



BBMRI-NL Workshop

Sustainable sample/data infrastructures

Rogier van der Stijl^{1,2}, Bart Scheerder^{2,3}, and Lisette Eijdemans^{1,2}

July 2018

¹Biobanking and BioMolecular resources Research Infrastructure, the Netherlands

²University Medical Center Groningen

³Biobanking and cohort studies Knowledge and Expertise Center

Table of contents

Management summary.....	3
Organisational context.....	5
Introduction	5
Methods.....	6
Results.....	7
Workshop 1: Round table – Sustainable sample/data infrastructures	7
Workshop 2: Value-model	10
Session 1: Stakeholders and their interests.....	10
Session 2: Value, value categories, and value parameters.....	13
Session 3: Stakeholder-value matrix.....	15
Workshop 3: Overarching solutions for sustainability.....	15
Session 1: Problem mapping.....	16
Session 2: Idea generation	19
Session 3: Idea development	20
Discussion	22
Acknowledgements	22
Supplementary information	23

Disclaimer

This report describes the outcomes of a BBMRI-NL workshop organised by the authors. The results, views, and opinions described in this report were made in the context of the workshop and do not necessarily reflect the position of the authors or the official policy or position of BBMRI-NL and its Partners.

Users may make free use of the information provided in this report under the condition that when such information is used, distributed, or reproduced –in current or altered form–, BBMRI-NL and this report are cited as a source.

Management summary

Background

Sample and data infrastructures in the medical sciences, such as biobanks and medical registries, need to operate for prolonged periods of time if they want to deliver results and fulfill their promise of personalized medicine. However, it is precisely this sustainable existence that is a constant challenge. And as such, sustainability is a pressing issue for individual infrastructures and for national and international policymakers and funders. To support the Dutch infrastructure field with this challenge we organised a meeting aimed at gathering and sharing experiences, challenges, and best practices, and finding main causes and overarching solutions.

Methods

During the meeting, the 22 participants were divided into three groups and rotated along three workshops: 1) a round table discussion on sustainability; 2) a session on stakeholders and value, and their relationships; and 3) a brainstorm session to find causes and overarching solutions. In workshops 2 and 3 each group would build upon the results of the previous group.

Workshop 1: Round table – Sustainable sample/data infrastructures

In workshop 1 the participants broadly discussed the subject of sustainability. Sample/data quality, financing, stakeholders, (re)use of samples/data, rules and regulations, and privacy were identified as main subjects linked to infrastructure sustainability. The main challenges with regards to sustainability consisted of finding sufficient funding, deciding on maintaining or disposing older sample/data collections, obtaining long-term stakeholder commitment, and handling changes in rules and regulations. In addition, the differences in access, review, and issuance policies between the various institutes were named as a hurdle. To become or remain sustainable infrastructures have to provide excellent services to their users and pay close attention to their communication and their relationship with key stakeholders.

Workshop 2: Value-model

In workshop 2 the aim was to build a value-model and combining stakeholders with the different forms of value contained within an infrastructure. First the participants identified the main stakeholders of sample/data infrastructures. In addition, they formulated perceived interests each stakeholder might have for collaborating with an infrastructure. Most stakeholders could be classified as user, supporter, or funder, with each group having comparable (types of) interests. Second, the participants determined which categories of value a sample/data infrastructure contains and which (adjustable) parameters subsequently determine this value. This resulted in an infrastructure value-model. What became clear is that most of the value categories are dependent on one another, and that performing poorly in a particular category diminishes the value of the infrastructure as a whole. The intended third step of combining the identified stakeholder interests with the different categories and parameters of value into a stakeholder-value matrix was not accomplished due to time restrictions.

Workshop 3: Overarching solutions for sustainability

In workshop 3 the aim was to work towards potential overarching solutions to improve the sustainability of the whole Dutch sample and data infrastructure landscape. To do this the participants followed the different phases of the brainstorm process, starting with formulating the following problem definition: infrastructures for sample / data collections cannot realise the necessary continuity.

As a first step a problem map containing the problem's causes and sub-causes was established. Six main causes were identified: 1) lack of a long-term business model; 2) variable sample/data quality; 3) current system rewards and focuses on new instead of reuse; 4) constantly

changing external factors; 5) too many infrastructures; and 6) mostly project-based funding instruments. Furthermore the participants discussed the poor image infrastructures currently have and the lack of incentive for researchers to share the collected samples and data.

Subsequently, solutions for the abovementioned main causes were gathered during a brainstorm session. The four most promising solutions were further elaborated:

- 1) The issuance of samples and data should be stimulated by:
 - a. giving credit for establishing the infrastructure and sharing the samples and data;
 - b. jointly establishing transparent public-private partnership models to enable research by private parties;
 - c. developing a quality mark for infrastructures ;
 - d. better communicating the value of infrastructures;
 - e. increasing visibility of already collected samples and data via an up-to-date catalogue;
 - f. and establishing funding incentives that encourage the (re)use of existing infrastructures.
- 2) A national NL-biobank would prevent further fragmentation and simplify standardisation.
- 3) A clear long-term national vision on sample/data infrastructures would provide a starting point for resolving outstanding policy issues and is necessary to work towards overarching solutions.
- 4) A long-term vision of research funders, linked to accompanying sustainability measures, would enable sample/data infrastructures to position themselves accordingly.

Discussion

The meeting was successful in bringing together a diverse group of participants to discuss issues related to sustainability. This provided a rich yield of information, presented in this report. What became clear is that, although each type of infrastructure also experiences its own individual challenges, there are many shared challenges that could be effectively addressed. The gathered results provide a promising starting point to start a dialogue with the various other stakeholders; with the goal of verifying and supplementing these results and working towards overarching solutions. In the end, a multiparty approach is essential to achieve lasting results and establish a future-proof Dutch sample and data infrastructure.

Organisational context

The meeting was organised by BBMRI-NL as part of work package 6 “interactive and sustainable biobanking”. Work package 6 focuses on the sustainability of sample and data collections and infrastructures and, what is seen as an important aspect of sustainability, the interactive involvement of stakeholders.

Introduction

Sample/data infrastructures enable the collection, management, and use of large amounts of samples and/or data for medical research and personalised medicine. The term “sample/data infrastructure” refers to population, clinical, and further-use biobanks, medical data registries, and centralised storage facilities managing multiple sample/data collections. We deliberately focus not only on biobanks but also include non-sample databases, as data is becoming increasingly important¹. The amount of gathered data is growing exponentially, in part due to developments in the imaging, sequencing, and –omics fields. In addition, the linking of existing biobanks and registries offer new opportunities for research. As a result, the fields of samples and data are becoming more and more intertwined.

Sample/data infrastructures need to be maintained for prolonged periods of time to allow the samples and data to be gathered and put to use by researchers. Getting actionable results takes more than a decade, and therefore such infrastructures need to be sustainable to deliver on their promise of personalised medicine. However, it is precisely this sustainability that is a constant challenge². And as such, sustainability is a pressing issue for individual infrastructures and for national and international policymakers and funders.

The sustainability of sample/data infrastructures depends on the interplay between social, operational, and financial aspects³. There is a need for best practices and shared solutions. However, both are hampered by the variety of organisational structures, individual contexts, and fragmentation of the sample and data field. Furthermore, it remains difficult for infrastructures to communicate their value to all stakeholders involved.

