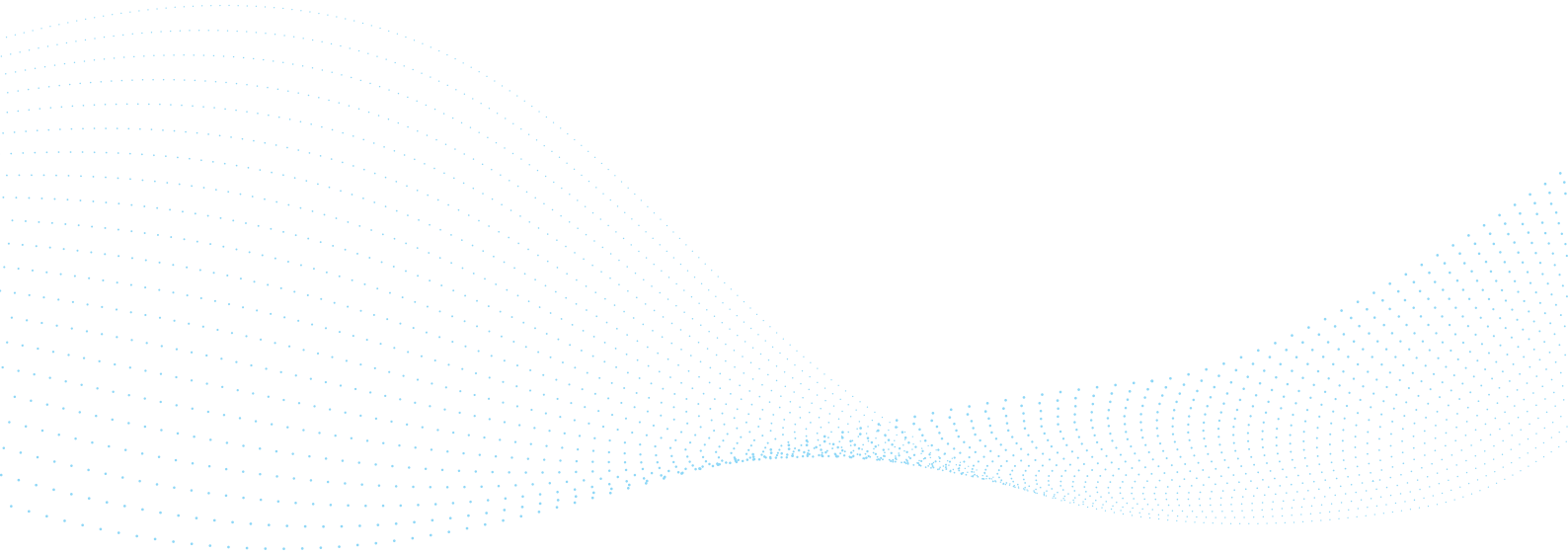
DNV

WHEN TRUST MATTERS

CHALLENGES, NEEDS AND OPPORTUNITIES IN FEDERATED HEALTH DATA NETWORKS

Summary of workshop hosted by DNV
April 2022

WHITEPAPER



Disclaimer

All information is correct to the best of our knowledge. Contributions by external workshop participants do not necessarily reflect the views of the editors and DNV.

Correspondance to: Sharmini.Alagaratnam@dnv.com
White Paper, DNV (2022).
Design and layout: Tress Design AS

Content

1. Executive summary

Given the challenges of sharing sensitive health data, federated health data networks (FHDNs) have emerged as an attractive alternative to pooled data storage and are seen as an enabler for clinical and population studies by facilitating decentralised data processing. However, the establishment and operation of such networks is not straightforward.

In order to identify the challenges, needs and opportunities in FHDNs, the Healthcare Programme from Group Research and Development, DNV hosted a workshop with participants representing eight different stakeholders operating within the FHDN domain. The work-shop findings revealed eight main challenges ranked as follows in order of importance or shared importance:
(I) trust, (II) infrastructure interoperability, (III) sustainability, (III) legal and regulatory, (IV) data interoperability, (V) network establishment, (V) scalability and (V) incentives. Potential approaches for addressing these challenges, and ideal requirements at each of the establishment, operation and expansion phases of a FHDN were discussed.

To deliver on the value of federated networks for efficient access to sensitive health data for all stakeholders, trust was identified as integral, both across the challenges and at the different phases. Importantly, the requirements from one challenge may have consequences on another challenge, and as such these may frequently not be solvable in isolation from each other.

