Research Ideas and Outcomes 10: e135978
doi: 10.3897/rio.10.e135978

Project Report

The key to bringing DNA collections to the next level

Emily Veltjen[‡], Pieter Asselman[§], Wim Baert[|], Steve Baeyen[¶], Lise Beirinckx[#], Liselot Breyne[¶], Dimitri Brosens[‡], Tim Claerhout[§], Sari Cogneau[¤], Karen Cox[‡], Laura Cuypers[«,»], Lynn Delgat[^], Philippe Desmeth[˅], Jordi de Raad[¦], Lore Esselens[ʔ], Maria-Rose Eves Down[‡], Philippe Helsen[¦], Frederik Leliaert[|], Kenny Meganck[ʔ], Zjef Pereboom[¦], Nathalie Smitz[ʔ], Gontran Sonet[»], Maarten Trekels[|], An Vanden Broeck[‡], Charlotte Van Driessche[‡], Aaike De Wever[‡]

‡ Research Institute for Nature and Forest (INBO), Brussels, Belgium
§ Ghent University, Ghent, Belgium
| Meise Botanic Garden, Meise, Belgium
¶ Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Ostend/Merelbeke, Belgium
# Vrije Universiteit Brussel, Brussels, Belgium
¤ Belgian Coordinated Collections of Microorganisms / Institute of Tropical Medicine - Mycobacteriology Unit, Antwerp, Belgium
« University of Antwerp, Antwerp, Belgium
» Royal Belgian Institute of Natural Sciences, Brussels, Belgium
^ Flanders Marine Institute, Ostend, Belgium
˅ Belgian Coordinated Collections of Microorganisms / Belgian Science Policy Office, Brussels, Belgium
¦ Royal Zoological Society of Antwerp, Antwerp, Belgium
ʔ Royal Museum for Central Africa, Tervuren, Belgium

Corresponding author: Emily Veltjen (emily.veltjen@inbo.be)

Reviewable v 1

Received: 30 Aug 2024 | Published: 08 Nov 2024

Citation: Veltjen E, Asselman P, Baert W, Baeyen S, Beirinckx L, Breyne L, Brosens D, Claerhout T, Cogneau S, Cox K, Cuypers L, Delgat L, Desmeth P, de Raad J, Esselens L, Eves Down M-R, Helsen P, Leliaert F, Meganck K, Pereboom Z, Smitz N, Sonet G, Trekels M, Vanden Broeck A, Van Driessche C, De Wever A (2024) The key to bringing DNA collections to the next level. Research Ideas and Outcomes 10: e135978. https://doi.org/10.3897/rio.10.e135978

Abstract

DNA collections are a valuable type of Natural Science collection, enabling the validation of past research, serving as a source for new genomic studies and supporting *ex situ* conservation. The DiSSCo Flanders DNA collection working group, aiming to advance and "unlock" their DNA collections, identified the need for: 1) actively sharing best practices regarding the management of DNA collections; and 2) providing guidance on how to bring theory into practice. By combining best practice examples from within the

working group with available literature and brainstorming ideas, the working group co-created two outputs, referred to as: the "Challenges" and the "Key". The Challenges are a list of obstacles to DNA collection management, which shape the structure of the linked Key and can also be used to spark discussion amongst stakeholders. The Key is a tool that guides users through the maturation process of their DNA collection in a standardised way. It stimulates holistic growth, breaks down the needed work into manageable steps and helps to decide priorities during the process. Furthermore, the Key facilitates communication with both internal stakeholders and external DNA collection managers. The Key distinguishes itself from other self-assessment tools in several ways: it includes (re)investigation of the collection's purpose and context; it is specialised for DNA collections; it delivers concrete goals linked to relevant information and shared experience; and it is inclusive, targeting all Natural Science DNA collections, regardless of their context or size.

Keywords

DNA specimens, Natural Science collections, DiSSCo, molecular collections, biodiversity biobanks, FAIR principles

Introduction

When referring to Natural Science collections (Addink et al. 2020), one might think of a myriad of pinned butterflies in a museum drawer, a huge reconstructed skeleton of a Brachiosaurus, a sparkling Lapis Lazuli or a beautifully dried and mounted herbarium specimen, to name but a few. A much less conspicuous and relatively young type of Natural Science collection is a DNA collection, preserving DNA extracted from (non-human) organisms or environmental specimens such as water, soil or air. A common synonym is "DNA bank(s)" (de Vicente and Andersson 2006, Corrales et al. 2023). A DNA collection can be categorised under the overarching concepts of a "biological or environmental repository" (ISBER 2023), a "biodiversity biobank" (Corrales and Astrin 2023) or a "molecular collection" (de Mestier et al. 2023). Natural Science DNA collections are typically associated with genetic laboratories within research institutes. The DNA specimens themselves are either extracted from museum specimens or actively gathered during sampling events by collecting tissue, organisms or environmental samples to extract the DNA from.

Natural Science DNA collections are considered valuable because:

- DNA specimens often contain much unexplored genetic information.
- DNA specimens often have enough volume left after their original research purpose for additional analyses.
- DNA specimens contain DNA molecules that are stable information carriers.

- DNA specimens avoid repeated destructive sampling of the source material.
- DNA specimens can be used to carry out alternative versions of genetic and genomic studies (e.g. molecular taxonomy, conservation genetic research, population genetics and phylogenetics) by using different techniques (e.g. other molecular markers, other sequencing techniques) than were used in the original study.
- DNA specimens allow research validation (i.e. repetition using the same techniques as the original study).
- DNA specimens allow investigation into completely new research questions (e.g. “museum genomics” (Card et al. 2021) and “museomics” (Raxworthy and Smith 2021).
- DNA specimens are extremely valuable when it comes to generating barcoding reference libraries (i.e. molecular vouchers (Astrin et al. 2013)).
- DNA specimens are crucial in the context of rapidly evolving research fields, such as eDNA (environmental DNA) biomonitoring (Jarman et al. 2018), metabarcoding and metagenomics (Ryan et al. 2021, Lazareva et al. 2022).
- DNA specimens allow storage of genetic diversity information during a global biodiversity crisis as a potential source of *ex-situ* conservation (Corrales and Astrin 2023).