By organising a meeting on sustainable sample/data infrastructures we intend to support the (Dutch) field in their sustainability challenge. The goal is to gather and share knowledge, experiences, individual challenges, and best practices on the broad subject of sustainability. In addition, we attempt to build an infrastructure value-model and work on overarching solutions for infrastructure sustainability. The results are made available for the biobanking community and used as input for subsequent stakeholder meetings to create a future-proof Dutch sample and data infrastructure.

¹ Actually some argue that a biobank could be seen as a data repository containing raw data (samples), sample-related metadata (e.g. processing and storage information), supplementary data (e.g. clinical annotations, questionnaires), data from analyses (e.g. DNA sequencing), and data linkages (e.g. links to other databases). (based on: Quinlan Philip R. et al., *The Informatics Challenges Facing Biobanks: A Perspective from a United Kingdom Biobanking Network*, *Biopreservation and Biobanking*, 2015 Oct 1; 13(5): 363–370, <http://doi.org/10.1089/bio.2014.0099>)

² Sustainability is broadly defined as the “capacity to endure”. This is too simplistic for sample/data infrastructures, as just enduring is not the intention. Still, existing or operating over a prolonged period of time is necessary for a sample/data infrastructure to fulfil its function, which is the gathering, processing, maintaining, and distributing of sample and data for current and future research. An agreed definition is missing, but sustainability for a sample/data infrastructure could be defined as: “being able to create and maintain value for its main stakeholders, now and in the future (or for how long is required)”.

³ Watson Peter H. et al., *A Framework for Biobank Sustainability*, *Biopreservation and Biobanking*, 2014 Feb;12(1):60-8, <http://doi.org/10.1089/bio.2013.0064>

Methods

The meeting took place on the 20th of April 2018 in Utrecht, The Netherlands. It consisted of three workshops on different subjects related to infrastructure sustainability. The twenty-two participants originated from twenty sample/data infrastructures, consisting mostly of population, clinical, and further-use biobanks and central storage facilities, supplemented with several data registries (see supplementary table 2 for a list of participants, their sample/data infrastructures and hosting institutions). The participants were spread across three groups of six to eight persons each. To obtain multiple perspectives each group contained a mix of organisational backgrounds.

The groups rotated along three workshops, with each workshop session lasting 45 minutes. Audio recordings were made of workshop 1 and 3. In workshops 2 and 3 each group would build upon the results of the previous group. Dutch was used as main language. The workshops were designed as follows:

Workshop 1: Round table – Sustainable sample/data infrastructures

The goal of this workshop was to gather and share challenges and best practices on sustainability. Each group discussed several pre-formulated questions that touched upon different subjects surrounding the sustainability of sample/data infrastructures. The workshop was chaired by Rogier van der Stijl.

Workshop 2: Value-model

This workshop was about the different concepts of value that biobanks offer. As value depends on the eye of the beholder the workshop also addressed the different stakeholder groups and their perceived interests. A sample/data infrastructure contains many forms of value. However, to be(come) sustainable an infrastructure has to convert its values into resources. The goal of this workshop was to establish what the participants see as value, who they see as important stakeholders, and how to connect these values to the different stakeholders in a stakeholder-value matrix. Each session tackled a different subject. The participants adjusted and supplemented pre-made and partially filled tables offered as starting material. The workshop was chaired by Bart Scheerder.

- *Session 1 – Stakeholders:* Which stakeholders exist around sample/data infrastructures, and what are their interests?
- *Session 2 – Value:* Which forms of “value” does a sample/data infrastructure contain? And what parameters define this value?
- *Session 3 –Stakeholder-value matrix:* Which values best match the interests of which stakeholders? And in what way do they match?

Workshop 3: Overarching solutions for sustainability

The goal of this workshop was to come up with solutions to improve the sustainability of the whole Dutch sample and data infrastructure. To arrive at concrete solutions, we followed the different phases of the brainstorm process (problem mapping, idea generation, and idea development), with each group dealing with a subsequent phase. The results of this workshop will be used as input for future stakeholder meetings. The workshop was chaired by Lisette Eijdens. In addition, Rick van Nuland, programme manager Health-RI, participated in all three sessions as linking pin for future developments and meetings. Pre-formatted posters and a starting problem definition⁴ were available to guide to participants.

- *Session 1 – Problem mapping:* What is the problem? And what are its (most important) causes and consequences?

⁴ Starting problem definition: Infrastructures for sample / data collections cannot realize the necessary continuity (In Dutch: Infrastructuren voor sample / data collecties kunnen niet de noodzakelijke continuïteit realiseren)

- *Session 2 – Idea generation:* What are potential solutions for the named causes?
- *Session 3 – Idea development:* What are the most promising solutions? What preconditions are needed? Which parties should be involved?

The results of each workshop and each session have been processed and are presented in this report. Before publication the participants received the opportunity to provide feedback on and further contribute to these results and accompanying discussions.

Results

Workshop 1: Round table – Sustainable sample/data infrastructures

During this workshop the participants had the opportunity to share and discuss their personal experiences, challenges and practices surrounding infrastructure sustainability. As each group discussed the same questions the results of all three sessions are combined and summarized below.

What does sustainability mean for sample/data infrastructures?

As a starting exercise, participants were asked to write down one to three subjects that they considered to be related to infrastructure sustainability. All groups named the subjects of quality, financing, and stakeholders. High quality of samples and data was considered a major requirement for sustainability. Standard operating procedures, uniform collection and storage conditions, and appropriate metadata on sample and data origin and processing were some of the mentioned indicators of quality. The gathered samples should be prepared and stored in such a way that they allow a wide range of follow-up measurements. High, comparable, and continuous quality improves the long-term usability of collected samples and data beyond the first round of research questions. Furthermore, researchers are more willing to use such high quality samples and data in their projects, which should eventually contribute to the infrastructure's overall image.

In addition to quality, sufficient funding is also a crucial part of sustainability. Participants stressed the importance of an adequate funding horizon⁵. A too short funding horizon forces an infrastructure to focus on short-term actions; instead of making the necessary investments for long-term sustainability. The other side of funding is costs. To keep costs manageable an infrastructure has to make decisions on which samples and data to collect, both in terms of the type and the appropriate amount. Also it has to decide which gathered samples and data to maintain, and for how long.

Furthermore, sustainability means working together with relevant stakeholders, for example by involving stakeholders in infrastructure governance or by direct collaboration between collectors, users, and facilitators. Stakeholder engagement and transparency should be the standards to ensure participant retention. All stakeholders should work together on finding joint solutions and aligning activities, thus preventing fragmentation and waste of resources.

Other subjects that were mentioned as part of sustainability were the (re)use of samples and data, for example via FAIR⁶ data management; the adherence to current rules and regulations; and the proper handling of participant privacy, informed consent and related issues. Clear Dutch informed consent guidelines that support both infrastructure sustainability and protect participant rights would be beneficial.

⁵ The funding horizon describes how far an infrastructure can plan ahead before the next round of funding is required.

⁶ FAIR stands for Findable, Accessible, Interoperable, and Reusable. For more information see: <https://www.dtls.nl/fair-data/fair-principles-explained/>

What are the most important challenges for long-term continuity of your sample/data infrastructure?

About half of the participants said they were worried about their infrastructure's sustainability. One of the main challenges was finding sufficient funding. It was recommended to start on time with finding the next round of financing. The field tries to further address this challenge by working together and aligning activities, hereby increasing efficiency. Funders benefit from these kind of collaborations, but at the same time also finance new initiatives, without the condition that these new initiatives collaborate, join and/or align with existing infrastructures. There is a clear role here for funders to prevent fragmentation, promote reuse, and increase efficiency.