2. Introduction

Health data are personal, highly sensitive, and subject to strict regulations and privacy rights. Making these data accessible for wider use can therefore be challenging, whether in primary healthcare, or in secondary use for research.

Due to the complex collection of national laws and shared EU frameworks that regulate the collection and processing of health data, collaboration between institutions in different jurisdictions is difficult. For example, the EU's General Data Protection Regulation (GDPR), despite having a primary purpose to provide standardised data protection laws, may be integrated and interpreted slightly differently into national legislation across jurisdictions. Clinicians, legal scholars, and researchers thus have different interpretations of the legality of providing access to patient data for primary or secondary analyses. For collaboration outside EU data sharing can be even more challenging.

Additionally, the lack of harmonised collection and curation of health data and governance and management frameworks have in many medical domains resulted in fragmented, unstructured and siloed data (1). However, in the Northern European countries, the creation of health registries such as cancer registries have attempted to counter these problems on a national level. Local and regional healthcare organisations and institutions have few incentives to share and make their data accessible, as this is a resource-heavy task that needs prioritisation from top management who already struggle with other more pressing challenges, such as a lack of resources and relevant competencies and a necessity to prioritise more tangible clinical needs. In addition, data protection and IT system security are complex to manage. The benefits of data access to enable transparency and improved clinical outcome for current patients from a specific hospital may not be immediately evident, instead longer term, less concrete value may come from incorporating this data for research purposes and diagnostics of future patients (2). As discussed above, legal barriers to moving health data from their site of production are particularly hard to overcome (3). In addition, discussions around IP rights, use of different software and different prioritisation of investment and running costs can add additional layers of complexity to the picture (4–6).

The global drive to improve health outcomes through digitalisation and data can only be achieved with improved access to data across institutions and jurisdictions. As an example, the implementation of artificial intelligence tools such as federated learning will require data of sufficient quality and volume to develop trustworthy and robust models. The urgency of overcoming these challenges for a rapidly evolving healthcare sector is demonstrated in the areas of cancer and rare diseases, where Federated Health Data Networks (FHDNs) have been suggested as a potential solution (see box for the hallmarks of a federated network, FN (7)). For example the Federated Tumour Segmentation initiative (8) and Beyond 1 million genomes project (9), combine previously isolated data to allow new understanding of disease trajectories and the biology underlying the disease, to ultimately guide and tailor clinical management. FHDNs have also been developed to answer needs in the COVID-19 pandemic, both for disease detection (10) and predicting clinical outcomes (11).

Multiple platform providers and research projects are focusing on this approach as a sustainable solution to enable research and innovation and eventually improve clinical care. To support this approach, the Healthcare Research Programme at DNV, through its mandate of exploring new assurance roles in healthcare, hosted a workshop with stakeholders operating within the FHDN domain to identify needs, challenges and approaches that they have encountered during establishment, operation and expansion phases. This collaborative approach aims to

FNs can be said to share the following common characteristics (7):

- Each node (partner with data) is semi-autonomous as they can make their own decisions on granting data access, however nodes are governed by a common framework agreed upon by all nodes.
- FNs are supported by a common infrastructure with harmonised interoperability standards and tools.
- Each node requires local computing capabilities to enable querying or processing to be performed locally.

develop a better understanding of how a 3rd party role could support all phases of a FHDN, across the ecosystem on both technical and non-technical fronts.

FHDNs can facilitate access to sensitive health data and have the potential to enable large cohort analysis across healthcare institutions and jurisdictions. FHDNs conform to the idea that where the data is located is where it should be used, and the concept is being developed for several health data platforms, including the European Medicine Agency's Data Analysis and Real World Interrogation Network (DARWIN EU®) (12). While the World Economic Forum has published an eight-step guide for building a FN consortium (13), its relative novelty means that the FHDNs established to date are mainly in the research domain, and few previous papers have explicitly addressed their operationalisation and expansion (14–16).

3. The workshop approach

A three-hour digital workshop was organised by the Healthcare Programme from Group Research and Development, DNV on 7th December 2021, bringing together global FHDN pioneers at the forefront of establishing and operating FHDNs. This workshop aimed to understand common challenges, opportunities and needs in establishing, operating, and expanding FHDNs.

Participants were selected and invited following a FHDN landscape mapping exercise, where relevant descriptions, experience, use cases and infrastructures were captured. The workshop invitation, design and delivery was a collaborative process. Prior to the workshop, participants were asked to prepare a scene setting introductory slide to their FHDN approach, highlighting their background, infrastructure, key activities, use cases, and partners for sharing and presenting in lightning fashion, at the workshop.