Of course, the value of each individual DNA specimen depends on its specific (data) context and quality (Veltjen et al. in preparation).

Although all types of Natural Science collections have their own associated management challenges, biodiversity biobanks, including DNA collections, have received extra attention in recent years. This is evident when looking at recent publications, such as the “Policies Handbook on Using Molecular Collections” (de Mestier et al. 2023), “Biodiversity biobanking - a Handbook on Protocols and Practices” (Corrales and Astrin 2023) and “Biodiversity biobanks: a landscape analyses” (Corrales et al. 2023). Similarly, there are specialised communities that focus on DNA collection and data management, such as the "International Society for Biological and Environmental Repositories (ISBER)", publishing the “ISBER Best Practices: Recommendations for Repositories” (ISBER 2023), the Global Genome Biodiversity Network (GGBN, created in 2013) and the Biodiversity Genomics Data Hub (Forsdick et al. 2023). These developments are largely expected for the three main reasons, discussed in the next paragraphs.

Firstly, given the context in which DNA specimens are generated:

- Genetic research is relatively young compared to other research disciplines, such as morphological studies (Durmaz et al. 2015).

- Genetic research is becoming more accessible and less costly (Ferreira et al. 2018).
- Identification by DNA sequencing is being increasingly used: the methods for DNA-based taxonomy and identification can largely be applied across taxonomic groups (however, see Keck et al. (2022)).
- Sequencing techniques continue to evolve quickly (Satam et al. 2023).
- Various high-profile agreements, such as the Convention on Biological Diversity (CBD), Access and Benefit Sharing (ABS) and the Nagoya protocol have attracted more attention to, and regulations for, the utilisation of genetic resources.
- DNA collections are a particularly challenging type of Natural Science collection, as their specimens are often linked to a wide variety of related data, such as:
    - The initial source sample from which the DNA was obtained, which can subsequently be linked to a specimen voucher (e.g. herbarium specimens, see Funk et al. (2017)).
    - The resulting DNA sequence data (e.g. deposited on ENA (Cummins et al. 2021)) and genetic/genomic datasets.
    - A match to a molecular voucher (and linked specimen voucher) from a reference database (Astrin et al. 2013).

Secondly, some DNA collections have come into being in an "ad-hoc" manner as a by-product of a genetic research laboratory, rather than being carefully planned. Sometimes, the DNA collections themselves have only recently been identified and treated as Natural Science collections.

Thirdly, the communities of researchers and laboratory technicians that manage DNA collections (especially those not linked to Natural Science museums, such as universities or governmental institutes), are often relatively new to, or not (yet) involved in the larger Natural Science collection community or the biobanking community. They may consequently miss certain useful guidance and knowledge when managing DNA collections.

The members of the DiSSCo Flanders consortium, with the common goal of maturing their DNA collections to be included in DiSSCo Europe, identified: 1) a need for actively sharing best practices regarding the management of (Natural Science) DNA collections and 2) a need for guidance on how to bring management theory into practice, given that each DNA collection varies in maturity, context and size. The context of DNA collections varies in terms of diverse factors, such as the available means (time, money, manpower), the geographical scope of the specimens, the organisation of the genetic laboratory associated with the DNA collection and the number of people (e.g. scientists, students, technicians) associated with the DNA collections that generate specimens for deposit. The consortium's Collection Self-Assessment Tool (CSAT) scores on the benchmark

statements of the molecular collections (CSAT taken around January 2023), supported this need and showed that many of the best practices circulating at the time of the assessment were not (yet) put into practice.

Using the shared experience of the consortium members and the available literature linked to DNA collection management, the consortium aimed to co-create a resource that addresses the question: "**How can a DNA collection be matured as efficiently as possible, taking into account the collections' diversity in context and size?".**

Material and Methods

During the DiSSCo Flanders project, the DiSSCo Flanders DNA collection working group was created. The working group is open to all Natural Science DNA collection associates in Belgium: researchers, lab technicians, collection managers, data managers etc. Around 50 people were brought together, with 13 different affiliations: Meise Botanic Garden, the Flanders Research Institute for Agriculture, Fisheries and Food, the Research Institute for Nature and Forest, the Belgian Coordinated Collections of Microorganisms, the Institute of Tropical Medicine, the Royal Zoological Society of Antwerp, KU Leuven, the Royal Belgian Institute of Natural Sciences, the University of Antwerp, the Royal Museum for Central Africa, Ghent University, the Flanders Marine Institute and the Vrije Universiteit Brussel. Most of the organizations represented here do not have a DNA-collection-only biodiversity biobank: tissues or (micro-)organisms are often stored in parallel. The various DNA collections are handled in a variety of contexts: universities, governmental and museum institutes; organised in both decentralised or centralised manners; using cold storage or room temperature storage; managed by an appointed curator or a dedicated lab technician and so on. The working group will have met 10 times over the course of 4 years (2021-2024, last meeting is in October 2024). Topics that have already been covered include: brainstorming and sharing of good practices, digital management, physical storage, data standards, legal challenges, standard operating procedures, linkage and the co-creation of this tool. During some of these topics, invited speakers joined in to share their expertise. The members of the working group also communicated interesting literature, events or courses via a mailing list and shared their workflows, to stimulate discussion and receive feedback from peers.

The working group co-created two main products: "the Challenges" and "the Key". Firstly, we compiled a "List of challenges for improving DNA collection management"; from this point onwards, these will be referred to as: the Challenges. Secondly, we co-created a tool to tackle these Challenges in a more strategic, standardised way: "The key to bringing DNA collections to the next level"; from this point onwards referred to as: the Key. These two products result from consulting and discussing current best practice literature and sharing in-house best practices and ideas.

