - 2

4: Identify meaningful sets of descriptors. Much of the data you'll be recording will be connected to individual accessions. To facilitate the operation and maintenance of the documentation system, you need to organise the descriptors into practical sets. You can think of these sets as separate books, folders or forms in a manual system or separate files in a computer system (e.g. "characterisation of T. *monococcum*", "viability testing"). These sets are practical in terms of how data are recorded and used and practical in terms of information retrieval.

5: Develop data formats and manual forms and/or computer screen entry forms which facilitate data recording and flexible information retrieval and that will be used at each stage of the documentation process. These forms should favour straightforward data entry, which minimize the risk of errors creeping into the system. If these forms are designed well, the accuracy of data entry is ensured.

6: Develop documentation procedures and implement the new system. A documentation system can be well designed and well-constructed but not used unless there are clearly defined documentation procedures and training is given in the use of the system.

General desirable features of a documentation system

Data integrity. Information retrieved from a documentation system must be accurate, reliable and up to date for it to be of value.

Fast information retrieval. If your system is well designed, retrieving information will be a simple and straightforward process. If it isn't well designed you may spend hours or won't be able to supply the information at all.

User-friendly operation. Data does not appear in a documentation system as if by magic – a CSB member or staff will have entered the data, devoting a fair amount of time filling digital or paper forms. Anything that reduces this workload is really helpful, making the documentation system user-friendly, minimising errors and requiring minimum training.

Flexible operation. The documentation system should not be rigid, but rather be flexible in order to be able to cope with different requests for information and to accommodate changes in CSB activities and routine procedures.

Data organisation. Data are not stored in a documentation system randomly, but rather organised into groups which are practical for data recording, storage, maintenance and information retrieval. Users' needs must be taken into account when organising data into *practical* groups. Part of the skill in designing a documentation system is to define groups which are practical for all operations that are likely to be performed. These practical groups are usually associated with the CSB's routine procedures.

Documentation procedures and priorities

CSB is managed through the execution of a (more or less complex) series of procedures, i.e. sequences of actions which collectively allow to accomplish a desired task. Priority procedures which will be reflected in the CSB's documentation system may include operational procedures and/or scientific procedures.

Examples of operational procedures	Examples of scientific procedures
Sample registration (passport and batch data), seed cleaning, seed moisture content determination, seed drying, seed viability testing, seed packaging, seed storage, seed monitoring and regeneration/multiplication, seed distribution	Germplasm characterisation and evaluation, traditional knowledge data, breeding and selection data

Data produced from *operational procedures* (which are often part of a sequence) are of vital importance for the management and continued viability of the collection, so a *high priority* is usually given to their documentation, even in very small CSBs. Not all CSBs will have the capacity to conduct scientific procedures, but when they do, the data generated is likely to be of interest to CSB users or external parties, which makes it important to devise ways of communicating such information using the documentation system itself (data ownership and sharing issues will be discussed later). The capacity to actively experiment with genetic diversity and to flexibly communicate the resulting data within (or even outside) the community are key elements of fulfilling the CSBs' role as catalysts of a truly dynamic and inclusive management model for agricultural biodiversity, rather than acting as simple repositories of static genetic diversity.

When designing a procedure, careful thought must be given to what is being achieved and how best to perform the task with the human and physical resources available. It is also important to take care of the possibility to easily update and modify changing data. Here are some examples of procedures, the activities they imply and the resulting information they entail.

PROCEDURE	ACTIVITY	INFORMATION
Registration	Allocate accession number	Passport data; characterisation data; other
Seed cleaning	Clean seed, remove debris	Comments on sample health
Seed drying	Dry seed to acceptable moisture content	Moisure content (fresh/final); 1000 seeds weight; drying method used
Seed viability	Check that seed has high viability	Date of test; % viability; method used; date of next test
Seed packaging and storage	Securely pack seed and store to ensure lon term viability	Weight of seed; % viability; minimum weight of seed allowed; date of next viability test; moisture content; location in store

Using descriptors - 1

Within the procedures of a CSB, you will be making many observations on different plant characteristics (weight of a batch of seeds, height of a plant, colour of flowers, etc.). These characteristics can be referred to as descriptors. Descriptors can refer to a quantitative measure (i.e. dealing with numbers such as 56gr, 42mm, 82%) or a qualitative characteristic (e.g. brown, hairy, horizontal). Each descriptor has a number of states which the descriptor can take, either categorical (flower colour) or continuous (seed weight or plant height).