This workshop created a meeting place for these FHDN pioneers to share their experiences, identify common challenges and needs in establishing and operating FHDNs, and discuss potential solutions that respond to these needs. The workshop participants represented different categories of FHDN stakeholders such as i) platform providers (both commercial and academic: Owkin, Vantage6, DataSHIELD, MedCO, Medical Informatics Platform); ii) European Commission (EC) funded projects (EUCANCan); iii) national cancer registries and institutions (Cancer Registry of Norway, Netherlands Comprehensive Cancer Organisation); and iv) an independent third-party assurance and risk management company (DNV). Brief details about the backgrounds, data types and infrastructures of these FHDNs can be found in Table 1.

Table 1: Platforms and institutions involved in the DNV hosted workshop on needs and opportunities in FHDNs

Workshop participant	Background	Data types (if relevant)	Commercial or academic	Operation
DNV	Founded in 1864, operating as a global independent third-party assurance and risk management company across multiple sectors. Developed a FHDN sandbox in collaboration with the Norwegian Cancer Registry.		Commercial	Operating globally. Healthcare Programme from Group Research and Development exploring enabling trust in the digitalisation of healthcare through third-party role in FHDNs.
DataSHIELD	Initiated in 2007, now offering a mature open-source platform with real time automated disclosure protection and over 100 statistical functions available on different (yet harmonised) data types.		Academic	Operating internationally within 14 research consortia.
Medical Informatics Platform	Developed to support the Human Brain Project in 2013, to link decentralised anonymised and harmonised data.	Data from brain science, clinical research, and patient care.	Both	Operating internationally with three active federations (and three under development) across 34 nodes.

Workshop participant	Background	Data types (if relevant)	Commercial or academic	Operation
EUCANCan	Originating from an EU Horizon 2020 funded project beginning in 2019.	Cancer genomics datasets.	Academic	Operating internationally between Europe and Canada with 3–5 nodes.
MedCo	Co-developed by EPFL and CHUV since 2017: operational system for secure exploration and analysis on distributed sensitive data. Supported since January 2022 by Tune Insight.	Main first application in oncology.	Both	Operating at three Swiss university hospitals, with nodes in US, and Italy.
Cancer Registry of Norway	Experimented with different FN technologies e.g., Sharemind, Vantage6 and DeCanSec (17) to enable privacy preserving statistical computation.	Epidemiologi-cal cancer data sets.	Academic	Operating between Europe and Nordic Cancer registries with at least one node per registry.
Owkin	Founded in 2016, offering its first HC application in 2018, with the goal to accelerate research along the whole drug discovery process.	Multiomics, histology and clinical data with small molecule assay data.	Both	Operating internationally across 20 clinical nodes (hospitals and SMEs) and 10 industrial pharma nodes (AWS) plus an aggregator.
Vantage6	Initiated in 2018, offering a privacy-preserving open-source platform based on user autonomy, heterogeneity and flexibility, particularly in the health/cancer data landscape.		Both	Operating internation-ally through technology development in NL, ad-hoc research projects and infrastructure across Europe and collaborations with other FHDNs (e.g., MedCO, DataSHIELD)

The digital workshop made use of the whiteboard software Miro. Following introductions to participants' FHDN approaches, participants were asked to use virtual post-it notes to list the challenges that they had encountered when establishing, operating and expanding a FHDN. The challenges were first clustered by plenary discussion, then collaboratively ranked by importance. Participants were then divided into three breakout rooms and six of these eight challenge clusters, and potential approaches to overcoming them, were discussed in more detail (as time allowed). The workshop concluded with a short session gathering the opinions of participants about the properties, needs and opportunities at the establishment, operation, and expansion phase of an ideal FHDN.

4. Challenges and opportunities identified in the workshop

Eight main challenges were identified by workshop participants when establishing, operating and expanding a FHDN, and were ranked as follows in order of importance or shared importance: (I) trust, (II) infra-structure interoperability, (III) sustainability, (III) legal and regulatory, (IV) data interoperability, (V) network establishment, (V) scalability and (V) incentives (Figure 1). This section of the paper describes each challenge in more detail both from the post-its and the breakout room discussions and includes brief descriptions of potential approaches proposed by the workshop participants.