Definitions used throughout this document are given in Suppl. material 1. The following works have inspired us to address multiple topics included in the main text, as well as the in-depth information given in Suppl. material 2: Zimkus and Ford (2014), Huxley et al.

(2020), Corrales and Astrin (2023), de Mestier et al. (2023), ISBER (2023), the SPNHC wiki and the GGBN wiki. The in-depth information given in Suppl. material 2 cites inspiring work such as: Access to Biological Collection Data task group (2007), Wieczorek et al. (2012), Guralnick et al. (2014), Droege et al. (2016), Deck et al. (2017), Petersen et al. (2018), CETAF Legislations and Regulations Group (2019), Pearlman et al. (2019), Crop Trust (2020), Sayers et al. (2020), Schoch et al. (2020), Güntsch et al. (2021), Luger et al. (2021), Santi et al. (2021), Ståhls et al. (2021), Grosjean et al. (2022), Hardisty et al. (2022), Hardisty et al. (2022), Woodburn et al. (2022), Abarenkov et al. (2023), Grosjean et al. (2023), Borisenko et al. (2024), Norton et al. (2024).

Results

RESULT 1: List of all the identified challenges for DNA collections (the Challenges). Each challenge is individually numbered purely for ease of identification: there is no associated priority.

1. Execute **strategic management of the DNA collection**: have a vision, define actions to achieve that vision, monitoring and managing of the progress.

2. Ensure **participation in DNA collection management** within institutes, including: the management of the institute, the external parties (e.g. visiting scientists) that generate and re-use DNA specimens, the students that generate and re-use DNA specimens, the researchers that generate and re-use DNA specimens.

3. Ensure a **mind shift** towards **data sharing** and FAIR data (Wilkinson et al. 2016) and working “as open as possible, as closed as necessary” (Landi et al. 2020).

4. Demystify **legal aspects** surrounding genetic research and DNA collections.

5. Establish a **good workflow** that ensures successful material and data **transfer** between the field, lab and archiving activities.

6. Formulate and maintain **successful communication** procedures which ensure that the workflow and responsibilities are clear at the start and throughout each project/collaboration.

7. **Centralise and structure** DNA collection data, for example in a Collection Management System (CMS) and/or Laboratory Information Management System (LIMS).

8. Select or customise **Collection Management Systems** to meet the needs of DNA collections.

9. **Digitally unlock DNA collection data and physically unlock the DNA specimens for re-use**, for example, via the GGBN portal, DiSSCo or GBIF.

10. Store (the large volumes of) **genetic/genomic data** (unpublished) in a standardised and centralised manner and properly link them with the DNA collection data.

11. **Unlock genetic/genomic data** (published) via dedicated (open) data repositories, such as ENA.

12. **Link** DNA sequence data and DNA specimens to all related specimens, such as source specimens and voucher specimens (Buckner et al. 2021, Groom et al. 2021).

13. Determine **minimal required data** at the start of research projects which will lead to the addition of new DNA specimens to the DNA collection.

14. Good **documentation** of all the material and methods preceding the addition of a DNA sample to a collection.

15. Ensure good **management of the documentation** surrounding DNA specimens, such as Standard Operating Procedures (SOPs), workflows and/or guidelines.

16. Manage the **rapid expansion** of the number of specimens within DNA collections.

17. Enable **efficient retrieval** of DNA specimens within the collection.

18. Mitigate the risks of DNA **sample degradation**.

19. Foresee **back-up** solutions for physical and digital storage.

20. Minimise the consequences of **destructive sampling** of DNA specimens: quality loss due to handling (e.g. freeze-thaw cycles) and (theoretically) ultimately consumed specimens.

21. **Track** sample quality, usage, storage location and data.

22. **Balance resources** (time, money, people) for achieving the best possible DNA collection management.

23. Make decisions on physical storage methods, based on **limited and/or scattered information** and data about their consequences on the quality of the DNA collection.

RESULT 2: The key to bringing DNA collections to the next level (the Key). An interactive format of the Key, including guidance documentation for the seven Key-questions and 11 Key-categories is available in Suppl. material 2.

Start situation: There is a DNA collection in your organisation that you would like to mature. (Tip: If you are looking into starting a new DNA collection, read: Harati et al. (2018) and/or ISBER (2023)).

1. Do you have an up-to-date overview of all direct, internal **stakeholders** of the institute's DNA collection and are you involving them in the (current) intent to "bring the DNA collection to the next level"?
    - Yes: continue to 2.
    - No or I need help: go to Suppl. material 2 "1. Involving the stakeholders".
2. Is preserving a DNA collection **within the scope of the institute** and is the DNA collection officially recognised within the institute?
    - Yes: continue to 3.
    - No or I need help: go to Suppl. material 2 "2. Having the DNA collection within the institute's vision".
3. Do you have, on paper, a clear description of the **scope** of the DNA collection?
    - Yes: continue to 4.
    - No or I need help: go to Suppl. material 2 "3. Outlining a scope".
4. Have you outlined the **current overarching workflow** of the DNA collection?
    - Yes: continue to 5.
    - No or I need help: go to Suppl. material 2 "4. Outlining the current workflow".
5. Go to Table 1, establish and log the **starting level**. Have you been able to establish the starting level and is the assessment properly logged?
    - Yes: continue to 6.
    - No or I need help: go to Suppl. material 2 "5. Establishing the starting level".
6. **Level up**, one level at a time and log the process. Have you reached all of the goals in level 3?
    - Yes: continue to 7.
    - No or I need help: go to Suppl. material 2 "6. Levelling up".
7. Do you have a **re-evaluation strategy** for the DNA collection?
    - Yes: Perfect, all done … for now!
    - No or I need help: go to Suppl. material 2 "7. Re-using the Key".

