SHiNeMaS: A web tool dedicated to seed lots history, phenotyping and cultural practices

Yannick De Oliveira (✉ yannick.de-oliveira@inrae.fr)
INRA  https://orcid.org/0000-0002-6902-0005
Laura Burlot
INRAE
Julie C. Dawson
University of Wisconsin Madison Department of Horticulture
Isabelle Goldringer
INRAE
Darkawi Madi
INRAE
Pierre Rivière
Self Employed
Delphine Steinbach
INRAE
Gaëlle Van Frank
INRAE
Mathieu Thomas
CIRAD

Software

Keywords: Database; Plant breeding; Software

Posted Date: June 20th, 2020

DOI: https://doi.org/10.21203/rs.2.21522/v3

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published on July 23rd, 2020. See the published version at https://doi.org/10.1186/s13007-020-00640-2.
SOFTWARE

SHiNeMaS : A web tool dedicated to seed lots history, phenotyping and cultural practices

Yannick De Oliveira¹*, Laura Burlot¹,², Julie C. Dawson³, Isabelle Goldringer¹, Darkawi Madi¹, Pierre Rivière⁴, Delphine Steinbach¹, Gaëlle van Frank¹ and Mathieu Thomas⁵,⁶

Abstract

Motivation: In 2005, researchers from the French National Research Institute for Agriculture, Food and Environment (Institut national de recherche pour l’agriculture, l’alimentation et l’environnement, INRAE) started a collaboration with the French farmers’ seed network Réseau Semences Paysannes (RSP) on bread wheat participatory breeding (PPB). The aims were: (1) to study on-farm management of crop diversity, (2) to develop population-varieties adapted to organic and low-inputs agriculture, (3) to co-develop tools and methods adapted to on-farm experiments. In this project, researchers and farmers’ organizations needed to map the history and life cycle of the population-varieties using network formalism to represent relationships between seed lots. All this information had to be centralized and stored in a database.

Results: We describe here SHiNeMaS (Seeds History and Network Management System) a web tool database. SHiNeMaS aims to provide useful interfaces to track seed lot history and related data (phenotyping, environment, cultural practices). Although SHiNeMaS has been developed in the context of a bread wheat participatory breeding program, the database has been designed to manage any kind and even multiple cultivated plant species. SHiNeMaS is available under Affero GPL licence and uses free technologies such as the Python language, Django framework or PostgreSQL database management system (DBMS).

Conclusion: We developed SHiNeMaS, a web tool database, dedicated to the management of the history of seed lots and related data like phenotyping, environmental information and cultural practices. SHiNeMaS has been used in production in our laboratory for 5 years and farmers’ organizations facilitators manage their own information in the system.

Keywords: Database; Plant breeding; Software

Background

Due to a lack of effort in breeding varieties adapted to organic and low-input agriculture [1] farmers currently have limited access to the proper crop diversity. Cultivated biodiversity has declined in the landscape for decades as a result of the industrialization of agriculture [2]. For these reasons, there is a need to deploy a cultivated diversity adapted to diverse practices and contexts.

In order to create diverse heterogeneous varieties of bread wheat adapted to farmers’ practices and needs, researchers from the French National Research Institute for Agriculture, Food and Environment (INRAE) and the wheat group of the French farmers’ seed network Réseau Semences Paysannes (RSP) started a collaborative program in 2005. The aims of this program are (i) to study on-farm management of crop diversity [3], (ii) to develop population-varieties adapted to organic and low-inputs agriculture in the context of a participatory plant breeding program involving a collective of farmers, facilitators and researchers [4], [5] and, (iii) to co-develop methods and tools such as data management tools, experimental designs and statistical methods adapted to on-farm experiments [6, 7], in order to foster genetic and social innovations.

