Achieving better outcomes from seed bank Data Management systems

ANNE FUCHS[1,3]*, LYDIA GUJA[2,3], TOM NORTH[2] AND AMELIA STEVENS[2,3]

[1] Biodiversity Informatics, Australian National Botanic Gardens
[2] National Seed Bank, Australian National Botanic Gardens
[3] Centre for Australian National Biodiversity Research
*Corresponding author: Anne.Fuchs@awe.gov.au

Introduction

There are many key activities that contribute to the successful management of a seed or gene bank. One such activity is data management. In undertaking research and collection management activities data are generated. The analysis and synthesis of this data help answer the questions such as how to achieve successful banking of seed, seed germination or alleviating dormancy. The data can also provide information about species characteristics to help understand their climate adaptability, or to inform ecosystem restoration.

This data encapsulates the knowledge and skills of the researchers, curators and technicians who contribute to it. To maximise the impact, data itself must follow the FAIR principles – be Findable, Accessible, Interoperable and Reusable. Furthermore, in a resource constrained environment organisations need to leverage skills in a collaborative way to avoid duplication of effort, provide timely outputs and maximise skills. The National Seed Bank (NSB) systems have been developed with these principles and objectives in mind. This article will present the approach and methods used to deliver better outcomes for the internal business systems and integration to national and international databases.

Approach

Leading up to 2012 a project was undertaken at the Australian National Botanic Gardens (ANBG), Canberra to bring together siloed systems that served the herbarium, seed bank and living plant collections. Work was undertaken to analyse and prepare a data model and an associated application which would combine these datasets and meet a number of principles, these being:

1. Enter data once and share – common data about the collection such as what do we have (stock), where is it (inventory), where was it collected and by whom (provenance) and what is its scientific name (identification) would be shared between herbarium, living and seed collections. This principle is also embedded in the objectives of the Centre for Australian National Biodiversity Research (CANBR) collaborative agreement between the Director of the National Parks (DNP) and CSIRO.
2. Support data and processes for specific areas – for the seed bank this included germination trials, seed pre-processing, characteristics such as thousand seed weight, or storage conditions. The model had to be adaptive with the ability to adopt innovations and additional data into the future.
3. Support analysis and decision making.
4. Adhere to FAIR principles – the data must be Findable, Accessible, Interoperable, and Reusable.
5. Adhere to international exchange standards such as Darwin Core to facilitate exchange of data between institutions such as the ANBG and the Millennium Seed Bank, Kew, and to aggregators such as the Atlas of Living Australia (ALA).
6. Adopt common patterns to reduce the IT overhead of introducing new methods and processes where possible (data and configuration over programming).

Implementation

This analysis led to the data model Figure 1 which became the basis of the database built to support the combined herbarium, living collections and seed bank. Much of the model which handles provenance,

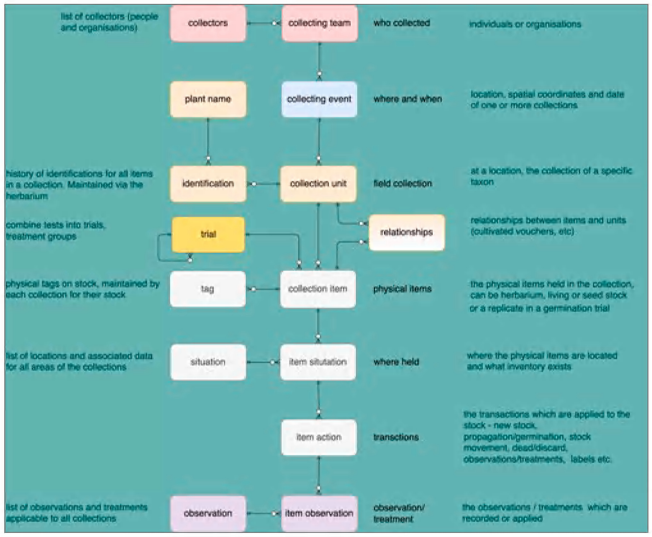


Figure 1. Data Model.