Discussion

The Challenges and Key are two outputs which the international community of DNA collections can apply where needed and expand upon going forwards. The Challenges can be used to spark debate on DNA collection management outside the DiSSCo Flanders working group. They can be adapted and complemented as the global discussion continues. DNA collections that use the Key will mature their DNA collection in a standardised way regardless of their context, considering all different aspects in DNA collection management (i.e. the categories), stimulating a holistic growth (i.e. respecting the levels), breaking down the maturation work into manageable steps (i.e. the goals)

and setting priorities (i.e. the categories with the lowest level). Additionally, the Key is a tool that can facilitate overall communication. Firstly, between different stakeholders, such as management and researchers, as the tool allows both a “helicopter view” of the maturation process (i.e. Table 1), as well as being able to zoom in on one concrete goal at a time (i.e. Suppl. material 2). Secondly, the tool allows clear communication across different DNA collections regarding their current maturation level and (near) future ambitions, which facilitates easier sharing of experiences that can inspire, give perspective on what is realistic or even allow different DNA collections to collaborate more intensively for a set period of time when working towards the same goal. This is especially valuable given the often encountered issues with limited availability of resources and personnel: better communication between collections allows for better resource pooling and avoidance of ineffective strategies.

Table 1.

DNA collection maturation chart. A DNA collection is considered to start at level 0 in all categories (rows). If the DNA collection meets all the goals within a level (column), it achieves that level (e.g. a collection conforming to all goals within "level 1" would be a level 1 collection). Making progress (e.g. reaching one specific goal or reaching a complete new level) in this chart is considered "maturing" the collection, with a fully matured DNA collection being "level 3". An interactive format of the Key, including guidance documentation for the seven Key-questions and 11 Key-categories, is available in Suppl. material 2.

Category	Level 1	Level 2	Level 3
Involvement of suppliers	Overarching workflow and expectations + communication strategy.	DMPs (+ versioning) as a tool for genetic research is standard practice.	A project conclusion protocol is in place.
Quality management	Quality meetings happen at fixed time intervals.	A written documentation policy is in place and executed.	Quality management is effective, up-to-date and FAIR.
Legal compliance	Informing on regulatory frameworks + central workspace for logging.	Five legal checkpoints are established.	The complete DNA collection is legally held.
Physical storage	A designated current space and a best-fit plan for the near future.	Best-fit storage method is functional for new DNA specimens.	The complete DNA collection is stored in the best-fit storage method.
Contingency planning	A basic contingency plan for the physical collection and data.	A final contingency plan for the physical collection and data.	The contingency plan is revised at agreed time intervals.
Identifiers	There is a unique ID-system for each physical sample and storage location.	There is a unique ID-system for each digital record + stable links.	Unique and persistent identifier-system within the global collection.
Digital management	Centralise and standardise data of the DNA collection + DMP.	Convert unstructured data to structured data.	Durable linkage to files and other data(bases).
Unlocking the collection	The DNA collection is on GrSCiColl and the institute’s website.	A test dataset in a repository + a publication strategy.	The complete DNA collection is in a public repository.

Category	Level 1	Level 2	Level 3
Loans	Responsible person appointed, public statement on how-to-loan and loan agreement template.	A FAIR loan policy + incoming and outgoing loan procedures.	Loaning policy and procedures are operational + a Loan Agreement breach protocol.
Stability	Financial responsibilities are outlined.	Optimised staffing.	The DNA collection has a stable budget and team.
Community	Become part of at least one community.	Active participation in co-creation.	Engagement in maintaining a community.

The Challenges and Key are focused on DNA collections only, yet it is important to acknowledge that DNA collections are often embedded in a wider context, in the first place often including a parallel “source material collection” (i.e. tissues, environmental specimens or organisms the DNA specimens were created from) and, in the second place, being part of a larger institutional Natural Science collection. On the one hand, many of the Challenges and goals in the Key can be generalised as being challenges or goals linked to managing molecular collections, (biodiversity and/or environmental) biobanks or even Natural Science collections: they are transferable to many other types of Natural Science collections. Yet, on the other hand, remaining specific to DNA collections increases the relevance of the tool for DNA collection managers and avoids "scope creep". Working with the broader scope of (biodiversity) biobanks has given more generalised results; for example, having to conclude that biobank curation practices vary greatly and that they experience more broadly defined challenges (Corrales et al. 2023). Notwithstanding that learning from different types of Natural Science or biodiversity biobank collections is extremely valuable, a next step is the translation of these results into tangible actions tailored to collection managers, which will further foster their implementation. Such translation is crucial for any recommendations to be applicable in real-life collection management, but it is also especially arduous given the ongoing limitations in available resources (e.g. personnel), one of the (current) challenges. The Challenges and Key, as presented in the results, are intended to empower one or a few individuals wanting to make a difference, in a manageable (i.e. step-wise), progressive and learning-while-doing manner. If a demand for it develops, the concept of the Challenges and Key can be recycled for other types of biodiversity biobank collections, such as living microorganisms, tissue collections, environmental specimen collections or even seedbanks. The resulting parallel tools can also link to the goals and guidance documentation (Suppl. material 2) of this document when they are overlapping, to avoid duplication of effort.

DNA collections, as well as all types of Natural Science collections, cannot be uncoupled from research practices. Research relies heavily on good practices being applied in collection management in order for the DNA specimens to be of value for reuse and vice versa: DNA collections rely heavily on good practices being applied during the collection of the specimens the DNA specimens are derived from, as well as good research practices overall. Research practices, such as the usage of proper sampling techniques, detailed (meta)data documentation, data management planning, project planning, fieldwork planning and application of the FAIR principles (Wilkinson et al. 2016), will

positively benefit the quality and the data management of DNA collections. This is also apparent from Corrales and Astrin (2023) dedicating multiple chapters of their book to best practices in specimen collection and preservation, as well as other literature advocating the need for investing in training on the importance of specimen collecting and linking to specimen vouchers (Lücking et al. 2020, Buckner et al. 2021, de Mestier et al. 2023).