In this context, participants have (i) developed, evaluated, and measured at global and individual levels new population-varieties [6, 8, 9] and (ii) fostered seed and knowledge circulation among participants in a large network of farms (more than 100 farms in 2018). This program therefore produces a lot of data, and, in particular, relational information between seed lots (trace-
ability) that can be represented as a network (figure 1). Thus several dozen users involved in the program, the researchers and facilitators needed a tool adapted to the monitoring of decentralised experimentation, accessible online and with user-friendly interfaces, to facilitate the management of the data produced. Moreover, it is important for us to develop a tool distributed under a free licence, free of charge and using free technologies. Some tools we identified like PlantDB [10], Phytotrack [11] or SeedUSoon [12] manages seed lots and their stock information but doesn’t fit our requirements as they are not adapted to field experimentation. For instance, PlantDB is mainly related to genotyping data management but doesn’t provide a system to collect data from the field. Phytotrack provides features to design plant crosses but no additional data (trait description) can be recorded. Moreover, Phytotrack and PlantDB use proprietary technologies and are operating system dependant. SeedUSoon fills this technological gap as it uses Java and MySQL DBMS, but it offers a limited number of scenarios. For instance, mixture, diffusion or selection can’t be handled in this tool. Finally, none of them provide feature to load data from a file, a mobile app or a web application making these tools unsuitable for field experimentation. Another one, GnpIS [13], manages seed lots and germplasm, but doesn’t ensure seed lots traceability, it doesn’t provide interfaces for data management and consequently, can’t be considered as a data management tool. The last, the Breeding Management System [14] provides a lot of useful features but does not ensure complete traceability of seed lots and a part of its component are under proprietary licence. To fill the gaps in these tools and to avoid expensive work to improve one of them to fill our requirements, we developed SHiNeMaS (Seeds History and Network Management System), a web tool to manage the network and history of seed lots and related data.

Implementation

SHiNeMaS has been created to support any kind of cultivated species such as cereals, fodder and vegetable plants. The web application has been designed to satisfy the needs of researchers, farmers’ organization facilitators, farmers and gardeners. Data access is protected by a system of user accounts with two levels of authorization: read/write user for data management and read-only user for data querying.

Results and discussion

SHiNeMaS has been created to support any kind of cultivated species such as cereals, fodder and vegetable plants. The web application has been designed to satisfy the needs of researchers, farmers’ organization facilitators, farmers and gardeners. Data access is protected by a system of user accounts with two levels of authorization: read/write user for data management and read-only user for data querying.

Database structure

The SHiNeMaS relational schema is provided as additional file 1. The first part of the schema describes the genetic material. Seed Lot is the main table of the schema while Update Quantity stores stock inventory events. Quantities are managed using gram as the unit. A seed lot is linked to the tables Germplasm and Location. The table Germplasm type manages the type of the genetic material. SHiNeMaS offers a flexible way to manage inbred lines, hybrids, populations or any kind of genetic material you want to describe. The name of the species is also an attribute of the genetic material. A second part of the schema describes the network of participants in the breeding program. The table Person describes the actors of the network, it can be farmers, facilitators, researchers etc. while the table Location describes an experimentation site, an institute, or any legal entity. A third part of the schema represents relations between seed lots with the Relation table. Each relation belongs to an event depicted in the tables Reproduction, Selection, Diffusion and Mixture. A relation is basically a double link to the Seed Lot table with one seed lot considered as a parent and the other one as a child. The quantity (in grams) used for the parent seed lot in a relation is stored in this table. This way to manage relations between seed lots ensures the ability of the tool to design any kind of crosses as long as they are described step by step. For instance, a back cross can be defined this way: $A \times B \rightarrow C$ in a first step, then $C \times B$ to ensure the backcross or a poly-cross can be handled like that: $A \times B \rightarrow C$, $D \times E \rightarrow F$ then $F \times C$ to ensure the poly-cross. Open pollinated event $A \times B \times C \times D \rightarrow E$ can be handled with a mixture.

---

[1] https://www.djangoproject.com/
The last part of the schema represents the data (Raw data table) related to the Relation table or to the SeedLot table. Each data is described according to a variable and a method.

Web application features

Manage data with files

SHiNeMaS provides interfaces to load (figure 2, step 1) relational information between seed lots and the associated data concerning phenotypes, environments and cultural practices with tabulated format files. SHiNeMaS accounts for five kinds of events to describe the type of relationship between seed lots: diffusion, reproduction, cross, mixture and selection resulting in five files format with specific and mandatory headers. A sixth file format describes individual measures. In addition to the mandatory headers, users can add headers corresponding to the variables measured (phenotyping, environment description, cultural practices).

The associated data submitted are linked to a relation, a sown seed lot, a harvested seed lot etc. according to the type of the variable and the type of event. A date can be associated to each measure respecting the ISO 8601[2] standard (YYYY-MM-DD). The format DD/MM/YYYY is also accepted in SHiNeMaS.

After each file submission, a report is displayed to inform the user whether the data has been successfully inserted into the database and, if not, what was wrong with the file. If the submission is valid, SHiNeMaS will create automatically the output seed lot of each relation (figure 2, step 2, for example, the harvested seed lot of a reproduction) with a specific pattern to create the name of the resulting seed lot.