Basic data structure for generic management of observations/treatments

OBS_CATEGORY	OBS_GROUP	OBS_TYPE	OBS_CODE	MEASUREMENT_1	UNIT_OF_QTY_M1	MEASUREMENT_2	UNIT_OF_QTY_M2	DESCRIPTION
Treatment	Propagation	Substrate	AGAR%		%			AGAR concentration
Treatment	Propagation	Substrate	Soil					Soil mix used for seed trial
Treatment	Propagation	Substrate	Filter paper					Filter paper.
Treatment	Propagation	Substrate	Towelling					Towelling.
Treatment	Propagation	Substrate	Moss					Moss
Treatment	Propagation	Substrate	AGAR+KN04					AGAR base with potassium nitrate added.
Treatment	Propagation	Substrate	Filter/towel					Filter/towel.
Treatment	Propagation	Substrate	AGAR+GA3	200	ppm			AGAR base with gibberellic acid added.
Treatment	Propagation	Chemical	H2SO4%		%		hours	Soaked in Sulfuric acid (H2SO4) at concentration for a period
Treatment	Propagation	Chemical	GA3sub		ppm			Gibberellic Acid applied to substrate at a concentration.
Treatment	Propagation	Chemical	KN04soak		ppm		hours	Soaked in potassium nitrate at a concentration for a period.
Treatment	Propagation	Chemical	GA3soak		ppm		hours	Soaked in gibberellic acid at a concentration for a period.
Treatment	Propagation	Chemical	KN04sub		ppm		hours	Potassium nitrate applied to the substrate at a concentration.
Observation	Seed Sample	Purity	PctPurity		%			% Purity.
Observation	Seed Sample	SW	SW		gm			Sample weight of seed (basis of TSW-m2 calculation)
Observation	Seed Sample	TSW	TSW-m1		gm			Thousand seed weight calculated from one sample.
Observation	Seed Sample	TSW	TSW-m2		gm			Thousand seed weight calculated from an average of 5 samples.
Observation	Seed morphology	Appendage	seed appendage					Seed appendages
Observation	Seed morphology	Length	seed length		?		?	Calculated length of seed (average, standard deviation) in microns
Observation	Seed morphology	LengthSample	seed sample(length)		ea			Number of seed used in length calculation

User entered values

Figure 2. Observation and Treatment examples.

occurrence and identification data was already well established and are able to be mapped to the Darwin Core biodiversity standard.

However, seed banks have several additional requirements which required an extension to these core data. Data associated with activities such as germination trials, testing results, pre-processing, standard seed bank metrics such as thousand seed weight and tracking of storage conditions lead to a design in which a generic model based on observations and treatments was adopted. An observation is something which is 'seen' and a treatment is something that is 'done'. These terms are a convenience to help categorise different codes assigned for data entry. Examples of this approach are presented in Figure 2. Overtime the list of codes has been extended to handle new requirements such as recording the morphological characteristics of seed derived from microscope images, the inclusion of a new x-ray test for seed fill and categorising seed storage behaviour (orthodox or otherwise). This ability to extend the application for new data with minimal programming is important for adoption of new methods in a timely way without extensive outlay of information technology resources.

In the seed bank application, to streamline data entry, reduce errors and support reporting requirements, the basic reference table for observation/treatment allows for two additional levels of categorisation, a unique code and two pairs of values (value, unit of qty) which can be associated with frequently used characteristics or test conditions. A more generalised model would provide ultimate flexibility however based on the analysed reporting requirements this level of categorisation has worked for our application.

The other feature of the data model and application is that it uses a set of inventory management tasks to track the stock quantities and location through 'actions' such as accessioning of new collections, recording the location and movement within the collection, how it leaves the collection – as a donation to another institution, deaccessioned, or consumed in germination trials.

This is coupled with 'the story' of the seed through its scientific journey – the observations and treatments made upon it – how it is pre-processed for storage, images are taken and characteristics measured, x-rays are taken to measure seed fill, seed is germinated in trials and the steps are recorded – the pre 'treatments', the trial conditions (day/night photoperiod, substrate), germination results, any re-treatments required and the

Figure 3. Data entry of inventory and science actions.

final cut tests. This mixture of inventory management and observation provides a flexible data set to meet reporting and analysis requirements. Examples of how inventory and science data is recorded is presented in Figure 3.

Supporting Seed Science

How do these data support seed science? By recording primary data, as opposed to summary results, and recording details relating to the data's origin (who created the record, when and through what action), this model provides the ability to report, collate, analyse, and export the data to support the activities of the seed bank both for management of the collection and contributing to scientific knowledge and sharing data.

Synthesis of this data allows a summary to be presented in real-time in the data entry system providing a quick overview of the data for a seed collection. The summary presents characteristics of the collection such as purity and seed fill, a snapshot of the seed's morphology and available microscope images, and germination statistics for each test that has been undertaken to quickly review germination results (Figure 4).

In addition to the seed science data, the inventory management aspect of the system and any related data from the living collection and herbarium can be presented in a consolidated view (Figure 5) which puts all related data at the user's fingertips.