A frequently mentioned challenge was on how to decide if an older collection should be maintained or disposed. This challenge is especially relevant to central infrastructures that manage multiple collections. Important factors in such a decision are the position of the responsible party, the overall state of sample quality, the presence and broadness of an informed consent, and the existence of a link between samples and associated (clinical) data. A first step to address this challenge could be to work towards a joint and clear decision protocol⁷. As such a decision is ultimately about value (does a collection or infrastructure still have value to stakeholders and is therefore worth maintaining?) the results of workshop 2 can serve as input.

Another challenge relates to data enrichment and follow-up. Sufficient metadata, data maintenance, and data enrichment are necessary to prevent data usability from going down over time, hereby negatively affecting long-term sustainability. However, this requires a considerable investment in time and expertise. This is especially challenging for clinical collections, as the responsible health care professionals do not have sufficient time to do this in addition to their clinical duties.

Other challenges that were discussed involved the difficulty of obtaining long-term commitment from the hosting organisation, often due to changes in higher management positions and priorities; the changes in laws and regulations (e.g. GDPR) and how these affect informed consent and an infrastructure's overall organisation; and the differences in application, review, and issuance procedures between different infrastructures and hosting organisations. This last challenge makes it difficult for researchers to obtain samples, hereby hampering the (re)use of existing infrastructures.

What must every sample/data infrastructure do to become sustainable?

Providing excellent services to the users should come first. Positive user experiences provide word of mouth and brand awareness. Through snowballing this should lead to new users and hereby contribute to infrastructure sustainability. Part of the user experience is determined by sample and data quality. Here as well, high, comparable, and continuous quality was emphasized as a requirement for sustainability. Furthermore, informed consent must be arranged in such a way that it supports sustainability, for example by allowing the infrastructure to use the gathered material for new research questions and to collaborate with private parties under set conditions.

Communication is also an important factor. An infrastructure needs to showcase its value by communicating to participants, the general public, and other stakeholders about what is being done with the gathered samples and data. Many participants want to receive information about research outcomes. Infrastructures could provide this information on a group level. And for some basic measurements (e.g. blood pressure, blood glucose) even on an individual level, without going into diagnosis. However, communication should be handled with care, especially related to privacy and consent, as misinterpretation can negatively affect an infrastructure's image. Therefore, in addition to following rules and regulations, an infrastructure should also take into account the prevailing

⁷ The Radboud Biobank, the centralised facility of the Radboud University Medical Center, has developed a flowchart to address this challenge.

public opinion. Involving non-researchers, such as patient or donor representatives, in the infrastructure's communication can prevent issues.

A sample/data infrastructure is only valuable if its samples and data are (re)used. What do you do to promote the use of samples and data?

All participants agree that, generally, infrastructures are insufficiently used. And most also acknowledge that their own infrastructure could be used more. Overall, data are used more than samples. This is in part because data can be infinitely duplicated and distributed, while samples are a finite and more costly resource. Often only a subset of the whole collection is (re)used. Beforehand, however, it is hard to predict which samples and data will be most requested. Within a constantly changing scientific field, specific samples and data can become relevant or irrelevant over time.

The participating infrastructures employ several methods to increase (re)use. Visibility within the own hosting organisation is often a priority as it represents the main user target group. This is done by organising and visiting internal events and through word of mouth. In addition, the researchers involved in collecting samples and data are often also their main users. To further increase external visibility infrastructures visit scientific conferences, present showcases, publish scientific papers, and maintain a website. Some use social media to reach the general public, for example to communicate research findings or relate the infrastructure to current events. Although such messages are read, it is a time-consuming process and it is unclear if it leads to additional users. BBMRI-NL aims to support infrastructure visibility through the BBMRI-NL Catalogue⁸ and BBMRI-NL Podium⁹. Infrastructures can also help each other by referring to each other's collections.

What do you expect from BBMRI-NL on the subject of sustainability?

To promote sustainability, the participants expect BBMRI-NL to provide the right tools (e.g. database and registration tools), and support these tools on the long-term. This offers certainty to infrastructures and enables them to work together more effectively because everyone is using the same tools. Also, participants expect an up-to-date and complete BBMRI-NL Catalogue that includes both samples and datasets, is easily accessible and linkable, and supports infrastructure visibility and user requests.

In addition to tools, the participants expect BBMRI-NL to organise relevant events and workshops to share experiences and, particularly, to jointly look for solutions for problems that hinder sustainability. This can be on policy issues (e.g. nWMO harmonisation¹⁰, joint application review procedures) and on more practical issues (e.g. biobank manager workshops, data handling workshops). For such events, BBMRI-NL should involve the appropriate external parties (e.g. COREON¹¹, DTL¹²).

⁸ BBMRI-NL Catalogue (publicly available), <https://catalogue.bbmri.nl/>

⁹ BBMRI-NL Podium, <https://www.bbmri.nl/podium/>, <https://podium.bbmri.nl/#/>

¹⁰ nWMO relates to medical research on human samples and data which is not covered by the Dutch Act on Medical-Scientific Research on Human Beings (WMO, Wet medisch wetenschappelijk onderzoek met mensen). WMO research is assessed by a medical ethical assessment committee according to standardised guidelines. nWMO research is not always assessed. And when it is assessed the guidelines and procedures can differ between institutions.

¹¹ COREON is the Commission for Research Regulation (COMmissie REgelgeving ONderzoek) coordinated by Federa (Foundation Federation of Dutch Medical Scientific Societies), <https://www.federa.org/federa-english>

¹² DTL, the Dutch Techcentre for Life Sciences, is a public-private partnership of more than 50 life science organisations in the Netherlands, <https://www.dtls.nl/>

Furthermore, BBMRI-NL should continue to counteract fragmentation in policy, logistics, and funding by further harmonising the differences between the hosting institutions and infrastructures, and by working together with funding agencies and the national government.

Finally, BBMRI-NL should continue to emphasise the importance of sample/data infrastructures, thus making sure this subject remains high on the agenda of key stakeholders.

Workshop 2: Value-model

Session 1: Stakeholders and their interests

Which stakeholders¹³ exist around sample/data infrastructures?

The sustainability of any organisation, including a sample/data infrastructure, depends on the long-term commitment of its most important stakeholders^{14,15}. A sample/data infrastructure exists in an extensive “ecosystem” of stakeholders. Different stakeholders are involved in running the infrastructure, collecting and using the samples and data, and providing funds and revenues. The identified stakeholders are described in table 1.

What are the interests of these stakeholders?

Each stakeholder has different interests for using, joining, or supporting a sample/data infrastructure. It is important to be aware of these interests as it enables an infrastructure to act on them, for example by customising their communication, prioritizing their actions, changing their governance structure, or adjusting their value proposition. Satisfying a stakeholder’s interests will help to ensure their (continued) commitment. The results of this session are therefore useful for sample/data infrastructures to position themselves accordingly. For each stakeholder group the workshop participants, based on their knowledge and experience, identified several presumed interests and needs. The results are shown in table 2. A next step would be to verify and supplement the results with each group of stakeholders.

When taking a closer look at the identified interests it is possible to make a subdivision between the various stakeholders. Researchers and private industry are part of a “user” group with partly comparable interests, for example on the quality of samples and data, and on saving time and money. However, even comparable interests may still result in different expectations per stakeholder. For example, private industry places high demands on samples and data, requiring strict quality assurance, collection protocols, and documentation before being able to use the samples and data in their drug discovery and development.