Currently, SHiNeMaS does not provide any native connection to mobile devices. However, the tool we describe here uses tabulated text files to record data. This kind of file can be easily transferred to any mobile device and can be filled with usual editors on this device; In this case offline mode is not a problem as all information needed is stored in the file and accessible in offline mode.

Also, a reproduction file can be used to delete (figure 2, step 5) a set of reproduction events. These events can be deleted alone or in cascade, which means that the event itself, all the subsequent events and all the data are deleted.

Prepare files with wizard tools

Because preparing a large file in a spreadsheet software can be a fastidious task, SHiNeMaS provides a helpful assistant to the creation of files using the seed lots or relations already recorded in the database (figure 2, step 3). By executing successive queries from a simple form, a user can build a list of seed lots or relations and export it in one of the file format described below. Thus, the user will only have to fill the remaining cells (figure 2, step 4) before the next submit.

Data management with single form

SHiNeMaS provides form interfaces to create single relation events (figure 3 shows a form to create a new reproduction event). As for the files, a different form has been designed for each event type. When creating a new single event, the user can also add all the data measured, like when submitting a file. Forms are also provided to manage any other type of information recorded in the database (projects, variables, methods, etc.).

Data querying

SHiNeMaS provides query interfaces to retrieve information about seed lots such as (1) parentage, (2) generation number (on the current location and in total) and the quality of information, (3) data regarding variables measured during a given period of time. The query can be filtered according to various criteria and the results are displayed in three tables showing the data related to seed lots, the data related to the relations and individuals data.

From any query result, the user can access the seed lot card (figure 4). It displays information about the seed lot (its name, stock location, creation year, germplasm etc.), how it was created and used and its stock evolution. The stock is computed using initial quantity of the lot, all quantities used in relations as a parent and the inventories of the stock. The user can also access to the relationship card (figure 5) between two seed lots. It displays all the data that has been measured for this event and the associated seed lots. From any card the data tables can be downloaded in tabulated text file.

Lastly, SHiNeMaS provides a global search bar enabling a quick access to the seed lots, relations or germplasms cards. This search bar is implemented with an auto-complete feature.

Support and community

SHiNeMaS source code is hosted on the French national forge for research and higher education, Source-Sup[3]. Through this forge, our aim is to provide tools to the community such as an issue tracker, forums or wiki site. This forge provides anyone who wants to contribute to the development of our project the usual

[2]https://www.iso.org/iso-8601-date-and-time-format.html

[3]https://sourcesup.renater.fr/
tools of popular forges. Currently, the web interfaces of SHiNeMaS are available in English or French but, with the help of the community, it would be easy to translate the application in any other language. A virtual machine is also available to run SHiNeMaS’ tutorial\(^4\).

Future developments
Major future developments of SHiNeMaS will focus on the management of new data types (climatic data and plant pictures). In addition to storing data on seed lots and relations between seed lots, it will be possible to store data on germplasm. A web service will be developed to provide features ensuring interoperability with external tools such as R packages\(^7\). Efforts will also be made in terms of controlled vocabulary by complying with the Plant Science community of Elixir \(^8\) and we aim to implement methods of the Breeding API\(^6\). The version 2.0 of this API defines new methods related to seed lot and seed lot transactions which would be relevant for our tool. Moreover, we will ensure an audit of submitted files to keep track of the submissions. In a further version of SHiNeMaS, we will focus on the use of controlled vocabulary and ensure the interoperability with standard ontologies. Using the same user role for database administration and data management can be a risk if the data manager is not aware of all these privileges. To fill this gap we will improve the data access control and define more user roles in a further SHiNeMaS release.

Conclusion
We developed SHiNeMaS to ensure seed lots traceability and to manage phenotypes, environments and cultural practices data for various genetic resources, varieties of many crop species. This web application is a key feature as it enables data administration at different scales (files, or single forms). SHiNeMaS is especially adapted for field experimentation in a network of farms or experimental stations and that requires tracking seed lots history. It is available under AGPL licence and uses free technologies. SHiNeMaS has been used in production in our laboratory for 5 years and some farmers’ organizations facilitate manage their own information in the system.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author’s contributions
YDO, LB and DM produces the code source of the software. PR, MT, GVF, IG, JD and DS contributed to the conception of the tool. They also contributed to the writing of the user documentation, translation, and tutorial.

Acknowledgements
The authors thank Eva Dechaux, Elodie Gras, Johann Joets, Sophie Pin and Mélanie Polart-Donat for testing, developing or advising along this work. The authors thank Marissa Nix for reviewing this manuscript.