As the tool draws from many of the recently-compiled literature resources on the topic, its first version is expected to be complete. Nonetheless, it is worth noting that there was not an exhaustive search of (best practice) literature of DNA collections, nor is the information written in the guidance documentation (Suppl. material 2) as comprehensive as certain more elaborate guidelines or handbooks (e.g. ISBER (2023), de Mestier et al. (2023), Corrales and Astrin (2023)). It is expected that this tool will evolve as it is used, with updates being published as new versions are enriched by users. It is therefore anticipated that any currently missing information will be identified and (links to) the missing information will be added.

The Key differentiates itself from the current plethora of self-assessment tools, such as the BBMRI Self-Assesment-Service (SAS), the ISBER Biobank-Assessment Tool (BAT) and the SYNTHESYS+ Collections Self-Assessment Tool in a number of ways. Firstly, the Key obliges the collection manager(s) to (re)investigate the purpose and hence importance of the collection, prior to even thinking about adhering to good or best practices or "levelling up", as well as re-evaluating the context of the collection (e.g. stakeholders and hosting institute). Secondly, levelling up is made as tangible as possible, facilitated by: 1) the narrow scope (i.e. its being limited to one type of Natural Science collection compared to other tools), 2) defining clear goals that are achievable for any DNA collection regardless of its context, 3) immediate linking to relevant information and 4) actively sharing experiences. Thirdly, the Key is openly published and can be used for free, building on the Open Science and community mindset and aiming for (Natural Science) DNA collections of all shapes, sizes and contexts to be included.

The next step is that the Challenges and Key are tested and improved, based on user feedback. As they are hereby published openly in their first version, all (user) feedback is welcomed. The Challenges and Key can be used independently in their current format (cf. self-assessments) or their usage can be embedded in community-driven initiatives. Such community initiatives could include the organisation of user-support-sessions which allow direct engagement with the users and create inspiring and supportive interactions between DNA collection managers. The DiSSCo Flanders DNA collections working group plans to adopt this community approach and invest in testing, updating and improving the Challenges and Key, whereby developing a more interactive format of the outputs will also be considered.

Conclusions

A holistic, yet realistic approach with tangible goals facilitates the maturation of a DNA collection in an efficient, durable manner, regardless of the collections' context, size and available resources. One way to carry out this approach is to use "The key to bringing DNA collections to the next level", presented here. During the maturation process, investment in (internal) stakeholder inclusion, as well as in exchanging experiences with the local or global DNA community, is considered important for achieving real, concrete progress in the management of a DNA collection, especially when the collection's resources are limited.

Acknowledgements

We'd like to thank all the participants of the DiSSCo Flanders DNA collection working group for their contributions during the on-site and online meetings, as well as sharing relevant information in between meetings about relevant courses or literature. We are also grateful to the invited speakers who have shared information on either different digital solutions or legal expertise. Lastly, we also want to thank the passive members of the working group for having interest in this initiative.

Funding program

Research Foundation Flanders - International Research Infrastructure (FWO-IRI)

Grant title

I001721: DiSSCo Flanders: Towards a collection management infrastructure for Flanders

Hosting institution

Meise Botanic Garden

Ethics and security

All shared experiences are included in this document with the informed consent of the co-authors and the collections they represent.

Author contributions

Author contributions follow the CRediT author statements (Allen, O'Connell & Kiermer, 2019). Participation in at least one working group meeting is considered "Investigation".

Participation in the working group meeting that conceptualised this output is considered “Conceptualisation”. **Emily Veltjen**: Conceptualisation, Investigation, Writing - Original Draft, Writing - Review & Editing, Supervision. **Pieter Asselman**: Investigation, Writing - Review & Editing. **Wim Baert**: Conceptualisation, Investigation, Writing - Review & Editing. **Steve Baeyen**: Investigation, Writing - Review & Editing. **Lise Beirinckx**: Investigation, Writing - Review & Editing. **Liselot Breyne**: Investigation, Writing - Review & Editing. **Dimitri Brosens**: Investigation, Writing - Review & Editing. **Tim Claerhout**: Investigation, Writing - Review & Editing. **Sari Cogneau**: Conceptualisation, Investigation, Writing - Review & Editing. **Karen Cox**: Investigation, Writing - Review & Editing. **Laura Cuypers**: Conceptualisation, Investigation, Writing - Review & Editing. **Lynn Delgat**: Conceptualisation, Investigation, Writing - Review & Editing. **Philippe Desmeth**: Investigation, Writing - Review & Editing. **Jordi de Raad**: Conceptualisation, Investigation, Writing - Review & Editing. **Lore Esselens**: Investigation, Writing - Review & Editing. **Maria-Rose Eves Down**: Investigation, Writing - Review & Editing. **Frederik Leliaert**: Investigation, Writing - Review & Editing, Funding acquisition. **Kenny Meganck**: Investigation, Writing - Review & Editing. **Zjef Pereboom**: Investigation, Writing - Review & Editing, Funding acquisition. **Nathalie Smitz**: Conceptualisation, Investigation, Writing - Review & Editing. **Gontran Sonet**: Conceptualisation, Investigation, Writing - Review & Editing. **Maarten Trekels**: Investigation, Writing - Review & Editing. **An Vanden Broeck**: Investigation, Writing - Review & Editing, Funding acquisition. **Charlotte Van Driessche**: Writing - Review & Editing. **Aaike De Wever**: Conceptualisation, Investigation, Writing - Review & Editing, Supervision, Funding acquisition.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Abarenkov K, Andersson A, Bissett A, Finstad AG, Fossøy F, Grosjean M, Hope M, Jeppesen TS, Kõljalg U, Lundin D, Nilsson H, Prager M, Provoost P, Schigel D, Suominen S, Svenningsen C, Frøslev TG (2023) Publishing DNA-derived data through biodiversity data platforms. GBIF Secretariat https://doi.org/10.35035/doc-vf1a-nr22
- Access to Biological Collection Data task group (2007) Access to Biological Collection Data (ABCD), Version 2.06. Biodiversity Information Standards (TDWG). http://www.tdwg.org/standards/115. Accessed on: 2024-8-26.
- Addink W, Hardisty A, Islam S, Weiland C, Glöckner F, Güntsch A (2020) Advancing the Catalogue of the World's Natural History Collections - 10 recommendations from DiSSCo. Zenodo https://doi.org/10.5281/zenodo.3949839
- Astrin J, Zhou X, Misof B (2013) The importance of biobanking in molecular taxonomy, with proposed definitions for vouchers in a molecular context. ZooKeys 365: 67-70. https://doi.org/10.3897/zookeys.365.5875