Donors, patients, patient advocacy groups, and citizens, and to a lesser extent doctors and the government, could be seen as a “supporter” group of stakeholders. They support the sample/data infrastructure each in their own way and for their own interests. These interests are often outcome-related. Something that separates this group from the previous group is that the outcomes they pursue are not directly useful for the individual stakeholder themselves, but more for society as a whole.

¹³ Stakeholders are individuals, groups, or organisations that are affected by or can affect a particular action undertaken by others (Roger Bjugn and Bettina Casati, *Stakeholder Analysis: A Useful Tool for Biobank Planning*, Biopreservation and Biobanking, 2012 Jun;10(3):239-44, <http://doi.org/10.1089/bio.2011.0047>).

¹⁴ Bourne L. *Why stakeholders matter*. In: Stakeholder Relationship Management. Farnham: Gower Publishing; 2009:11–27.

¹⁵ Bryson JM. *Clarifying organizational mandates and mission*. In: Strategic Planning for Public and Nonprofit Organizations, 4th ed. San Francisco: Jossey-Bass; 2011:117–149.

The last group that can be identified is the “funder” group, consisting of funding agencies and charities. They have largely comparable interests, focussed on efficiency and quality. The remaining three stakeholders –health insurers, the biobank itself, and the hosting organisation– cannot be grouped.

Table 1. Identified stakeholders and their description (ordered alphabetically)	
Stakeholder	Description*
Charities	General and disease specific foundations, such as the Dutch Cancer Society, that often act as funders of research and related infrastructure.
Citizens	The general population.
Medical doctors** (non-researchers)	Medical staff, often involved in infrastructure initiation, governance, participant recruitment, and sample and/or data collection.
Donors	Healthy participants that voluntarily donate samples and/or data to the infrastructure.
Funding agencies	Organisations, often governmental, that fund research and related infrastructure. Examples from the Netherlands are ZonMW and NWO***.
Government	Regional and national governmental authorities and agencies which can act as funders, regulators and beneficiaries.
Health insurers	Play an important role in the Dutch health care system as care purchaser, in between patients and health care providers. They aim to provide efficient quality health care to their clients.
Hosting organisation	A sample/data infrastructure is often embedded in a larger organisation, such as a research institute or hospital.
Patients	Participants with a specific diagnosis that voluntarily donate samples and/or data to the infrastructure.
Patient advocacy groups	An organisation representing a specific group of patients.
Private industry	This is a diverse group containing pharmaceutical and biotech companies.
Researchers	Researchers that use the samples and/or data to do research.
The biobank	The samples/data infrastructure itself. The interests will depend on the adopted organisational structure and the position of the infrastructure within the hosting organisation (e.g. central facilitating storage infrastructure, department collection, population cohort, commercial biobank).
* Descriptions were not made during the workshop, but added afterwards.	
** This group does not use the gathered samples and/or data themselves. If they did, they would transfer to the “researchers” group. It is possible for individuals to change from one stakeholder group to the other or belong to multiple stakeholder groups.	
*** NWO, Netherlands Organisation for Scientific Research.	

Table 2: Who are the biobanking stakeholders? And what are their interests?

Stakeholders	Interests							
Charities	Control costs	Prevent research waste	Increase quality	FAIR for many stakeholders	Optimal use of invested funds	Transparency	Health care innovations	
Citizens	Health care innovations	Prevention	Lower health care costs	Evaluation of treatment effect	Knowledge on disease	Personalized medicine	Find predictors of disease	
Doctors (non-researchers)	New biomarkers to monitor, diagnose, or forecast disease	Personalized medicine						
Donors	Altruism	Contribute to science	Knowledge on disease	Incidental findings	Medical check-up	Transparency	Health care innovations	
Funding agencies	Control costs	Prevent research waste	Increase quality	FAIR for many stakeholders	Optimal use of invested funds	Transparency		
Government	Data for policy	Health care innovations	Increase employment	Strengthen competitive position	Societal interest			
Health insurers	E-Health	Registration at the source	Improved treatment	Prevent over-utilization	Prevention	Lower health care costs		
Hosting organisation	Increase research opportunities	Create brand awareness	Strengthen competitive position	Attract research funds	More and improved publications			
Patients	Health care innovations	Evaluation of treatment effect	Knowledge on disease and its course	Personalized medicine	Find predictors of disease			
Patient advocacy groups	Transparency	Health care innovations	Knowledge on disease and its course	Personalized medicine	Find predictors of disease	Health care innovations	Evaluation of treatment effect	
Private industry	Answer hypothesis	Control studies and validation	Selection for phase 3 trials	Quality samples and data	Service	Save time & money		
Researchers	Translational, prospective, and longitudinal research	Find causal insights	Save time & money	Outsource logistics to focus on research	Service (“additional hands”)	Raising funds for research	Quality samples and data	More and improved publications
The biobank	Provide samples and data to relevant research	Increase efficiency (CI)	Ensure and increase quality (CI)	Optimal use of collected samples and/or data				

CI: Central infrastructure that facilitates samples and/or data related operations within a hosting organisation

Session 2: Value, value categories, and value parameters

To ensure stakeholder commitment, a sample/data infrastructure needs to create “value” for each of its stakeholders. However, value is a broad and subjective concept; its meaning depends on perspective and therefore varies between stakeholders. In this session, the participants determined which forms of value a sample/data infrastructure contains and from which (adjustable) parameters this value is built. For each type of sample/data infrastructure (e.g. clinical biobanks, central storage infrastructure) there will be different categories of value and different accompanying parameters. For this session, the participants adopted the viewpoint of a “de novo”¹⁶ biobank. Partly pre-filled tables were used as starting material.

Which forms of value does a sample/data infrastructure contain?

The participants discussed the starting material and decided on the value categories listed in table 3.

Table 3. Infrastructure value categories	
Value category	Description*
Organisation	The value inherent to an infrastructure’s organisation and activities.
Participants	The scientific and economic value represented by the population characteristics of included participants.
Samples	The usefulness for research purposes of particular samples.
Data	The usefulness for research purposes of particular data.
Output	Scientific, societal or commercial output produced, directly or indirectly, by making use of assets of a sample/data infrastructure.
Services	Value generated by providing services.
Education and training	Added value through the generation and subsequent transfer of knowledge.
Brand value	The value derived from having a well-known brand name, which results in positive brand recognition and awareness.
Economic value	The value the infrastructure provides for its users, investors, and the general economy.
(Re)use	The typical goal of a sample/data infrastructure is to store the right samples/data under the right conditions for the right amount of time and make sure they are (re)used as much as possible for relevant research.
*Descriptions were not made during the workshop, but added afterwards.	

The discussion on the education and training value category led to new insights on how an infrastructure can be valuable. By training personnel, students, and users, on for example how to correctly collect and process large amounts of samples and/or data for scientific research, the infrastructure is of value to both the hosting institution and to society at large. The training of (former) biobanking staff will reflect on their future activities, hereby raising the overall level of scientific research.

The total value of an infrastructure is determined by the sum of its parts. The value categories are intertwined and/or directly dependent on each other. As an example, an infrastructure can have data and samples of the highest quality, but if access is not well-organised their scientific and societal impact will be restricted, lowering the overall value of the infrastructure.

¹⁶ De novo biobanks are typically prospective data and sample collections established only for research purposes.

Which parameters determine the value of each value category?