Availability
- Project name: SHiNeMaS
- Project home page: https://sourcesup.renater.fr/projects/shinemas
- Operating system(s): Platform independent
- Programming language: Python
- Other requirements: Python3, Django 2.0
- License: Affero GPL
- Any restrictions to use by non-academics: None

Funding
This work has been supported by INRAE. It has been funded by (1) the European Community’s Seventh Framework Program (grant agreement No 245058, Solibam), (2) the European Union’s Horizon 2020 research and innovation program (grant agreement No 633571, DIVERSIFOOD), (3) the Fondation de France, the French Region Centre (SPEAL project), the French Ministry of Agriculture and Food (ProABioDiv and Coevolution CASDAR projects).

Author details
1 Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQF - Le Moulon, 91190 Gif-sur-Yvette, France. 2 AgroBioPerigord, 24000 Périgueux, France. 3 Department of Horticulture, University of Wisconsin–Madison, Madison, Wisconsin, 53706, USA. 4 Réseau Semences Paysannes, 47190 Aiguillon, France. 5 CIRAD, UMR AGAP, F-34398 Montpellier, France. 6 Univ Montpellier, CIRAD, INRAE, Montpellier SupAgro, Montpellier, France.

References
1. Wolfe, M.S., Baresel, J.P., Desclaux, D., Goldringer, I., Hoad, S., Kovacs, G., Löschbenberger, F., Miedaner, T., Østergård, H., Lammers van Bueren, E.T.: Developments in breeding cereals for organic agriculture. Euphytica 163(3), 323 (2008). doi:10.1007/s10681-008-9960-9
2. Bonnin, I., Bonneuil, C., Goffaux, R., Montalent, P., Goldringer, I.: Explaining the decrease in the genetic diversity of wheat in France over the 20th century. Agriculture, Ecosystems & Environment 195, 183–192 (2014). doi:10.1016/j.agee.2014.06.003
3. Thomas, M., Demoulen, E., Dawson, J.C., Khan, A.R., Galic, N., Jouanne-Pin, S., Remoue, C., Bonneuil, C., Goldringer, I.: On-farm dynamic management of genetic diversity: the impact of seed diffusions and seed saving practices on a population-variety of bread wheat. Evolutionary Applications 5(8), 779–795 (2012). doi:10.1111/j.1752-4571.2012.00257.x. Accessed 2013-02-27
4. Dawson, J., Rivière, P., Berthelot, J.-F., Mercier, F., de Kochko, P., Galic, N., Pin, S., Serplay, E., Thomas, M., Giuliano, S., Goldringer, I.: Collaborative plant breeding for organic agricultural systems in developed countries. Sustainability 3(8), 1206–1223 (2011)
5. Rivière, P., Goldringer, I., Berthelot, J.-F., Galic, N., Pin, S., De Kochko, P., Dawson, J.: Response to farmer mass selection in early generation progeny of bread wheat landrace crosses. Renewable Agriculture and Food Systems 30(2), 190–201 (2015)
6. Rivière, P., Dawson, J.C., Goldringer, I., David, O.: Hierarchical bayesian modeling for flexible experiments in decentralized participatory plant breeding. Crop Science 55(3), 1053–1067 (2015)
7. Rivière, P., Frank, G.V., David, O., Mulioz, F.: PPBstats : An R package for Participatory Plant Breeding statistical analyses. https://github.com/prieviere/PPBstats
8. Goldringer, I., van Frank, G., Bouvier d’Yvoire, C., Forst, E., Galic, N., Garnault, M., Locquille, J., Pin, S., Bailly, J., Baltassat, R., Berthelot, J.-F., Caizergues, F., Dalmasso, C., de Kochko, P., Gascuel, J.-S., Hyacinthe, A., Lacanette, J., Mercier, F., Montaz, H.,...
Ronot, B., Rivière, P.: Agronomic evaluation of bread wheat varieties from participatory breeding: A combination of performance and robustness. Sustainability 12(1) (2019). doi:10.3390/su12010128
9. David, O., van Frank, G., Goldringer, I., Rivière, P., Delof, M.T.: Bayesian inference of natural selection from spatiotemporal phenotypic data. Theoretical Population Biology 131, 100–109 (2020). doi:10.1016/j.tpb.2019.11.007
10. Exner, V., Hirsch-Hoffmann, M., Gruissem, W., Hennig, L.: Plantdb – a versatile database for managing plant research. Plant Methods 4(1), 1 (2008). doi:10.1186/1746-4811-4-1
11. Nieuwland, J., Sonnay, E., Marchbank, A., de Graaf, B.H., Murray, J.A.: Phytotracker, an information management system for easy recording and tracking of plants, seeds and plasmids. Plant Methods 8(1), 43 (2012). doi:10.1186/1746-4811-8-43
12. Charavay, C., Segard, S., Pochon, N., Nussaume, L., Javot, H.: Seedusoon: A new software program to improve seed stock management and plant line exchanges between research laboratories. Frontiers in Plant Science 8, 13 (2017). doi:10.3389/fpls.2017.00013
13. Steinbach, D., Alaux, M., Amelem, J., Choixne, N., Durand, S., Flores, R., Kellet, A.O., Kimmel, E., Lapalu, N., Luyten, I., Michotey, C., Mohelibi, N., Pommier, C., Reboux, S., Valdenaire, D., Verdelet, D., Queeneville, H.: Gnips: an information system to integrate genetic and genomic data from plants and fungi. Database 2013, 058 (2013). doi:10.1093/database/bat058
14. The Integrated Breeding Platform: The IBP Breeding Management System Version 3.0.9. https://www.integratedbreeding.net/breeding-management-system
15. Stevens, T.J., Boucher, W.: Python Programming for Biology: Bioinformatics and Beyond. Cambridge University Press, ??? (2015). doi:10.1017/CBO9780511843556