Some examples are given below:

DESCRIPTOR	DESCRIPTOR STATES
Flower colour	white, cream, yellow, orange, green, dark green, red, dark red
Collection source	wild habitat, farm land, farm store, backbyard, village market, commercial market, institute, other
Altitude of collection site	continously variable
Monthly rainfall of collection site	continously variable
Seed germinating in ripe fruit	absent, present

The use of numeric codes greatly facilitates the simple and accurate scoring of descriptors. For instance, it would be time consuming if you had to type or write “village market” as a collection source for a whole series of accessions: it would take up a lot of space on the forms you are using or in the computer database and spelling mistakes might creep in. It is useful to assign a numeric code to each descriptor state: it's a lot easier to type or write “5” (and to later search and filter for this category) than to write “village market”. There are occasions where non-numeric, letter codes are used (an example is the 3 letter codes for country names recommended by the Statistical Office of the United Nations).

Under this system, all gene bank procedures will have their associated set of descriptors (see table below), ensuring uniformity and reducing error, while also allowing data sharing.

ACCESSION DESCRIPTORS	SEED CLEANING DESCRIPTORS	SEED VIABILITY TESTING DESCRIPTORS	SEED DISTRIBUTION DESCRIPTORS
Accession number	Accession number	Accession number	Accession number
Scientific name	Batch reference	Batch reference	Batch reference
Variety name	Date seed cleaned	Collection type (the collection from which the sample was derived, e.g. base, active)	Date of supply
Donor name	Reference to method	Reference to method	Amount of seed sent
Donor identification number	Estimate of total seed	Date of viability test	Reference to recipient address
Other numbers associated with the accession	Proportion of empty seed	Viability (%)	Phytosanitary certificate number
Acquisition date	Treatment of seed	Operator	Export permit number
Date of last regeneration or multiplication	Operator		Recipient's import permit number
			Mail registration number

While CSBs may choose to devise and adopt their own set of descriptors, many will benefit from using standard descriptors devised for institutional gene banks or other genetic resource related fields, such as those developed by Bioversity International (https://www.bioversityinternational.org/e-library/publications/descriptors/) wich are used by a number of seed networks and CSBs in Europe already.

Using descriptors – 2 (heterogeneous and special data)

When some samples in a collection are not genetically uniform but contain a certain amount of variation, how can it be adequately reflected in the documentation system? How can one score the descriptors related to this material? This is an important question for many of the varieties often maintained in a CSB, either landraces or mixtures and evolutionary populations, which are by nature much more variable than commercial or modern varieties. In this particular case, but not only, it is very important to keep track of the raw experimental data (i.e. the original observations). If this is not possible or practical, there are a number of other approaches, such as:

- **Record as "variable".** This is probably the easiest way out because it ignores the problem entirely.
- **Record the mean or most frequently occurring state.** This is another easy way out but loses all the information on the extent of variation.
- **Record the mean and standard deviation.** This is probably the best approach for continuous scale data: it gives an idea of the midpoint of the range of values for a particular descriptor and the extent of variation shown in the sample. It is useful for normally distributed data as useful predictions can be made
- **Record the frequency of each descriptor state.** In the case of flower colour in an accession, the record could be something like: 25% yellow, 10% cream, 50% orange, 15% red and 0% purple. Genetic data often take this form too.
- **Record the range of variation.** For example plant height for a particular accession would be stored as "0.75-1.2m"; this would say something about the total extent of variation but nothing about where most values lie.
- **Score using a binary scale.** This will indicate if an accession is heterogeneous but it will not be able to say how heterogeneous the sample is.

Other data types which are specific to a CSB context are those related to traditional knowledge and to quality characteristics of any given accession (e.g. organoleptic or nutritional traits). However, standard descriptors for the gathering, storage, retrieval and exchange of farmers' knowledge of plants exist, such as those developed by Bioversity International; such list is a first attempt to combine a documentation system traditionally used in controlled environments (gene banks, breeding institutes) with an approach that involves people and their knowledge 'in the field'.

However, another possibility adopted by some European CSBs is to develop their own list of descriptors for the features related to farmers' knowledge and/or qualitative traits which are more relevant to that specific CSB or context. It has to be kept in mind that this limits what type of data can be shared across CSBs and communities.

Data entry forms

Many documentation procedures of a CSB rely on the use of manual forms for recording raw data or scoring descriptors even if a computerized documentation system is (or will be) used. Indeed, sometimes direct data entry is impractical (you are in a boiling hot and very sunny field with no access to electricity), or some data analysis is required before documentation, or some operational difficulties exist (if you are recording data from a chain of procedures you may have to wait for the previous steps to be completed).