Each value category consists of a number of value parameters¹⁷. Each of these parameters can be adjusted in different ways to increase the overall value of that category. Table 4 shows the (adjustable) value parameters that make up each value category. This list is a good starting point for a discussion on what determines the value of a sample/data infrastructure, and should be further supplemented in the future. Insight in these parameters will help infrastructures to review and improve their own value towards different stakeholders and will assist stakeholders in estimating the value of a given infrastructure.

Table 4. Value categories with their associated value parameters	
Value category	
Value parameter	Matters that determine to the outcome of this particular parameter
Organisation	
Informed consent	Broad, narrow, or dynamic consent; commercial use; data linkage; opt-in/opt-out; national, European, or international use
Access	Transparency of process; user availability; delivery and service speed; governance; data access committee; user costs; online catalogue; (de)central storage
Standard documents	Material transfer agreement; consent; baseline paper
Response time	Application; assessment; issuance; (de)central storage
Quality system	Accreditation; certification; standard operating procedures; issuance testing; quality assurance personnel
Protocols	Number of participants; study design; frequency of measurements
Networks & consortia	Number, size, and budget of networks & consortia
Participants	
Inclusion criteria	Bias; scientific ambition
Population	Representativeness; number of participants; participant loyalty
Patient	Diagnosis; stage of disease
Events & efficacy endpoints	New diagnoses; hospital admissions
Samples	
Type of biomaterial	Rarity; uniqueness
Sample quality	Completeness log; degree of standardisation; standard operating procedures per biomaterial; pre-analytical sample processing speed; quality control storage and issuance; metadata pre-analytical phase; registration of deviations; links to clinical and other data
Quantity	content has not been discussed
Sample frequency	Before, during, and after event X; amount of resampling; number of measurement time points; longitudinal
FAIR samples	Biobank Information Management System; catalogue; access protocol
Sample storage	Safe and retrievable storage; biobank management system; continuous temperature monitoring; type of storage related to future analyses
Associated costs	content has not been discussed

¹⁷ A parameter is any characteristic that can help in defining or classifying a particular system. That is, a parameter is an element of a system that is useful or critical when identifying the system, or when evaluating its performance, status, etc. (source: <https://en.wikipedia.org/wiki/Parameter>).

Data	
Type of data	Clinical data; real world data; socio-economic data; link to associated samples
Links to other data sources	Personal health train concept; internal data sources; external data sources
Data quality	Complete codebook; accuracy; data origin and history; collection protocol; enrichment; description; completeness
Data frequency	Before, during, and after event X; number of measurement time points; longitudinal
FAIR data	Use of metadata standards; access protocol
Data storage	Validated data management system; sustainable and secure archiving
Associated costs	content has not been discussed
Output	
Scientific publications	Impact of publication, translatability towards clinic
Publications (other)	content has not been discussed
Patents	content has not been discussed
Improved and/or new...	Treatment; diagnostic procedure; (clinical) guideline; decision support system; prevention
New research	National or international collaboration; infrastructures; grants
Services*	
Education and training*	
Brand value*	
Economic value*	
(Re)use*	
*These value categories and their associated value parameters were not discussed during the workshop due to time restrictions. Supplementary table 1 shows both the workshop results and the value categories and value parameters that were created by the authors as starting material.	

Session 3: Stakeholder-value matrix

Which values best match the interests of which stakeholders? And in what way do they match?

In this session the participants aimed to determine which value categories were most valued by each of the identified stakeholder groups, based on their respective interests. Although a decent start was made, the available time was too limited to achieve a comprehensive result. Furthermore, the entered results were sometimes additional stakeholder interests or value parameters. Consequently, we decided to move those additional interests and parameters to their respective tables. We aim to supplement the current stakeholder-value matrix by gathering input from each specific stakeholder group.

Workshop 3: Overarching solutions for sustainability

In this workshop the perspective shifted from the individual sample/data infrastructure to a more general, national perspective. The Netherlands contains a high concentration of sample/data infrastructures, which, to a greater or lesser extent, are faced with sustainability issues. Can we create a national environment that promotes and supports the sustainability of individual

infrastructures? What are potential overarching solutions and required preconditions? To arrive at potential solutions, we followed the different phases of the brainstorm process (problem mapping, idea generation, and idea development), with each session dealing with a subsequent phase. The results will serve as input for future meetings with national stakeholders.

Session 1: Problem mapping

What is the problem?

To come up with the right solutions it is imperative to understand the problem. Therefore, the first phase of the brainstorm process is problem mapping. This means getting consensus on the problem to be solved and listing as many causes and consequences as possible. After getting a clear overview, the group selects the most important and/or solvable causes to address in the next phase.

The starting problem definition was: “infrastructures for sample / data collections cannot realise the necessary continuity”¹⁸. Despite discussions in all three sessions, the problem definition was ultimately unchanged. However, some nuance was added by reframing the problem as a challenge. This challenge forces infrastructures to be critical on what they collect and how they spend their resources, hereby preventing so called “data graveyards”, infrastructures that are maintained but not used.

What are the (most important) causes?

The participants focused mainly on finding potential causes and associated sub-causes. Their results were supplemented with remarks made in other sessions, resulting in an extensive problem map (see figure 1). At the end of the session six main causes were identified and taken to the next session:

1) *Lack of a long-term business model*

This cause is related to: *insufficient thought on long-term sustainability at the start of infrastructures*. A positive example named by the participants was the large European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, which spent a lot of time beforehand on matters related to sustainability. This is now paying dividends, as the EPIC collection is now used for studies on additional diseases, such as dementia and cardiovascular disease.

2) *Sample/data quality is too variable*

The quality of samples and data must be high, continuous and comparable. Variable sample and data quality leads to non-reproducible results, creating uncertainty about the reliability and usefulness of sample/data infrastructures. This problem is evident when companies, for example, attempt to validate original research outcomes on (routine) clinical samples and data of variable, and often lower, quality.

¹⁸ There is a difference between sample and data collections established for a specific delineated research question (studies) and collections established for a broader, more open research goal including future research questions. The stated problem definition is more relevant for the latter group. The former group can often be financed and executed as a project; and terminated after its goals are achieved. These study-specific collections do not necessarily need to think about long-term continuity or sustainability (apart perhaps from FAIR data management), as it is unlikely whether anyone will be interested in using, or even is allowed to use, the leftover material for new research. This does not however mean that valuable collection with a lot of potential should be discarded without careful consideration. And it is still advised for such studies to consider future use and sustainability beforehand.

3) *Reward and focus on new instead of reuse*

In the current system researchers are mostly rewarded –via high impact papers, grants, and/or awards– for working on new and hot subjects, instead of validating previous research and establishing robust results and standards. This does not stimulate the (re)use of already existing sample/data infrastructures and their standards and standard operating procedures, but instead promotes fragmentation by starting ever new collections and new methods. ZonMW, a Dutch funding agency, tries to counteract this by making the reuse of existing infrastructures mandatory for particular grant applications.

4) *External factors change*

This cause relates to the changing environment in which sample/data infrastructures operate. For example, changing political priorities outside or within the own organisation can result in continuity problems. In addition, changes in the scientific field due to advancing knowledge or technologies can render an infrastructure (partly) irrelevant. This creates uncertainty about what should be collected to best serve user needs. As infrastructures require substantial investments and operate on a long timescale, reducing uncertainty and anticipating future changes becomes ever more important. Linking of infrastructures, connecting to additional data sources, and integration in the care process can help expand an initial sample and data collection and increase adaptive capacity.