- Borisenko A, Young R, Hanner R (2024) A lab-centric, workflow-based data management system for environmental DNA research. Research Ideas and Outcomes 10 https://doi.org/10.3897/rio.10.e120483
- Buckner JC, Sanders RC, Faircloth BC, Chakrabarty P (2021) The critical importance of vouchers in genomics. eLife 10 https://doi.org/10.7554/elife.68264
- Card D, Shapiro B, Giribet G, Moritz C, Edwards S (2021) Museum Genomics. Annual Review of Genetics 55 (1): 633-659. https://doi.org/10.1146/annurev-genet-071719-020506
- CETAF Legislations and Regulations Group (2019) CETAF Code of Conduct and Best Practice for Access and Benefit Sharing, Annex 5: Practice Guidance. URL: https://cetaf.org/resources/best-practices/
- Corrales C, Astrin JJ (2023) Biodiversity Biobanking – a Handbook on Protocols and Practices. Pensoft, Advanced Books https://doi.org/10.3897/ab.e101876
- Corrales C, Luciano S, Astrin J (2023) Biodiversity biobanks: a landscape analysis. ARPHA Preprints https://doi.org/10.3897/arphapreprints.e103105
- Crop Trust (2020) Seeds for Resilience Genebank reviews. https://www.croptrust.org/work/projects/seeds-for-resilience/#c7192. Accessed on: 2024-8-19.
- Cummins C, Ahamed A, Aslam R, Burgin J, Devraj R, Edbali O, Gupta D, Harrison PW, Haseeb M, Holt S, Ibrahim T, Ivanov E, Jayathilaka S, Kadhirvelu V, Kay S, Kumar M, Lathi A, Leinonen R, Madeira F, Madhusoodanan N, Mansurova M, O'Cathail C, Pearce M, Pesant S, Rahman N, Rajan J, Rinck G, Selvakumar S, Sokolov A, Suman S, Thorne R, Totoo P, Vijayaraja S, Waheed Z, Zyoud A, Lopez R, Burdett T, Cochrane G (2021) The European Nucleotide Archive in 2021. Nucleic Acids Research 50 https://doi.org/10.1093/nar/gkab1051
- Deck J, Gaither M, Ewing R, Bird C, Davies N, Meyer C, Riginos C, Toonen R, Crandall E (2017) The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples. PLOS Biology 15 (8). https://doi.org/10.1371/journal.pbio.2002925
- de Mestier A, Mulcahy D, Harris D, Korotkova N, Long S, Häffner E, Paton A, Schiller E, Leliaert F, Mackenzie-Dodds J, Fulcher T, Stahls G, von Rintelen T, Martín M, Lücking R, Williams C, Lyal C, Güntsch A, Aronsson H, Castelin M, Pielach A, Poczai P, Ruiz-León Y, Sanmartin Bastida I, Thines M, Droege G (2023) Policies Handbook on Using Molecular Collections. Research Ideas and Outcomes 9 https://doi.org/10.3897/rio.9.e102908
- de Vicente MC, Andersson M (Eds) (2006) DNA banks-providing novel options for genebanks? Topical Reviews in Agricultural Biodiversity. International Plant Genetic Resources Institute, Rome. URL: https://cgkb.cgiar.croptrust.org/images/file/learning_space/dnabanks.pdf
- Droege G, Barker K, Seberg O, Coddington J, Benson E, Berendsohn WG, Bunk B, Butler C, Cawsey EM, Deck J, Döring M, Flemons P, Gemeinholzer B, Güntsch A, Hollowell T, Kelbert P, Kostadinov I, Kottmann R, Lawlor RT, Lyal C, Mackenzie-Dodds J, Meyer C, Mulcahy D, Nussbeck SY, O'Tuama É, Orrell T, Petersen G, Robertson T, Söhngen C, Whitacre J, Wieczorek J, Yilmaz P, Zetzsche H, Zhang Y, Zhou X (2016) The Global Genome Biodiversity Network (GGBN) Data Standard specification. Database 2016 https://doi.org/10.1093/database/baw125