Manual forms hence become the basis for the files or tables in the database: the sections of table which always hold the same descriptor are called fields, while records hold different descriptors which relate to a single element. See the following example (referring to seed drying descriptors).

FIELDS

RECORDS	Accension number	Batch reference	Date of final moisture determination	1000 seed weight (g)	Final moisture content
→	1363	15/10/18	17/09/19	527	7
→	1427	08/11/2018	17/09/19	692	6

When you are recording a lot of data for a particular accession, a page layout is more appropriate, since it provides more freedom in where you can place descriptors on the page. With page layouts you can use a mixture of columns or text boxes, multiple choice questions and/or comments to assist the user (see example below).

COLLECTION									
Accession number					Origin				
Scientific name									
Acquisition date					1000 seed weight				
Donor ID									
Batch date	BATCH 1			BATCH 2			BATCH 3		
Amount of seed									
Moisture content									
Location									

VIABILITY TESTING											
Batch			Date of test			% viability			Date of next testing		


FRASSINETO 405
GENTIL ROSSO 48
AUTONOMIA A
AUTONOMIA B
VILLA GLORI
96
Mara (MM) (1947)
Abbondanza (MM) (1950)
Legenda

Computerised databases - 1

In this section, we will focus on the steps to create a database that's functional yet simple to handle. Computerized tools are far more flexible and by now universally available and accessible to any user in a European context, and indeed many of existing European CSBs already use some form of computerized documentation system.

A computer database can be defined as a well-organised set of interrelated data held in one or more files or tables (see an example on the previous page) which are capable of being managed by the same software. Files managed by different softwares are completely separate databases. Most database management softwares allow you to perform the following basic activities:

- Enter new data
- Modify or delete data
- Search and retrieve data for reports
- Sort data
- Import and export data
- Modify the structure of a file in response to changing information needs

The performance of a database will depend on how it is designed, as well as on the capabilities and features of the management software used. Financial aspects (costs of the software and usability for non-technical people) should also be kept in mind, particularly in the context of the resources available to many smaller CSBs. Two main categories of database management software exist: *flat file managers* and *relational database managers.*

Flat file managers are the more simple type of database managers. They consist of a simple, two-dimensional table structure made up of rows and columns, similar to what you might create in Microsoft Excel. Typically they are saved as simple plain text files, and use distinct separation characters called delimiters to define where each column starts and stops. Each database consists of a single file. The basic activities listed above can be performed on each file in a flexible way, but usually on one file at a time, which may be a limitation when you need to work with data from different files at the same time. However, with many currently available flat file managers, it is possible to set up the system so that different files make reference to each other and can be worked with at the same time.

Relational database managers are designed specifically to work on more than one file at the same time. This is achieved by linking two or more separate files through a field which is shared between files (i.e. producing a *relationship* between the files). The shared field is only stored once and the linked files can then be worked on simultaneously. The theory of relational database managers is fairly complex but the above simplified description should be sufficient for our current needs.


Sachet GRATUIT
Semailles

Computerised databases – 2

In a genetic resources documentation system, the accession number is a field often used to link files together (here it links three files, each related to a different procedure)

Registration	Passport	Characterisation
Accession number*	Accession number*	Accession number*
Scientific name	Collector's number	Batch reference
Crop name	Collecting insitute	Sowing date
Donor number	Date of collection	Soil type
Acquisition date	Country of collection	soil ph
Other fields	Other fields	Other fields

One can use multiple fields to establish links. Through linked files, you may use one file to modify data in another as well as to retrieve information stored in different files, which can then be manipulated into the form you need. One must think carefully about which fields will be used to link files so to avoid duplicating data unnecessarily.

With a computerized system (e.g. FilemakerPro, Mysql), when new records are allocated an accession number, they are simply added to the end of the file, meaning that records can theoretically appear in any order. This random ordering is not a problem for a computerized system as it would be for a manual system: before information is retrieved, the computer can be instructed to sort the records into a specific order, based on a single or multiple fields of choice, for example by accession number, or by batch reference (useful to know which accessions need regeneration) or by both accession number and other fields (e.g. date of last viability test). Records are commonly sorted in alphabetical, numeric (ascending or descending) or date order but different database software may have additional sorting capabilities.

Before building the main database file, it is useful to produce a "data dictionary", which describes the records' structure and the characteristics of each field, with the following information:

- Full descriptor name – e.g. "Accession number"
- Field name (usually an abbreviation) – e.g. "acc_no" (make it unique, simple, descriptive!)
- Field type – e.g. character, numeric, logical, date, etc.
- Field description – an explanation of how the field should be used, incl. type of data and its form
- Data validation rules – any rules for data validation which apply to the field
- Index – is the field indexed, i.e. used for linking with other fields in the database?
- Field width (where appropriate): e.g. species = 24 characters; subspecies = 26 characters, etc.