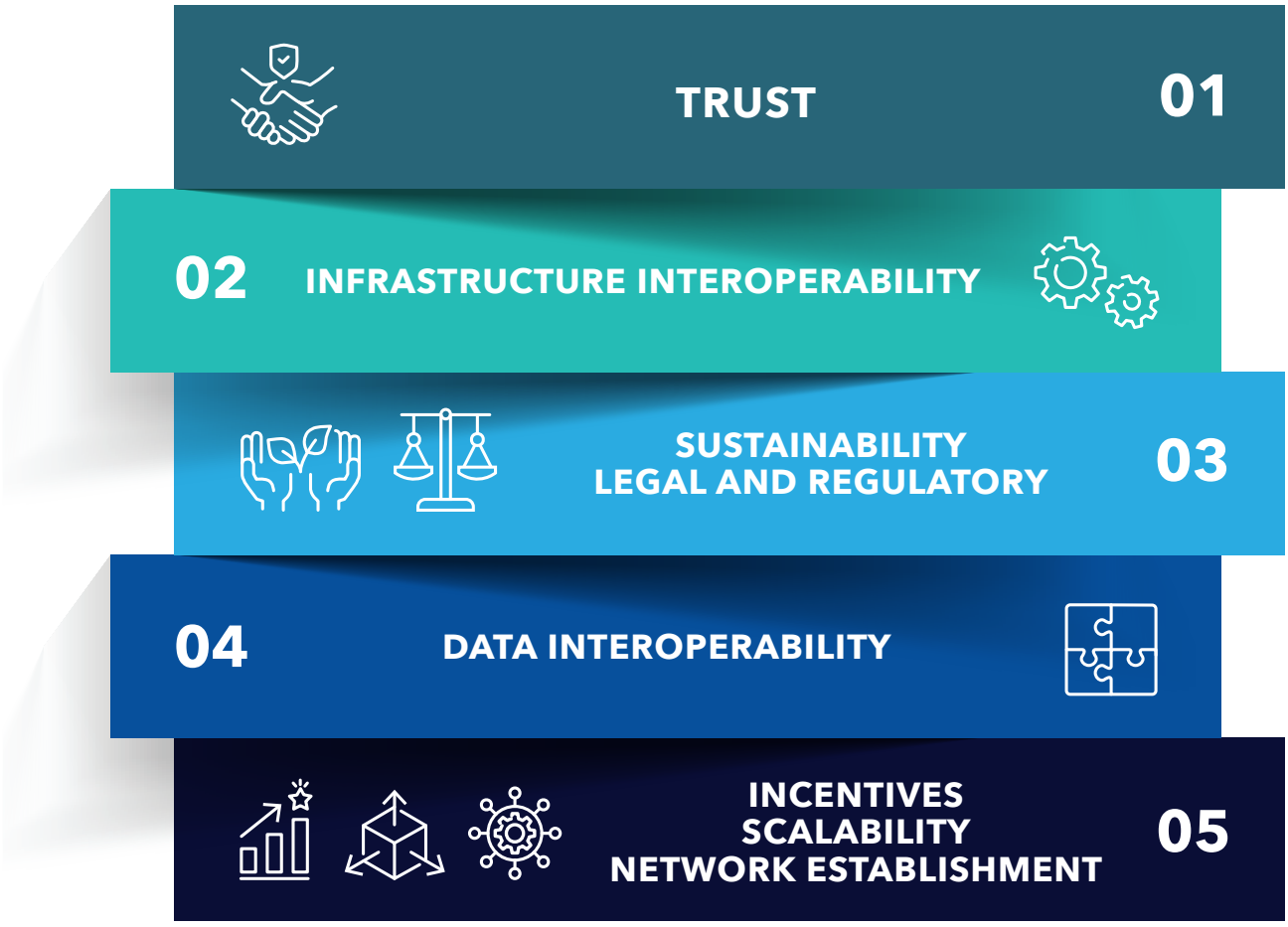


Figure 1: Challenges encountered when establishing, operating, and expanding a FHDN ranked in order of importance

4.1 Trust

Trust in the context of FHDNs includes technical (security and privacy), statistical (collection, organisation, analysis, interpretation and presentation of data), and non-technical (risk and assurance) aspects. Common to all these, however, is the challenge of determining what level of trust is needed for each aspect, and how this can effectively be built, maintained, and demonstrated to and by each FHDN partner.

FHDNs also need strategies for building trust to accommodate different types of partners, e.g., research, academic and commercial, some of whom may traditionally be competitors, or cross-disciplinary partners that may metaphorically or literally speak different languages. Different levels of trust are required between different partners in a FHDN. A balance between incentives to collaborate and competition is needed, and an emphasis on reinforcing trust should be considered when translating technical details in a comprehensible way to those with a different knowledge base.