5) *There are too many sample/data infrastructures*

The Netherlands contains more than 240 sample infrastructures¹⁹ and a large number of data infrastructures. As a consequence, the available infrastructural resources need to be distributed among a large number of infrastructures, resulting in increased competition and potentially lower funding per infrastructure. BBMRI-NL and Health-RI aim to address this cause by linking infrastructures.

6) *Funders mostly have project-based instruments*

As a result, sample/data infrastructures are run as projects, rather than as infrastructures. The corresponding short funding horizon will mean that an infrastructure is only concerned with short-term actions and acquiring the next round of project funding, instead of focussing on its infrastructural tasks and investing in measures that support long-term sustainability. A short funding horizon can also impede commitment and collaboration, as stakeholders are uncertain about the infrastructures future survival. A positive example of an infrastructure outside the life sciences is CERN (The European Organization for Nuclear Research), where users and partners are certain that it will continue to exist for more than 30 years. The Dutch Cancer Society is currently working on an infrastructure-specific call²⁰ to support cancer-related research infrastructures with temporary funding.

Several subjects were discussed at length during the sessions. One of these subjects was the poor image that sample/data infrastructures have with specific stakeholder groups (e.g. hosting institutions, funders, researchers, companies). This is related to a presumed low return on investment (investment versus output in the form of e.g. publications, clinical results). In addition, reproducibility issues experienced in translational research negatively affect the overall image of biomedical research infrastructures. Furthermore, an often heard comment is that the initiators of infrastructures are more interested in collecting samples and/or data, than in using them. Workshop

¹⁹ Source: BBMRI-NL Catalogue, <https://catalogue.bbmri.nl/>

²⁰ <https://www.kwf.nl/english/poi-english/Pages/Infrastructural-initiatives.aspx>

participants identified a poor image as both a cause and a consequence of the main problem, hereby creating a vicious cycle.

Another discussed subject was the lack of incentive for researchers to actively disseminate the collected samples and/or data to others. This contributes to the insufficient use of existing infrastructures. The lack of incentive is partly caused by the current scientific system which (a) promotes competition between researchers on research funding, hereby frustrating collaboration and sharing of resources and (b) rewards only (high-impact) papers and not data sharing. Researchers that collect samples and/or data often spend years building an infrastructure but receive little credit or reward for their efforts. However, subsequent users gain credit for publications based on the gathered samples and/or data, without putting in the effort of collecting them. This results in protective ownership, as keeping the samples and data for oneself allows for more high-impact publications by the collecting researcher, while sharing the data will only result in other researchers getting high-impact publications. What complicates matters is that most researchers that collect samples and data see them as *their* samples and data. This is not the case as they are often collected with public funding. On top of that it is currently impossible in the Netherlands to even exercise property rights when it comes to bodily materials²¹. In the end, this conflict between individual and societal interests needs to be solved to increase (re)use and unlock the full potential of sample/data infrastructures.

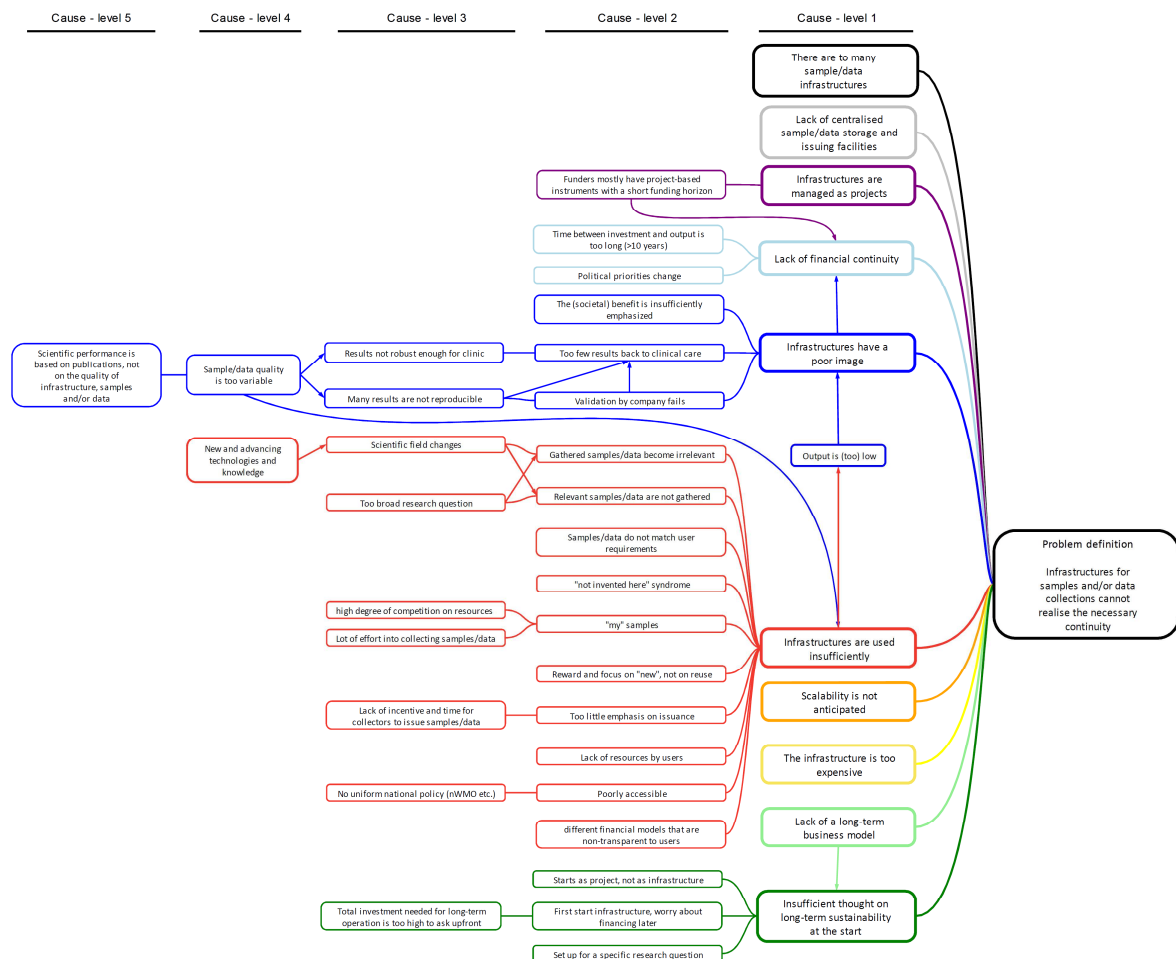


Figure 1. Problem map: problem definition with identified causes and sub-causes

²¹ Aart C. Hendriks and Rachèl E. van Hellemond, *Regulating Privacy and Biobanks in the Netherlands*, Journal of Law, Medicine & Ethics, 2016 Mar;44(1):68-84, <https://doi.org/10.1177/1073110516644200>

Session 2: Idea generation

What are potential solutions for the most important causes?

In this session participants were asked to come up with as many solutions as possible for the six identified main causes. The results are visualised in figure 2. Several of these solutions are already being implemented. For example, BBMRI-ERIC works on improving sample and data quality by standardisation of protocols.

In addition to these solutions, the discussion highlighted the importance of being flexible as an infrastructure by adapting to developments in the field and in the clinic. Furthermore, participants pointed out that stimulating the (re)use of gathered samples and data is critical to justify the investments made. Researchers should be encouraged to distribute samples and make their collections more visible.