The following example refers to the accession number descriptor:

Descriptor name	Accession number
Field name	acc_no
Data type	Numeric
Data validation	Between 1 and 999
Field description	Unique identifier assigned to the accession when it enters the collection
Field width	6
Index	v

Exchanging, exporting and analysing data

The ability to exchange data between the documentation system and databases on separate computers or sites or between different applications is a very important consideration, since data exchange considerably increases the value of any documentation system. Different CSBs may have different needs. While smaller or starting CSBs may just need to keep track of seed movements in and out, other CSBs with an extended range of activities and engaged in active experimentation and innovation around seeds, may wish to perform more complex operations on the data they host and generate. Quality and security assurance checks are essential in both cases, but more so when the operations become more elaborate. In general, one may wish to:

- Export/import data to/from other database systems
- Export data to a word processor or desktop publishing software for the preparation of reports or catalogues
- Export data to a spreadsheet or to statistical software in order to perform complex calculations
- Import data from spreadsheet or statistical software

When allowing for data exchange between different databases, it is important that the same descriptors with the same descriptor states are used in the different databases (it is not impossible to exchange data when descriptors and descriptor states are different, but the process is more complicated). For example, if a descriptor has been recorded in separate files using different scales or a different set of codes, before you can exchange data it needs to be translated to the system of scales or codes used by the recipient file.

One of the advantages of a computerized database is the possibility to flexibly export data for further elaboration, analysis and reporting. This allows a CSB to share the achievements of its biodiversity management system within its own constituency or even beyond it, generating greater awareness on the importance of conserving, documenting and sustainably using seeds. The most commonly used tools for performing a whole variety of statistical and mathematical calculations are spreadsheets, which allow more complex calculations than are possible with a database alone. For example, one could export seed viability and storage data from a CSB's database to a spreadsheet, and calculate the effects of varying the storage parameters on seed viability. Spreadsheets can also be used to print reports or generate graphs using the data they hold. They are also the most common input formats for other statistical software and/or programming tools (including those free of charge, such as R) which allow further elaboration of the data exported from a database..

Keep track of raw data!

Notebooks are often used to record the data on a day to day basis from CSB operations, which are then transferred to the documentation system. These notebooks should be kept, not thrown away, and filed according to subject and date so that the raw data can be located later on. The time period that each notebook or file represents should be clearly marked on the cover of the notebook/ file (e.g. March 2019 – September 2020). It is also important to make a reference to the source of the raw data in your system. For many operations (regeneration, characteristics, evaluations) you can store the date of the tests in your documentation system so that you can locate the raw data (and its hosting notebook) at a later date.

seed savers

seed saver is maintaining the variety and propagating it for the seed library

seed saver is offering seeds for free in the seed variety finder (catalogue)

Person is selling variety (seedlings, products) without conservation (making own seeds)

first provenience of the seeds. It must be an approved source (= prospecierara. label owner)

last check (checklist control)

Rebekka Herzog

5277 Hottwil

ID 16024

Email für Webseitenlogin: lycopersicon@bluewin.c

Checkliste | Gütesiegel | Doku

Erhaltungsliste | Sortenliste | nur Tiere | Archiv anzeigen | nur mit Erhaltung

☐ Ackerpflanzen ☐ Beeren ☐ Gemüse ☐ Kartoffeln ☐ Kräuter ☐ Obst ☐ Reben ☐ Weiden ☐ Zierpflanzen