- Durmaz AA, Karaca E, Demkow U, Toruner G, Schoumans J, Cogulu O (2015) Evolution of Genetic Techniques: Past, Present, and Beyond. BioMed Research International 2015: 1-7. https://doi.org/10.1155/2015/461524
- Ferreira CM, Sabino-Marques H, Barbosa S, Costa P, Encarnação C, Alpizar-Jara R, Pita R, Beja P, Mira A, Searle J, Paupério J, Alves PC (2018) Genetic non-invasive sampling (gNIS) as a cost-effective tool for monitoring elusive small mammals. European Journal of Wildlife Research 64 (4). https://doi.org/10.1007/s10344-018-1188-8
- Forsdick N, Wold J, Angelo A, Bissey F, Hart J, Head M, Liggins L, Senanayake D, Steeves T (2023) Journeying towards best practice data management in biodiversity genomics. Molecular Ecology Resources https://doi.org/10.1111/1755-0998.13880
- Funk V, Gostel M, Devine A, Kelloff C, Wurdack K, Tuccinardi C, Radosavljevic A, Peters M, Coddington J (2017) Guidelines for collecting vouchers and tissues intended for genomic work (Smithsonian Institution): Botany Best Practices. Biodiversity Data Journal 5 https://doi.org/10.3897/bdj.5.e11625
- Groom QJ, Dillen M, Huybrechts P, Johaadien R, Kyriakopoulou N, Fernandez FJQ, Trekels M, Wong WY (2021) Connecting molecular sequences to their voucher specimens. Biohackrxiv https://doi.org/10.37044/osf.io/93qf4
- Grosjean M, Høfft M, Gonzalez M, Hahn A, Robertson T (2022) GRSciColl: What is the Global Registry of Scientific Collections and How to Contribute? Biodiversity Information Science and Standards 6 https://doi.org/10.3897/biss.6.93875
- Grosjean M, Robertson T, Gonzalez M, Høfft M, Hahn A (2023) Ongoing Work with the Global Registry of Scientific Collections. Biodiversity Information Science and Standards 7 https://doi.org/10.3897/biss.7.110932
- Güntsch A, Groom Q, Ernst M, Holetschek J, Plank A, Röpert D, Fichtmüller D, Shorthouse DP, Hyam R, Dillen M, Trekels M, Haston E, Rainer H (2021) A botanical demonstration of the potential of linking data using unique identifiers for people. PLOS ONE 16 (12). https://doi.org/10.1371/journal.pone.0261130
- Guralnick R, Conlin T, Deck J, Stucky B, Cellinese N (2014) The Trouble with Triplets in Biodiversity Informatics: A Data-Driven Case against Current Identifier Practices. PLOS ON 9 (12). https://doi.org/10.1371/journal.pone.0114069
- Harati M, Williams R, Movassaghi M, Hojat A, Lucey G, Yong W (2018) An Introduction to Starting a Biobank. Methods in Molecular Biology7-16. https://doi.org/10.1007/978-1-4939-8935-5_2
- Hardisty A, Addink W, Glöckler F, Güntsch A, Islam S, Weiland C (2022) A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). Research Ideas and Outcomes 7 https://doi.org/10.3897/rio.7.e67379
- Hardisty AR, Ellwood ER, Nelson G, Zimkus B, Buschbom J, Addink W, Rabeler RK, Bates J, Bentley A, Fortes JAB, Hansen S, Macklin JA, Mast AR, Miller JT, Monfils AK, Paul DL, Wallis E, Webster M (2022) Digital Extended Specimens: Enabling an Extensible Network of Biodiversity Data Records as Integrated Digital Objects on the Internet. BioScience 72 (10): 978-987. https://doi.org/10.1093/biosci/biac060
- Huxley R, Quaisser C, Butler CR, Dekker RW (2020) Managing Natural Science Collections. 1. Routledge https://doi.org/10.4324/9780429426599
- ISBER (2023) ISBER Best Practices: Recommendations for Repositories. 5. ISBER, Vancouver, 19 pp. URL: https://www.isber.org/page/BPR

- Jarman S, Berry O, Bunce M (2018) The value of environmental DNA biobanking for long-term biomonitoring. Nature Ecology & Evolution 2 (8): 1192-1193. https://doi.org/10.1038/s41559-018-0614-3
- Keck F, Couton M, Altermatt F (2022) Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Molecular Ecology Resources 23 (4): 742-755. https://doi.org/10.1111/1755-0998.13746
- Landi A, Thompson M, Giannuzzi V, Bonifazi F, Labastida I, da Silva Santos LOB, Roos M (2020) The “A” of FAIR – As Open as Possible, as Closed as Necessary. Data Intelligence 2: 47-55. https://doi.org/10.1162/dint_a_00027
- Lazareva T, Barbitoff Y, Changalidis A, Tkachenko A, Maksiutenko E, Nasykhova Y, Glotov A (2022) Biobanking as a Tool for Genomic Research: From Allele Frequencies to Cross-Ancestry Association Studies. Journal of Personalized Medicine 12 (12). https://doi.org/10.3390/jpm12122040
- Lücking R, Nadel M, Araujo E, Gerlach A (2020) Two decades of DNA barcoding in the genus Usnea (Parmeliaceae): how useful and reliable is the ITS? Plant and Fungal Systematics 65 (2): 303-357. https://doi.org/10.35535/pfsyst-2020-0025
- Luger T, Maes S, Milkers A, Bergmans B (2021) Een collectieplan schrijven? Dat doe je zo! [Writing a collection plan? Like this!]. OBT, leiden, 88 pp. URL: https://www.cultureelerfgoed.nl/publicaties/publicaties/2021/01/01/een-collectieplan-schrijven-dat-doe-je-zo [ISBN 978-90-5799-343-5]
- Norton B, Woodburn M, Grant S, Jones J, Webbink K, Vincent S, Trekels M, Buschbom J, Baskauf S, Biodiversity Information Standards (TDWG) (2024) Latimer Core (LtC) GitHub Repository. https://github.com/tdwg/ltc. Accessed on: 2024-4-09.
- Pearlman J, Bushnell M, Coppola L, Karstensen J, Buttigieg PL, Pearlman F, Simpson P, Barbier M, Muller-Karger F, Munoz-Mas C, Pissierssens P, Chandler C, Hermes J, Heslop E, Jenkyns R, Achterberg E, Bensi M, Bittig H, Blandin J, Bosch J, Bourles B, Bozzano R, Buck JH, Burger E, Cano D, Cardin V, Llorens MC, Cianca A, Chen H, Cusack C, Delory E, Garello R, Giovanetti G, Harscoat V, Hartman S, Heitsenrether R, Jirka S, Lara-Lopez A, Lantéri N, Leadbetter A, Manzella G, Maso J, McCurdy A, Moussat E, Ntoumas M, Pensieri S, Petihakis G, Pinardi N, Pouliquen S, Przeslawski R, Roden N, Silke J, Tamburri M, Tang H, Tanhua T, Telszewski M, Testor P, Thomas J, Waldmann C, Whoriskey F (2019) Evolving and Sustaining Ocean Best Practices and Standards for the Next Decade. Frontiers in Marine Science 6 https://doi.org/10.3389/fmars.2019.00277
- Petersen M, Glöckler F, Kiessling W, Döring M, Fichtmüller D, Laphakorn L, Baltruschat B, Hoffmann J (2018) History and development of ABCDEFG: a data standard for geosciences. Fossil Record 21 (1): 47-53. https://doi.org/10.5194/fr-21-47-2018
- Raxworthy C, Smith BT (2021) Mining museums for historical DNA: advances and challenges in museomics. Trends in Ecology & Evolution 36 (11): 1049-1060. https://doi.org/10.1016/j.tree.2021.07.009
- Ryan MJ, Schloter M, Berg G, Kostic T, Kinkel LL, Eversole K, Macklin JA, Schelkle B, Kazou M, Sarand I, Singh BK, Fischer D, Maguin E, Ferrocino I, Lima N, McClure RS, Charles TC, de Souza RS, Kiran GS, Krug HL, Taffner J, Roume H, Selvin J, Smith D, Rybakova D, Sessitsch A (2021) Development of Microbiome Biobanks – Challenges and Opportunities. Trends in Microbiology 29 (2): 89-92. https://doi.org/10.1016/j.tim.2020.06.009
- Santi I, Casotti R, Comtet T, Cunliffe M, Koulouri P(, Macheriotou L, Not F, Obst M, Pavloudi C, Romac S, Thiebaut E, Vanaverbeke J, Pade N (2021) European Marine