Figure 2. Brainstormed solutions for the six identified main causes

Session 3: Idea development

What are the most promising solutions? What preconditions are needed? Which parties should be involved?

In the final session of this workshop the participants selected and elaborated on the most promising solutions (see table 5). Stimulating the issuance and use of samples and data would go a long way in improving the continuity of infrastructures. The participants came up with four approaches to achieve this.

First, give some form of reward or credits to collecting parties for their hard work in establishing the infrastructure and for sharing their samples and data. Building a large, high quality collection is a science in itself and should receive proper recognition. A potential method would be via authorships. This will provide individual incentives for issuance; prevent collectors from keeping hard earned samples and data to themselves; and provides a motivation for investing in a well-used, sustainable infrastructure.

Second, stimulate the use of samples and data by companies. The resulting public-private partnerships are complex and can be a sensitive issue as they involve the use of publicly funded resources by private parties. Therefore, governance models that make sure sufficient benefits flow back to the public domain are required. Transparency is essential to preserve participant trust. Establishing such governance models should be a joint effort between the infrastructure field and (overarching representatives of) private parties.

Third, develop a quality mark or quality standard for sample/data infrastructures, so users, funders, and other stakeholders can determine an infrastructure's quality. It could be that only infrastructures of a certain quality level are represented in the BBMRI-NL Catalogue, assuring public and private users of quality samples and data. Quality monitoring could be organised similar to the medical diagnostics field. Here different diagnostic laboratories send each other samples to be rechecked for accuracy and quality. Developing such a standardised quality mark will however be difficult as it requires consensus on the question: "what is quality?".

Fourth, communicate the value of infrastructures to public and private users, for example via successful use cases. This should also help in improving the overall poor image of sample/data infrastructures, a main cause for continuity problems.

Two additional approaches to stimulate issuance were mentioned in other sessions. The first was increasing the visibility of already collected samples and data, for example by updating the BBMRI-NL Catalogue and further expanding BBMRI-NL Podium. The second was promoting the (re)use of samples and data, for example by increasing the chance for funding when (re)using existing sample/data infrastructures. The Dutch funding agency ZonMW is already implementing the latter approach. This underlines the important role of research funders in creating a sustainable national sample/data infrastructure.

The participants also discussed the potential of establishing a Dutch national biobank. The idea was that merging smaller infrastructures into one national biobank would prevent further fragmentation and simplify standardisation. Furthermore, it would take logistics and management out of the hands of researchers, letting them focus on what they do best. In such a setup, individual infrastructures should not be forced to join, but instead the benefits of joining should be too good to resist. There was no clear consensus if such a solution would be desirable and feasible. If such an initiative would take shape it would need to be in a form that safeguards sample quality, for example via a virtual format with regional sample hubs at the major hospitals.

There was a clear call for a long-term vision from funding agencies and the national government. This would allow infrastructures to position themselves accordingly and invest in appropriate sustainability measures. For funding agencies such a vision should focus on the (re)use of infrastructures and strengthening the current sample/data ecosystem, for example by requiring newly established infrastructures to contact and support the overarching initiatives BBMRI-NL and Health-RI. For the national government such a vision and accompanying leadership should focus on

addressing outstanding policy issues (e.g. nWMO, incidental findings) that hamper infrastructure sustainability.

Table 5. Development of four selected solutions	
1. Solution: Stimulate issuance of samples and data	
How	With whom?
<ul style="list-style-type: none"> Credits and reward for “the dirty work” of collecting and maintaining high quality samples and/or data. Improve the benefits and incentives for sharing samples and data. 	Local biobanks, Association of Universities in the Netherlands, Netherlands Federation of University Medical Centres
<ul style="list-style-type: none"> Governance models for public-private partnerships, where yields flow back to the public domain. 	First BBMRI-NL / Health-RI with individual sample/data collectors and infrastructures, next with private parties
<ul style="list-style-type: none"> A standardised and straightforward quality mark, containing different levels, enabling users and funders to judge infrastructure quality. Quality could be monitored by letting infrastructures test each other’s samples and data. 	BBMRI-NL / Health-RI
<ul style="list-style-type: none"> Communicate successful use cases to show the value of and need for sample/data infrastructures. 	BBMRI-NL / Health-RI
2. Solution: National NL-biobank infrastructure	
How	With whom?
<ul style="list-style-type: none"> Virtual NL-biobank with samples and data collections 	String of Pearls Institute and Durrer Center (already operating at multiple locations), Centralised UMC biobanks
<ul style="list-style-type: none"> NL-biobank with regional sample hubs at major hospitals 	String of Pearls Institute, Centralised UMC biobanks
<ul style="list-style-type: none"> Uniformity in (IT) systems across NL-biobank organisations 	e.g. Netherlands Society for Clinical Chemistry and Laboratory Medicine, String of Pearls Institute, Durrer Center, suppliers
3. Solution: Long-term national vision on sample/data infrastructures	
How	With whom?
<ul style="list-style-type: none"> Such a national vision provides a starting point for jointly resolving outstanding policy issues (e.g. nWMO, incidental findings) and work towards lasting, overarching solutions that support sustainability of infrastructures. The field should work on joint lobby and draft proposals. 	e.g. Ministry of Health, Welfare, and Sport; Ministry of Economic Affairs and Climate Policy; Ministry of Education, Culture, and Science; Netherlands Organisation for Scientific Research; Association of Universities in the Netherlands; Netherlands Federation of University Medical Centres
4. Solution: Long-term vision of research funders	
How	With whom?
<ul style="list-style-type: none"> Follow up on vision with clear actions. 	Funding agencies, charities, government
<ul style="list-style-type: none"> Forward each (new) research project that is related to sample/data infrastructures to BBMRI-NL / Health-RI. 	Funding agencies, Netherlands Federation of University Medical Centres, Association of Universities in the Netherlands
<ul style="list-style-type: none"> Reserve part of allocated funding for individual sample/data infrastructures for funding of BBMRI-NL / Health-RI as overarching initiatives. 	Funding agencies, Netherlands Federation of University Medical Centres, Association of Universities in the Netherlands
<ul style="list-style-type: none"> Include a mandatory section in research proposals where researchers explain how they plan to use, interact, or collaborate with existing infrastructures. 	Funding agencies, charities, government

Discussion

The workshop groups consisted of participants from different types of sample/data infrastructures. This provided a mix of knowledge and experience; enriching the discussions and allowing the participants to learn from each other. The participants were however representatives of one specific stakeholder group, creating a one-sided perspective. The results should therefore be interpreted with that in mind, as there is a chance for assumptions and bias. To correct for this the results should be verified and supplemented by other stakeholder groups.

What stood out during the workshops was the focus on sample and data quality. A prevalent thought was that a high-quality infrastructure with great samples and data would result in a high number of users; build it and they will come. However, although high quality is the major requirement for attracting users, it is not a guarantee. Quality just increases the odds of them coming. The odds can be further increased by adopting a business-oriented approach and starting a conversation with intended users from an early stage. In this way sample/data infrastructures make sure that they build the high quality that their intended users need.

Another point raised during the workshops was how infrastructures should handle scientific hypes. Currently there is a high demand for microbiome samples, peripheral blood mononuclear cells, and organoids. However, the handling costs for such samples are high and, because there is a lack of standardised and validated protocols, quality is questionable. As a result, collecting such samples poses a risk to the infrastructure in terms of reputation and resources. What speaks in favour is that the high demand will mean that the samples and data are likely to be used.