124

+ PSR-Nr. Sorte	Sorte / Rasse	Art-Untergruppe / Kat.	Erhaltung	SF Angebot	Verm.	Nr. Herk./Quell.	Herkunftsname / Quelle	Pflanze/Ort	Seit	Aktiv.meld.	OK	unklar	PSR-Check
GE-15	Chioggia	Rande/Rote Bete	nein			12171	Wyss Samen- und	Zuchwil	2013	17.12.18			
GE-26	Küttiger Rüebli	Karotte	nein			2577	Zollinger Bio	Les	2013	17.12.18			
GE-82	Halblange	Wurzelpetersilie	nein			3667	Sativa Rheinau AG 154	Rheinau	2012	17.12.18			
GE-109	Berner Landfrauen	Stangenbohne	nein			16024	Rebekka Herzog 98	Hottwil	2008	17.12.18			
GE-151	Gezahnte Bührer-Keel	Fleischtomate	nein			16024	Rebekka Herzog 98	Hottwil	1999	17.12.18			
GE-158	Baselbieter Röteli	Kirschtomate	nein			16024	Rebekka Herzog 98	Hottwil	2002	17.12.18			
GE-185	Matina	Mittelfruchtige Tomate	ja			A-GE-000185	Matina		2007	03.09.18			
GE-233	Weinländerin	Stangenbohne	nein			16024	Rebekka Herzog 98	Hottwil	2012	17.12.18			
GE-277	Rotgelb gestreifte Runde	Mittelfruchtige Tomate	nein			16024	Rebekka Herzog 98	Hottwil	1998	17.12.18			
GE-347	Gelbe Cherry	Kirschtomate	ja			A-GE-000347	Gelbe Cherry		2002	03.09.18			
GE-348	Gelbe Birnentomate	Kirschtomate	ja			A-GE-000348	Gelbe Birnentomate		2001	03.09.18			
GE-405	Golden	Rande/Rote Bete	nein			12171	Wyss Samen- und	Zuchwil	2013	17.12.18			
GE-406	Wiener Riesen	Knollensellerie	nein			206835	Sativa Rheinau AG	Rheinau	2018				
GE-413	Howard Deutsche Tomate	Mittelfruchtige Tomate	ja			A-GE-000413	Howard Deutsche		2001	03.09.18			
GE-414	Amish Pasta	Fleischtomate	ja			A-GE-000414	Amish Pasta		1999	03.09.18			
GE-438	Green Zebra	Fleischtomate	ja			A-GE-000438	Green Zebra		1999	03.09.18			
GE-574	Ampeltomate himbeerfarbig	Kirschtomate	ja			A-GE-000574	Ampeltomate		2014	03.09.18			
GE-576	De Paudex	Fleischtomate	nein			16024	Rebekka Herzog 98	Hottwil	2012	17.12.18			
GE-581	Landsorte Chez-le-Bart	Kopfsalat	nein			3667	Sativa Rheinau AG 154	Rheinau	2013	17.12.18			
GE-586	Schweizer Riesen	Kefe/Zuckererbse	nein			16024	Rebekka Herzog 98	Hottwil	2010	17.12.18			
GE-589	Jaune longue du Doubs	Karotte	nein			3667	Sativa Rheinau AG 154	Rheinau	2006	17.12.18			
GE-610	Mandarin	Fleischtomate	ja			A-GE-000610	Mandarin		2002	03.09.18			

Building and managing the system

Implementing the documentation system involves drawing together the work described until now, to form a fully operational physical system that meets the specification developed in your analysis. System implementation is a big project, particularly if the system will be used by a number of people or if it will be operating on more than one site. Some CSBs are run in a decentralized manner by small groups of staff/volunteers operating in different areas of a given country, so this issue is rather relevant. It is usually essential to have someone accept responsibility for the system's supervision and management (i.e. a *system manager*) who is available to respond to comments and suggestions from internal and external users, and to deal with any problems.

The following is a simple list of operations that need to be undertaken to complete the implementation of the documentation system:

- Develop a screen form for each procedure (makes data entry more user friendly)
- Group relevant procedures into logical groups which become "menus"
- Develop reports for information required regularly (define criteria for searches and information retrieval)
- Explore and implement the data management features of the chosen software (e.g. protect certain fields, identify duplicate records)
- Implement data quality and security checkpoints and procedures
- Produce a detailed documentation for your system (i.e. a MANUAL!)

Naturally, before launching the system, all those who will be using it actively within the CSB should be appropriately trained in its functionalities! Ideally, training should be an ongoing process. As staff/volunteers leave and new ones join, training will be required. As new computer equipment is added to your system and software updates are implemented, users will require training in the use of this new equipment and software.

Looking forward

CSBs may expand over the years. As they do, internal policies change, new accessions are added, new areas of work undertaken and others phased out, new collaborations are entered into. All these changes will have an effect on the information requirements and the operation of the documentation system. For example, as the areas of work change, new documentation procedures will need to be developed and existing procedures modified. It may be necessary to develop new reports for the changing information needs, new screen forms or menus, to include new procedures or exclude redundant ones.

It would be important to keep the possibility of these changes in mind as much as possible when designing the system, so that the naturally evolving needs of the system can be accommodated as easily as possible in the future.