To supplement presentation in the data entry system, standard outputs and exports from the data are provided. Scripts have been written to generate required reports or views of the seed bank data. Each script can be run for various parameters which allows the data to be reported pertinent to the questions *e.g.,* which seed collections had a germination test in a particular year and what was the result?, what are the results of germination tests undertaken for *Eucalyptus rossii*? The output of each script can be downloaded to Excel for further work. These statistics are generated on the base data which allows, over time, for the data to be analysed in different ways.

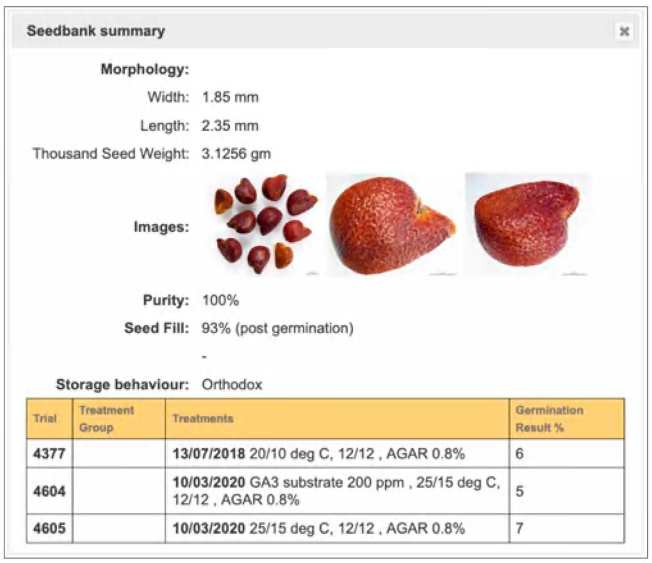
An overview of the analysis, exports and potential usages of some of the seed bank data is presented in Figure 6.

Specific examples of reports currently generated from the system include:

Analysis

Germination trial outcomes

Provides a summary of all results for a germination trial with data on number of replicates, total number of germinates; number of empty, mouldy, full seeds after cut-test; % seed fill, final %germination, %viability adjusted germination, days to first germination and total number of days in the test. This can be requested based on a number of criteria such as a taxon name to compare tests for different collections of the same taxon.



Trial	Treatment Group	Treatments	Germination Result %
4377		**13/07/2018** 20/10 deg C, 12/12 , AGAR 0.8%	6
4604		**10/03/2020** GA3 substrate 200 ppm , 25/15 deg C, 12/12 , AGAR 0.8%	5
4605		**10/03/2020** 25/15 deg C, 12/12 , AGAR 0.8%	7

Figure 4. Seed Collection Summary.

Figure 5. Consolidated view of shared data services.

Testing and viability schedules

Collate data to support the scheduling of seed testing programs – when was the seed originally collected?, what germination testing has occurred?, has any previous testing resulted in a 'successful' germination test (>75%)? The data collected from all previous tests and general provenance data are used to report the current status and help inform decision making. This is supplemented with analysis of the storage conditions of the seed – how long has it been stored under different conditions, age of the seed and linked to previous germination testing results.

Data gap analysis

Assess the seed collection to identify those collections which do not have all the required seed bank metrics. This is done by reviewing whether germination testing has been conducted, that the seed has been weighed and a thousand seed weight has been recorded. This allows work to be scheduled within seed bank resources to ensure minimum levels of collection and data quality.

Annual reporting

Collate statistics for annual reports such as numbers of new seed collections, and through matching the seed bank data to other datasets such as the Species Rare and

Threatened (SPRAT) provide the number of collections of EPBC priority taxa and overall percentage of EPBC priority taxa represented in the collection.

Data Exchange

To support the exchange of data to other data aggregators, institutions or researchers the FAIR principles must be met. A large part of ensuring this principle is adoption of existing biodiversity standards such as those provided by the Biodiversity Information Standards (TDWG). In the biodiversity community standards for occurrence data are well articulated, such as Darwin Core, and form the basis for delivery of occurrence and related data to products such as the Australian Virtual Herbarium (AVH).

However definitions and standards for the exchange of seed bank characteristics are a work in progress and are being addressed via workshops such as the Seed Trait Workshop, Perth, 2016 (Saatkamp *et al*. 2019), in the proposed Seed Germination Database (pers. comm. 2022) and work by ALA and Global Biodiversity Information Facility (GBIF) piloting changes to the current Darwin Core and extensions (pers. comm. 2022). The implication of this is that the standards for the exchange of trait data are a work in progress.