Figures

**Figure 1 Seed lots network.**
A seed lot is represented by a colored plain circle. A different color means a different germplasm. A relation is depicted by an arrow whose color indicates the type of relation.

**Figure 2 Data flow in SHiNeMaS.**
SHiNeMaS has been designed to work efficiently with files. When submitting a file (1) the application will create the resulting seed lots for each event (2). These seed lots can be used later to create and export a new file (3). Thus, the user only need to fill the remaining cells in a spreadsheet software (4) before a new submission. A user can also export a file with recorded relations and use it to run a deletion process (5).

**Figure 3 Reproduction creation form.**
Form to create a new reproduction. The form is divided in three fieldsets. The first one requires sovn seed lot information. The second requires to fill the reproduction information. The last one contains the data related to the reproduction.

**Figure 4 Seed lot card.**
The seed lot card is divided in four frames:
- The seed lot information with its name, species, germplasm etc.
- The history of the seed lot, how it has been created and how it has been used
- The data linked to the seed lot
- The stock level and change history

**Figure 5 Relation card.**
The relation card is divided in two frames:
- The information about this relation
- The data related to this relation at global and individual level

Additional Files
Additional file 1 — Database schema
Relational schema of SHiNeMaS' database.
Seed lots network. A seed lot is represented by a colored plain circle. A different color means a different germplasm. A relation is depicted by an arrow whose color indicates the type of relation.
Data flow in SHiNeMaS

Data ow in SHiNeMaS. SHiNeMaS has been designed to work efciently with les. When submitting a le (1) the application will create the resulting seed lots for each event (2). These seed lots can be used later to create and export a new le (3). Thus, the user only need to ll the remaining cells in a spreadsheet software (4) before a new submission. A user can also export a le with recorded relations and use it to run a deletion process (5).
Figure 3

Reproduction creation form. Form to create a new reproduction. The form is divided in three fieldsets. The first one requires sown seed lot information. The second requires to fill the reproduction information. The last one contains the data related to the reproduction.
**Seed lot profil : RxY PEM 2015 0001**

Seed lot informations

Name: RxY PEM 2015 0001
Species: Bread-wheat
Germplasm: RxY
Owner: PEM
Creation year: 2015
Projects involved in: PPB

Seed lot history

Creation history:
- RxY PEM 2014 0001 → RxY PEM 2015 0001 in a Reproduction event

Use history:
- RxY PEM 2015 0001 → RxY PEM 2016 0001
- RxY PEM 2015 0001 → RxY STS 2015 0001
- RxY PEM 2015 0001 → RxY JOD 2015 0001

Seed lot data

No data for this seed lot

Stock information

Seed lot still available? Yes
Remaining quantity: 688.0 g

Stock evolution:
Initial quantity: 888 g

---

**Figure 4**

Seed lot card. The seed lot card is divided in four frames: - The seed lot information with its name, species, germplasm etc. - The history of the seed lot, how it has been created and how it has been used - The data linked to the seed lot - The stock level and change history
Figure 5

Relation card. The relation card is divided in two frames: - The information about this relation - The data related to this relation at global and individual level

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.jpg