Software, licences and data management rules

The DIVERSIFOOD project (Horizon2020, 2014-2018), surveyed some of the European seed networks about the types of databases they used. The following list describes the different software and license used by five European seed networks for constructing their CSB databases (details of each are not given here, for sake of simplicity and space, but can be easily found on the web or by consulting computer/programming experts):

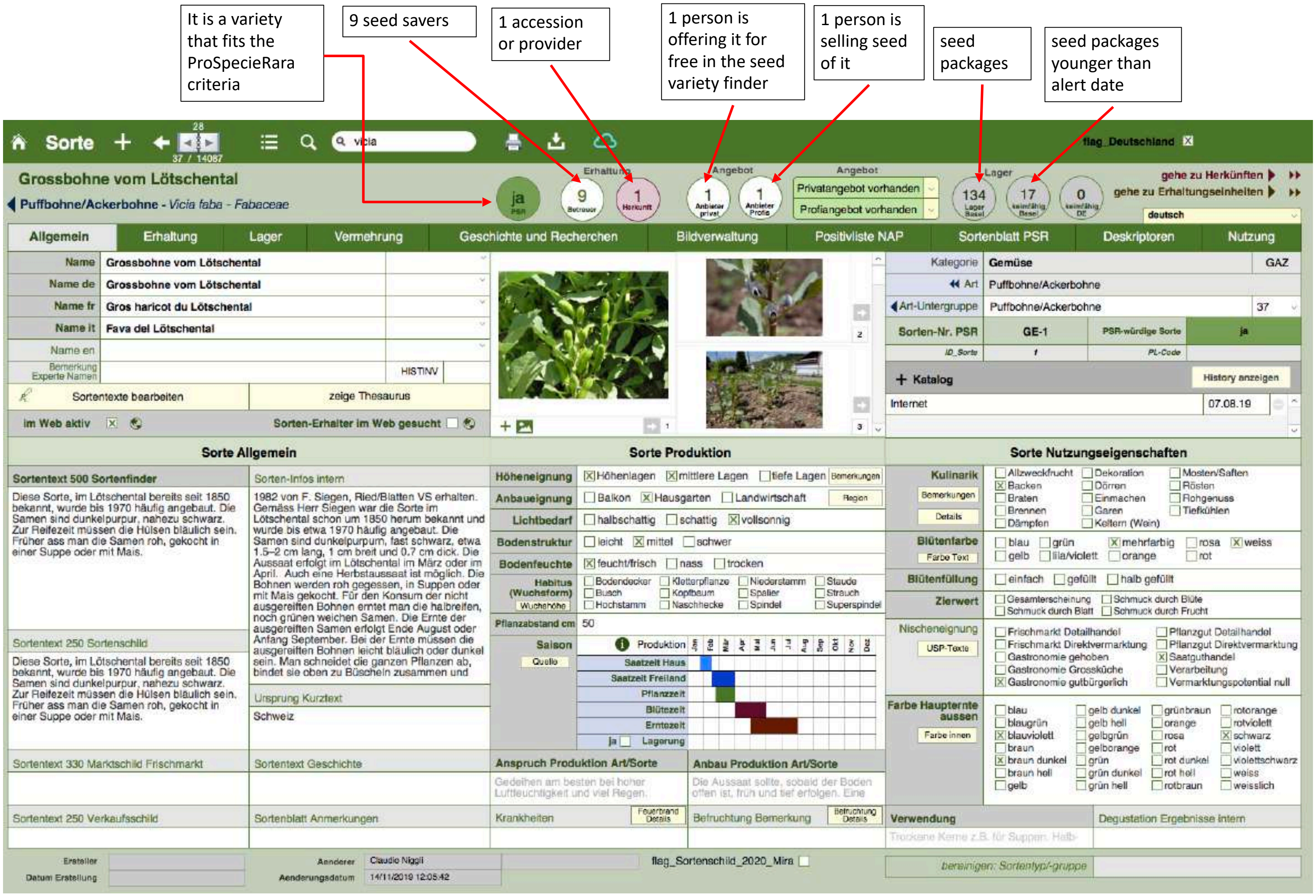
- Réseau Semences Paysannes (France) uses two databases with different objectives: *Spicilege* which is an open access database online to display characteristics regarding varieties cultivated by farmers or gardeners; *SHiNeMaS* is dedicated to internal CSB use for data management regarding participatory plant breeding programmes and reproduction, selection, diffusion, mixture of seed lots. *ShiNeMaS* uses postgresql, python and django software, under the Affero General Public license.

- Rete Semi Rurali (Italy)'s database is constructed under Mysql and handled under an open source licence.