The approach taken to date is to adopt a standard where it is available and where not, adhere as much as possible to an existing standard but with a focus of delivering the data required for shared seed banking and research. As biodiversity standards for this data mature they will be adopted.

Australian Seed Bank Partnership (ASBP)

The ASBP mission is a national effort to conserve Australia's native plant diversity through collaborative and sustainable seed collecting, banking, research and knowledge sharing. The NSB support this mission in multiple ways including through the provision of data to the Australian Seed Bank online portal hosted at the Atlas of Living Australia (ALA).

On the inception of the ASBP project in 2011 one of the first tasks was to prepare a data exchange standard (ASBP Shared Data Definitions v0.2) which meet the above principles but also considered various levels of raw and summarised seed bank data available in different institutions. The NSB has provided data, using these standards, to the ASBP on an ad-hoc basis since December 2012.

AusTraits

AusTraits is an open-source, harmonised database of Australian plant trait data and integrates plant trait data collected by researchers from diverse disciplines, including functional plant biology, plant physiology, plant taxonomy, and conservation biology (Falster *et al*. 2021). The NSB contributes data to this project using the same data exchange files as provided to ASBP, providing reusability and common data to both aggregators.

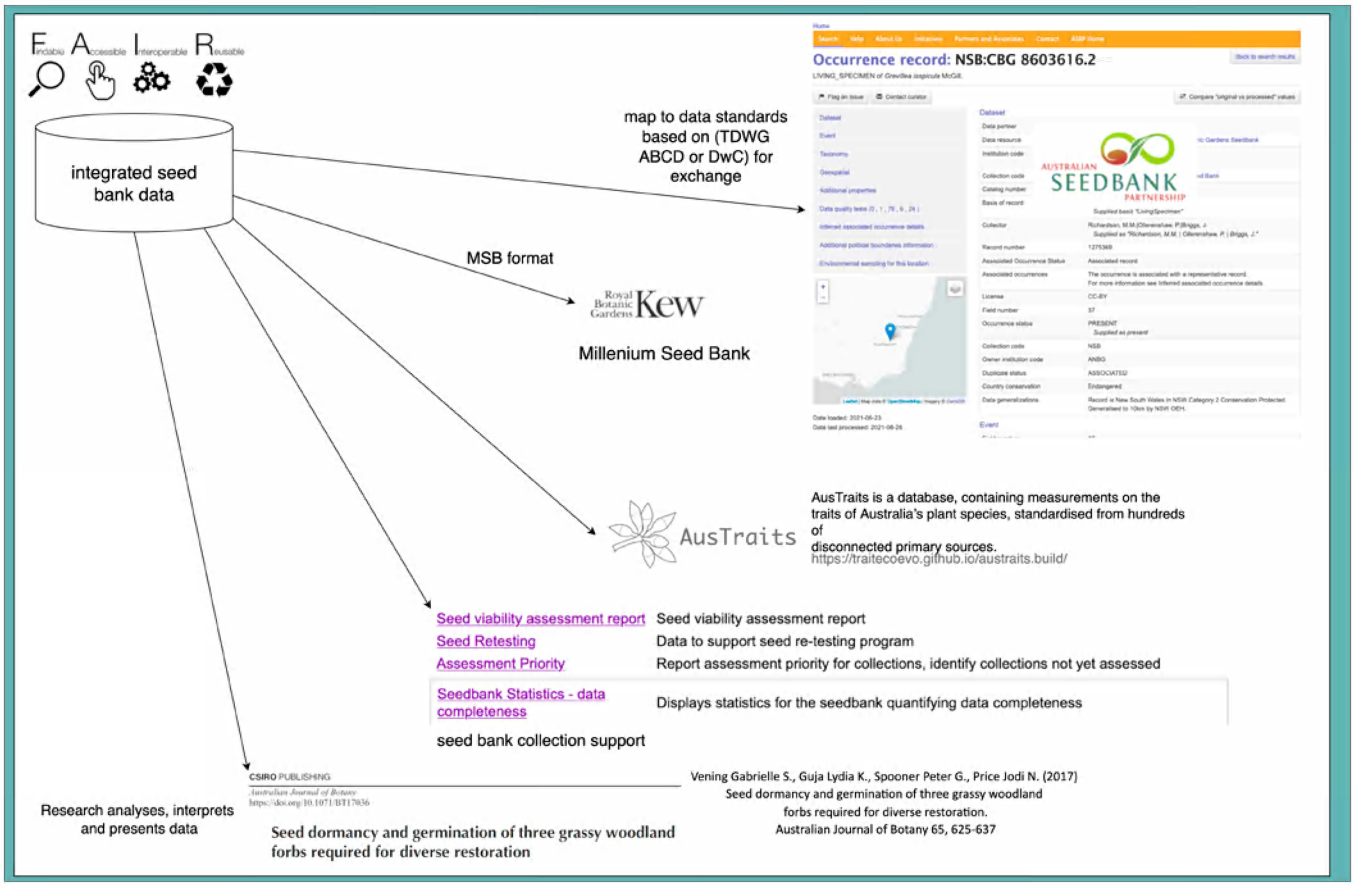


Figure 6. Analysis, exchange and FAIR use of data.