Potential approaches

Approaches to create trust should cater to the technical, statistical, and non-technical needs of the different partners. What reinforces trust for one partner can be less or ineffective for another, so a multi-faceted approach is appropriate. This could include robust testing and evidence of measures to garner trust, such as comprehensive and documented governance framework for the FHDN. Governance can be considered as both high-level data asset management and stewardship, and formulation of processes for ensuring compliance, quality, and security across an ecosystem. It can include considerations from consortium agreements or other legal framework agreements, stakeholder mapping, identification of regulations and best practices, data architecture, quality and processes, role designations, (continuous) risk management, distributed ledger technology for operational traceability, and a strategy for including new and/or unknown users. DNV, as an independent third-party, can enable this trust through its broad expertise in governance, cybersecurity, assurance, risk assessment and data quality, coming from service offerings in digital solution domains and explorations in international clinical projects.

When legislation is familiar e.g., the GDPR within the EU, successful incorporation of a FHDN could provide the incentive and know-how to promote expansion and uptake of FNs in other health domains. Additionally, previously established 'offline trust', in this case the propensity to trust in partners prior to and as an aside to the FHDN, may provide a good basis for establishing and operating a FHDN, helping to mitigate and prevent technical challenges (and consequent trust issues).

Distributed ledger technology was also touted as a potential approach for helping to increase transparency in the infrastructure; providing visibility to interactions through the logging of operations and queries, and helping clinicians and researchers achieve credit for their efforts. Even though this technology might be premature in the healthcare sector, and lack incentives or knowledge about positive healthcare uptake, it is seen as a promising technology to enable trustworthy data processing and management.

4.2 Infrastructure interoperability

There are multiple aspects around infrastructure interoperability which must be taken into account when considering the establishment, operation and expansion of a FHDN. Considerations include hardware, software, network, institutional affiliation and geographical node, each with their own specific challenges. The different partners (e.g., clinics, health registries, research centres, data repositories and companies) will have different pre-existing infrastructure solutions, needs (e.g. conventional central storage and/or cloud-based approaches) and resources, resulting in heterogeneous starting points for establishing a FHDN.

Challenges at the establishment phase may be more hardware and institutional focused due to the need to be compatible with local institutional architecture, whilst at the operation phase the challenges may be more software, hardware and network related. The infrastructure provider must be prepared to adapt and adopt frequent and potentially irregular updates, which can be problematic within clinical environments where external software and federated data access is actively de-prioritised, due in part to a lack of resources, relevant competencies, understanding of legal implications, unclear authorisation/decision making structure and prioritisation of clinical needs.

Additionally, discrepancies exist and arise between user needs and network properties offered and developed, which are not always adequately conveyed and addressed. Communication and follow-up between developers, infrastructure management and users can be fragmented, with poor follow-up of requests and difficulties in contacting the responsible partner.

Potential approaches

There is a need to make FHDN platforms interoperable with potential partners' infrastructure, and both top-down and bottom-up approaches can be considered to achieve this, depending on the starting point for formation of the FHDN. A top-down approach focuses on standards and software that are community developed and harmonised. However, for successful establishment of a FHDN the top-down approach still requires alignment, and subsequent avoidance of incorrect protocol assumptions relating to a partners' network properties and deployment environments. A bottom-up approach targets interoperability efforts at local institutional level first, through the creation of more common and flexible APIs across competitive platforms. Federation on top of federation (i.e., federating platforms) and decoupling of the data and user interfaces from proprietary federated engines and algorithms, was also touted as a potential approach to enable retention, interoperability and use of platforms and data on a global scale. Value could also be gained from the integration of a single, dedicated channel for communication with users for co-ordination of the FHDN.

4.3 Sustainability

The majority of FHDNs to date have been established through externally funded research projects to drive collaboration between partners, and to demonstrate proof of concept prior to scaling and potential commercialisation. These FHDNs often operate in 'closed' environments that connect silos of researchers, and as such effective sustainability plans, i.e. those related to long-term resilience, system longevity, durability, adaptability, and energy/resource consumption (18) are often not prioritised in terms of funding and strategy. Additionally, partners in a FHDN have different needs, where sustainability across the network and infrastructure will invariably be interpreted differently. For example, some users may want only to consume data, others may be concerned with data versioning issues, and yet others may want to declare IP/patent rights to restrict data. Strategies for sustainability can be developed and implemented later; however, they are not trivial, and can frequently be de-prioritised as other pressing operational challenges arise.