- Arche-Noah (Austria) and Red Andaluza de Semillas (Spain) use Microsoft Access, an easy-to-use tool to quickly create customizable database applications

- Pro Specie Rara (Switzerland) uses a FileMaker Pro12 database under a dedicated license for Non-Profit/Education purposes.

Any CSB should decide how to manage access to their data, ranging from leaving it fully open to anyone interested to restricting it to a more or less narrow group of users. Data security is an important consideration here: an effective way of ensuring data security is to restrict the number of users who have access to the system, most commonly by using passwords. Indeed, among European CSBs, most organisations place some form of restrictions on who can access their databases, providing passwords and different authorisation levels to different user groups. At the same time, most European CSBs actively and openly disseminate reports, summaries and research results deriving from their work with the genetic resources conserved in their collections.

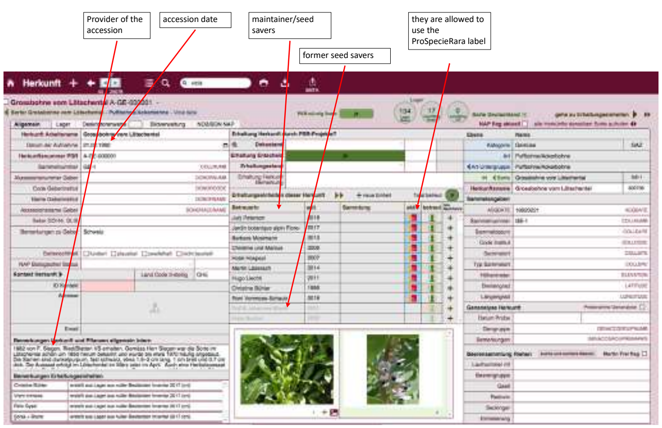

It is a variety that fits the ProSpecieRara criteria
9 seed savers
1 accession or provider
1 person is offering it for free in the seed variety finder
1 person is selling seed of it
seed packages
seed packages younger than alert date
Sorte
Grossbohne vom Lötschental
Puffbohne/Ackerbohne - Vicia faba - Fabaceae
Allgemein
Erhaltung
Lager
Vermehrung
Geschichte und Recherchen
Bildverwaltung
Positivliste NAP
Sortenblatt PSR
Deskriptoren
Nutzung
Sorte Allgemein
Sorte Produktion
Sorte Nutzungseigenschaften

The inclusive database system of Pro Specie Rara

Pro Specie Rara is a network of about 3500 activists, with seed conservation carried out by 650 private and voluntary seed savers trained through a four days seed propagation course and follow-up yearly training. PSR has integrated most of its material in the Multilateral System of the International Treaty on Plant Genetic Resources (see Manual 3 for details) and the genbank stores duplicates as a backup in a black box system.

The data management system of Swiss-based seed network and CSB Pro Specie Rara (PSR) is a very inclusive example, involving all the stakeholders engaged in activities connected to the use of plant genetic resources, from gardeners to farmers, breeders and retailers. It provides a tool for the description of the seeds, as well as for the monitoring of their conservation and diffusion. Its main components (based on FileMaker) allow describing the seed lots in the collection through increasingly detailed levels of information: species, variety, accession, conservation unit (i.e. the number of accessions cultivated on a specific site or at a specific address). Each new seed lot entering the system is linked to a specific conservation unit and registered in the database. Each outgoing seed lot becomes a new conservation unit when sent to a new seed saver.

While the core of the PSR database follows a linear and hierarchical structure, there are other modules of documentation interlinked with it. Some examples beside the mentioned contacts/sites are projects, field collections, cultural regions, literature, photos, descriptors, recipes, management of mixtures and evolutionary populations. The DB is connected to the PSR website, as well as to the national database (ndb), to which passport data about the PSR collection are uploaded regularly. Besides the international coding system, the DOI[1] system has been partly implemented. Accession and variety status, and their availability, can be checked real time. Dynamic population management activities like mixing genepools and fostering adaptation processes on different sites have recently been integrated in the monitoring system. For the time being, PSR accepts populations or mixtures into their collection by creating a new accession, and attaching all the documentation related to the breeding process that leads to its creation to the passport data. However, it will be particularly important to devise ways to link this information to data management and statistical analysis tools.



This table shows how a conservation unit (how many trees of an accession of a given variety are present in one orchard at a specific locality) and its associated information is described in PSR's db. Other tables in the system describe the activity of individual members of the network, and their seed/seedling offer.

[1] Digital Object Identifier

Glossary

Accession: a distinct, uniquely identifiable sample of seeds representing a cultivar, breeding line or a population, which is maintained in storage for conservation and use.