Conclusion

The adoption of robust data structures that have proven adaptable to new requirements, with minimal programming effort, have allowed the seed bank to quickly implement new methods and procedures and capture the data electronically. This has allowed standardised reports to be created to support management of the collection and the sharing of seed science data. From the standard data inputs information can be provided to service different requirements at different scales, for example informing the seed bank team's curation and research work, the ANBG's living plant collection maintenance, the national picture of conservation collections or plant traits and the global understanding of seeds and biodiversity.

References

Falster, D., Gallagher, R., Wenk, E.H. *et al.* (2021) AusTraits, a curated plant trait database for the Australian flora. *Sci Data* **8:** 254. https://doi.org/10.1038/s41597-021-01006-6

Saatkamp, A., Cochrane, A., Commander, L., Guja, L., Jimenez-Alfaro, B., Larson, J., Nicotra, A., Poschlod, P., Silveira, F.A.O., Cross, A., Dalziell, E.L., Dickie, J., Erickson, T.E., Fidelis, A., Fuchs, A., Golos, P.J., Hope, M., Lewandrowski, W., Merritt, D.J., Miller, B.P., Miller, R., Offord, C.A., Ooi, M.K.J., Satyanti, A., Sommerville, K.D., Tangney, R., Tomlinson, S., Turner, S. and Walck, J.L. (2019). A research agenda for seed-trait functional ecology. N*ew Phytol*, 221: 1764-1775. https://doi.org/10.1111/nph.15502

Websites

Atlas of Living Australia https://ala.org.au

Australian Seed Bank Partnership https://www.seedpartnership.org.au

Australian Seed Bank online https://asbp.ala.org.au/

Australian Virtual Herbarium https://avh.chah.org.au/

AusTraits https://austraits.org/

Global Biodiversity Information Facility (GBIF) https://gbif.org

TDWG https://www.tdwg.org

Darwin Core https://www.tdwg.org/standards/dwc/

News from the ASBP

Extract from the Partnership's Annual Report 2021–22

BRADLEY DESMOND[1]*, DAMIAN WRIGLEY[1], MATHEW STRAY[2], ANDREW CRAWFORD[3], DAN DUVAL[4], ANDRE MESSINA[5], JAMES WOOD[6], TOM NORTH[7], PETER CUNEO[8], JASON HALFORD[9] AND AMELIA MARTYN YENSON[10]

[1] Australian Seed Bank Partnership
[2] Western Australian Seed Centre, Kings Park
[3] Western Australian Seed Centre, Kensington
[4] South Australian Seed Conservation Centre, Botanic Gardens and State Herbarium
[5] The Victorian Conservation Seedbank, Royal Botanic Gardens Victoria
[6] Tasmanian Seed Conservation Centre, Royal Tasmanian Botanical Gardens
[7] National Seed Bank, Australian National Botanic Garden
[8] The Australian PlantBank, The Royal Botanic Gardens and Domain Trust
[9] The Queensland Herbarium, Queensland Department of Environment and Science
[10] Australian Network for Plant Conservation
*Corresponding author: coordinator@seedpartnership.org.au

The Australian Seed Bank Partnership are pleased to share highlights from over the last year.

The Western Australian Seed Centre, Kings Park

Over the course of 2021–22, field collection activities included trips to Western Australia's Mid-West Region, Stirling Range and Kimberley as well as various day trips made in the areas surrounding Perth. With 45 days spent in the field, 186 collections of 122 species were made, most of which have now undergone germination testing and are safely stored in the seed bank as conservation collections. These are also now available for use in research activities. Collections were made across a broad cross-section of the species' ranges to gather a suitable representation of the genetic diversity within a species, and future collecting activities will focus on increasing the number of populations from which the seeds are sourced.

With the benefit of additional funding made available by the Australian Seed Bank Partnership through the UK Government Emergency Bushfire Fund, Kings Park was able to send a team of four staff to collect within bushfire-prone areas of Western Australia's Stirling