The discussions during the workshop showed that each type of infrastructure has its own specific challenges, views, and interests. This stems from their different roles, responsibilities, and organisational setups. The biggest differences were between biobanks and infrastructures that focus solely on data. In the end, each subset of infrastructures will require tailor-made solutions to address their specific sustainability challenges. However, there is likely more that connects than separates the different infrastructures; and as such the field should focus on finding joint solutions that benefit most infrastructures.

We can conclude that the meeting was a success given that twenty-one participants from nineteen organisations met to discuss matters related to sustainability. It was even more a success given the rich yield of information and shared experiences; benefitting both individual participants and the field. The results provide a promising starting point for next steps that focus on verifying and supplementing the outcomes among the various stakeholders; further developing the stakeholder-value matrix; and using the problem map and developed solutions to start a dialogue with relevant stakeholders on overarching solutions, while keeping in mind the diversity among infrastructures. A multiparty approach is essential to achieve lasting results and work towards a future-proof Dutch sample and data infrastructure.

Acknowledgements

We want to thank all the participants who invested their time and effort in this meeting and in providing feedback on this subsequent report. Furthermore, we want to thank Rick van Nuland from Health-RI/Lygarture for his support and input, and his chairmanship. We are grateful for the organisational support provided by Léone Böck and Erna Erdtsieck-Ernste, both from BBMRI-NL/Lygarture.

Supplementary information

Supplementary table 1. Value categories with their associated value parameters

This table contains both the content discussed during the workshop and the content which was not discussed due to time restrictions. This non-discussed content was constructed beforehand by the authors to serve as starting material during the workshop. The non-discussed categories/parameters include services, education and training, brand value, economic value, and (re)use.

Organisation	
Informed consent	Broad, narrow, or dynamic consent; commercial use; data linkage; opt-in/opt-out; national, European, or international use
Access	Transparency of process; user availability; delivery and service speed; governance; data access committee; user costs; online catalogue; (de)central storage
Standard documents	Material transfer agreement; consent; baseline paper
Response time	Application; assessment; issuance; (de)central storage
Quality system	Accreditation; certification; standard operating procedures; issuance testing; quality assurance personnel
Protocols	Number of participants; study design; frequency of measurements
Networks & consortia	Number, size, and budget of networks & consortia
Participants	
Inclusion criteria	Bias; scientific ambition
Population	Representativeness; number of participants; participant loyalty
Patient	Diagnosis; stage of disease
Events & efficacy endpoints	New diagnoses; hospital admissions
Samples	
Type of biomaterial	Rarity; uniqueness
Sample quality	Completeness log; degree of standardisation; standard operating procedures per biomaterial; pre-analytical sample processing speed; quality control storage and issuance; metadata pre-analytical phase; registration of deviations; links to clinical and other data
Quantity	content has not been discussed
Sample frequency	Before, during, and after event X; amount of resampling; number of measurement time points; longitudinal
FAIR samples	Biobank Information Management System; catalogue; access protocol
Sample storage	Safe and retrievable storage; biobank management system; continuous temperature monitoring; type of storage related to future analyses
Associated costs	content has not been discussed
Data	
Type of data	Clinical data; real world data; socio-economic data; link to associated samples
Links to other data sources	Personal health train concept; internal data sources; external data sources
Data quality	Complete codebook; accuracy; data origin and history; collection protocol; enrichment; description; completeness
Data frequency	Before, during, and after event X; number of measurement time points; longitudinal
FAIR data	Use of metadata standards; access protocol
Data storage	Validated data management system; sustainable and secure archiving
Associated costs	content has not been discussed

Output	
Scientific publications	Impact of publication, translatability towards clinic
Publications (other)	content has not been discussed
Patents	content has not been discussed
Improved and/or new...	Treatment; diagnostic procedure; (clinical) guideline; decision support system; prevention
New research	National or international collaboration; infrastructures; grants
Services	
Consultancy	content has not been discussed
Additional analysis	content has not been discussed
Storage, management and issuance	content has not been discussed
Medical check-up	content has not been discussed
Training (for third parties)	content has not been discussed
Education and training	
Training of personnel	content has not been discussed
Internships and students	content has not been discussed
Education of users	content has not been discussed
Brand value	
Communication resources	content has not been discussed
Network	content has not been discussed
Exposure	content has not been discussed
Brand awareness	content has not been discussed
Economic value	
Number of jobs/FTE	content has not been discussed
Investment multiplier	content has not been discussed
Return on investment	content has not been discussed
Number of spin-offs	content has not been discussed
(Re)use	
Number of issued samples/data	content has not been discussed
Number of customers	content has not been discussed
% returning customers	content has not been discussed

Supplementary table 2. List of workshop participants			
Participant	Sample/data infrastructure	Organisation	Group
Dr. Rick van Nuland*	Health-RI	Lygature	
Mrs. Claudia Benschop	Data Interstitial Lung Disease Biobank	St. Antonius Ziekenhuis	3
Dr. Henk van Kranen	Doetinchem Study	National institute for public health and the environment	1
Dr. Wanda Hermans- van Ast	Durrer Center Biobank	Netherlands Heart Institute	1
Dr. Rob Hoofst	ELIXIR-NL	Dutch Techcentre for Life Sciences	1
Dr. Peter (P.H.J.) Riegman	Erasmus MC Tissue Bank	Erasmus Medical Center	1
Dr. Carine (C.J.M.) van der Vleuten	Hecovan Biobank	Radboud University Medical Center	1
Dr. Aafje Dotinga	Lifelines	Lifelines	1
Prof.dr.ir. Hein (H.W.) Verspaget	LUMC Biobank	Leiden University Medical Center	3
Dr. Carla (C.J.H.) van der Kallen	Maastricht Study	Maastricht University Medical Center	1
Dr. Nadine (N.S.M.) Offermans	Maastricht Study	Maastricht University	2
Mrs. Marion (A.H.) Feijge	Maastricht Study	Maastricht University	3
Dr. Annegien Broeks	Molecular Pathology & Biobanking Core facility	Netherlands Cancer Institute	2
Prof. dr. Dorret (D.I.) Boomsma	Netherlands Twin Register	VU University Amsterdam	3
Dr. Annette (A.H.) Gijsbers-Bruggink	PALGA, Network and registry of histo- and cytopathology in the Netherlands	PALGA, Network and registry of histo- and cytopathology in the Netherlands	2
Dr. Elena (A.N.) Shumskaya	Nijmegen Brain Imaging Genetics / Cognomics	Radboud University Medical Center	2
Mrs. Janny Nagel	OncoLifeS, Oncological Life Study	University Medical Center Groningen	2
Dr. Geraldine (G.R.) Vink	Prospective Dutch colorectal cancer cohort	Netherlands Comprehensive Cancer Organisation	3
Dr. Peggy Manders	Radboud Biobank	Radboud University Medical Center	3
Dr. Johan (J.W.N.) Lagerberg	Sanquin Bloodbank	Sanquin	3
Drs. Gerard van Grootheest	The Netherlands Study of Depression and Anxiety	GGZ inGeest; Amsterdam UMC	3
Dr. Robert (R.A.) Verheij	Nivel Zorgregistraties eerste lijn	Nivel, Netherlands institute for health services research	2
* Dr. Rick van Nuland participated in all sessions of workshop 3 to serve as linking pin towards Health-RI and for future developments and meetings.			