Omics Biodiversity Observation Network (EMO BON) Handbook (Version 1.0). EMBRC-ERIC https://doi.org/10.25607/obp-1653

- Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare R, Banday S, Mishra A, Das G, Malonia S (2023) Next-Generation Sequencing Technology: Current Trends and Advancements. Biology 12 (7). https://doi.org/10.3390/biology12070997
- Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, Karsch-Mizrachi I (2020) GenBank. Nucleic Acids Research 49 https://doi.org/10.1093/nar/gkaa1023
- Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, Sharma S, Soussov V, Sullivan JP, Sun L, Turner S, Karsch-Mizrachi I (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database 2020 https://doi.org/10.1093/database/baaa062
- Ståhls G, Aleixo A, Hyvärinen M, Kuusijärvi A, Myllys L, Väre H, Sihvonen P, Luttinen A, Kröger B, Oinonen M, Schulman L, Juslén A (2021) The Genomic Resources Collection Policy of the Finnish Museum of Natural History. Research Ideas and Outcomes 7 https://doi.org/10.3897/rio.7.e61664
- Veltjen E, Brosens D, Cools C, De Cuyper B, De Vos B, D'hondt B, Herr C, Lambeens I, Maebe K, Lettens S, Pollet M, Soors J, Teunen L, Vanden Broeck A, Van Driessche C, Verstraeten A, De Wever A (in preparation) INBO-beslissingsboom: Archiveren van natuurwetenschappelijke specimens (of niet)? INBO.
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1). https://doi.org/10.1371/journal.pone.0029715
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1). https://doi.org/10.1038/sdata.2016.18
- Woodburn M, Buschbom J, Droege G, Grant S, Groom Q, Jones J, Trekels M, Vincent S, Webbink K (2022) Latimer Core: A new data standard for collection descriptions. Biodiversity Information Science and Standards 6 https://doi.org/10.3897/biss.6.91159
- Zimkus BM, Ford LS (2014) Best practices for genetic resources associated with natural history collections: Recommendations for practical implementation. Collection Forum 28: 77-112. https://doi.org/10.14351/0831-0005-28.1.77

Supplementary materials

Suppl. material 1: Definitions doi

Authors: Emily Veltjen, Pieter Asselman, Wim Baert, Steve Baeyen, Lise Beirinckx, Liselot Breyne Dimitri Brosens, Tim Claerhout, Sari Cogneau, Karen Cox, Laura Cuypers, Lynn Delgat, Philippe Desmeth, Jordi de Raad, Lore Esselens, Maria-Rose Eves Down, Philippe Helsen, Frederik Leliaert, Kenny Meganck, Zjef Pereboom, Nathalie Smitz, Gontran Sonet, Maarten Trekels, An Vanden Broeck, Charlotte Van Driessche, Aaike De Wever

Data type: Definitions

Brief description: Definitions used throughout the manuscript and guidance documentation.

Download file (89.09 kb)

Suppl. material 2: Interactive format of the Key including the guidance documentation doi

Authors: Emily Veltjen, Pieter Asselman, Wim Baert, Steve Baeyen, Lise Beirinckx, Liselot Breyne Dimitri Brosens, Tim Claerhout, Sari Cogneau, Karen Cox, Laura Cuypers, Lynn Delgat, Philippe Desmeth, Jordi de Raad, Lore Esselens, Maria-Rose Eves Down, Philippe Helsen, Frederik Leliaert, Kenny Meganck, Zjef Pereboom, Nathalie Smitz, Gontran Sonet, Maarten Trekels, An Vanden Broeck, Charlotte Van Driessche, Aaike De Wever

Data type: Guidance documentation

Brief description: For each of the seven main questions of the Key and each of the 11 categories in Table 1 of the Key, guidance documentation is provided. The 18 different guidance documentation chapters are each time structured in four sections: 1) Repetition of either the question (1-7) or the category and its three goals (8-18); 2) Explanation: giving more in depth information on the specific question, category or goals; 3) Importance: explaining why the question or category at hand is important; 4) Experiences: providing an overview of experiences (synonyms in this context: “stories”, “use-cases”) to inspire tackling a specific question, category or goal.

Download file (623.98 kb)