Agricultural biodiversity or Agrobiodiversity: the variety and variability of animals, plants and microorganisms that are used directly or indirectly for food and agriculture, forestry and fisheries. It comprises the diversity of genetic resources (varieties, breeds) and species used for food, fodder, fibre, fuel and pharmaceuticals. It also includes the diversity of non-harvested species that support production (soil microorganisms, predators, pollinators), and those in the wider environment that support agro-ecosystems (agricultural, pastoral, forest and aquatic).

Genetic diversity: the genetic variability among or within a sample of individuals of a variety, population or species.

Seed system: an ensemble of individuals, networks, organizations, practices and rules that provide seeds for plant production.

Food system: collaborative network that integrates all components from food production through food consumption based on ecological, social and economic factors and values of a region or sub-region.

Variety: a plant or group of plants selected for desirable characteristics and maintained in cultivation. It may be traditional (i.e. a landrace) and maintained by farmers, or modern and developed through deliberate scientific breeding programmes (i.e. a commercial variety). **Landraces** harbour a degree of genetic variability with a certain genetic integrity that has evolved in cultivation, usually in a traditional agricultural system over long periods, and has adapted to a specific local environment or purpose. They are usually not registered in formal variety lists or registers for commercialisation. **Commercial varieties** are characterised by greater genetic uniformity and are registered in formal (official) variety lists although some which were developed in the past may have since been delisted and made redundant (and can be known as “old” or “historical” varieties).

Population (or ‘**crop population**’): the term is used here to generally refer to a (large) number of plants in one location (field), in which individual plants are not genetically identical to each other. Two special cases of populations are Composite Cross Populations (CCPs) and varietal mixtures, differing in the way in which they were created, i.e., by crossing in the case of CCPs, and by physical mixing seed of existing varieties in the case of varietal mixtures. Depending on the genetic variation available and the strength and direction of environmental variables, the frequencies of different genotypes in the population will change from season to season, thus CCPs and varietal mixtures are evolving populations.

Plant breeding: the science of changing the traits of plants in order to produce desired characteristics. Plant breeders strive to create a specific outcome of plants and potentially new plant varieties. **Participatory plant breeding** is a form of plant breeding in which farmers, as well as other partners (extension staff, seed producers, traders, NGOs) participate in the development of a new variety. The objective is to produce varieties adapted not only to the physical but also to the socio-economic environment in which they are utilized. In **evolutionary plant breeding**, crop populations with a high level of genetic diversity are subjected to the forces of natural selection: year after year, those plants favored under prevailing growing conditions are expected to contribute more seed to the next generation than plants with lower fitness, thus, evolving crop populations have the capability of adapting to the conditions under which they are grown.

Organic farming: an agricultural method that aims to produce food using natural substances and processes, limiting the environmental impact of food production and encouraging responsible use of energy and natural resources, maintenance of biodiversity and fertility and preservation of regional ecological balances.

Genebank: a type of biorepository which preserves genetic material. For plants, this is done by stocking the seeds (e.g. in a seedbank), or through in vitro storage, or freezing cuttings from the plant.

***In situ* conservation**: the conservation of ecosystems and natural habitats and the maintenance and recovery of viable populations of species in their natural surroundings. In the case of domesticated or cultivated species, it refers to conservation in the surroundings where they have developed their distinctive properties. **On farm conservation** is a dynamic form of crop and animal genetic diversity management in farmers' fields, which allows the processes of evolution under natural and human selection to continue.

***Ex situ* conservation**: the conservation of components of biological diversity outside of their natural habitats.

Recommended readings

DIVERSIFOOD (2018) Community Seed Banks in Europe. Report from a DIVERSIFOOD stakeholder workshop in Rome on September 21st, 2017. http://www.diversifood.eu/community-seed-banks-in-europe/

Vernooy, R.; Sthapit, B.; Bessette, G. (2017). Community seed banks: concept and practice. Facilitator handbook and three associated manuals. https://www.bioversityinternational.org/e-library/publications/detail/community-seed-banks-concept-and-practice/

Vernooy, R.; Shrestha, P.; Sthapit, B. (eds)(2015). Community seed banks: origins, evolution and prospects. Earthscan/Routledge, USA and Canada. https://www.bioversityinternational.org/e-library/publications/detail/community-seed-banks-origins-evolution-and-prospects/

Resources on descriptors can be found at: https://www.bioversityinternational.org/e-library/publications/descriptors/ and https://www.upov.int/portal/index.html.en

DYNAVERSITY

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n. 773814

LET'S LIBERATE **DIVERSITY!**