Potential approaches

FHDNs should ideally map sustainability needs already when establishing the network. One way to target sustainability, and ensure it remains a priority is to assign responsibility to one specific partner, who could ensure management of operations, users and expectations, along with other sustainability necessities. Distributed ledger technology could be an additional approach to ensuring sustainability in the FHDN , especially in terms of operation management (compliance checking, data standardisation, AI integration, transparency) (19) and data minimisation (encryption, masking and access delegation) (20).

FHDN sustainability can also be achieved by ensuring processes exist to enable all nodes to undergo harmonised maintenance across users as required over time. In this way, out-of-date/incompatible software versions become obso-lete and network usability is maintained.

The success of a FHDN rests on building and supporting a community that benefits all partners in the ecosystem, potentially through incentivisation and rewards where required. In order to offer a return on investment within a short time frame, the value of a FHDN must be demonstrated and conveyed early on.

4.4 Legal and regulatory

The complex interplay of actors, technology, data ownership and knowledge transfer in FHDNs can make interpretation of legislation and regulatory requirements difficult. For FHDNs a gap must be bridged between technological capabilities and legal clarity. Different partners involved in FHDNs, such as developers, administrators, clinicians, researchers, and legal scholars, have different knowledge bases. As such, they may place varying levels of importance on the same issue, and in some cases, may lack the competency to fully understand the issues at hand. A specific example discussed during the workshop was the question: *How can informed decisions relating to state-of-the-art technologies such as privacy-enhancing technologies be evaluated by legal advisors who may not have the necessary competence to do so?*

In addition to this, local legal and regulatory policies are known to differ due to complex internal risk pictures, enforcing limitations on what can be achieved consistently within institutions of the same jurisdiction. These inherent resulting differences are even more pronounced between jurisdictions, which has large implications for international FHDNs.

Potential approaches

Legal scholars and data protection officers must be educated to make decisions based on an understanding of the technology, and conversely developers and users must understand the relevant limitations determined by the law. This knowledge exchange will benefit both parties and ensure confidence that the establishment, operation and expansion of a FHDN, its technology, and the use of the data adheres to legal and regulatory policies. Continuous engagement of partners, communication with regards to the topics of privacy and security, and alignment with local and national legal requirements are as important as ensuring that partners preserve autonomy and control at their node.

4.5 Data interoperability

The curation and standardisation of currently available and future data sets to ensure interoperability can be a challenge, even across departments within the same institution. Decisions and consensus for optimising data interoperability should be made between partners within the FHDN, however, due to limitations in assigning resources this is not always achieved.

Although FHDN nodes generally maintain responsibility and ownership for their data, they are not always the users of the data. As such, nodes may feel less responsibility for aligning curation and quality of their data with the needs for data use at other nodes if it requires a significant investment of resources. Continuous assessment and assigning responsibility for who, how and when the data quality is checked for use within the FHDN may be a prerequisite for ensuring ongoing value from the data.

Potential approaches

Decentralised data requires more investment to achieve interoperability, and the necessity to meet this investment need generally remains the responsibility of the node, although involvement and investment of other partners can aid this process. As such, the agreement of harmonised standards should ideally occur as early in the establishment of the FHDN as possible. Networks and data formats ideally should learn from common models and best practices of existing international and interdisciplinary collaborations involving large scale health data, where Informatics for Integrating Biology & Bedside (i2b2) (21), FAIR guiding principles (22) and the Observational Health Data Sciences and Informatics (OHDSI) (23) were named as examples.

Data quality and provenance can be achieved through integration of data catalogues, or organised inventories of data assets within each node, which promote understanding about each partners' data. Data quality could additionally be assured prior to the data being made available, and throughout the data pipeline, by implementing scripts and libraries.

4.6 Network establishment

For FHDNs, the basis must be built not just on shared goals, but also on trust, agreements, mutual benefits, communication and enthusiasm, all of which can facilitate initial network establishment and an environment that encourages the expansion of the network. Insufficient management or maintenance of communication between current and/or future partners may result in misalignment and ineffective coordination of the needs of users and developers. The benefits for partners joining or establishing a network might not be clearly demonstrated and therefore adoption by all relevant users within an institution (and not just the early technology adopters) is not prioritised. Additionally, discussions dwelling on determining assignment of intellectual property may hamper data access and resulting output.

Specific issues raised during the workshop included: *How can enthusiasm and communication from and between different partners be sufficiently maintained? How can a community of users and developers be coordinated to address effective network establishment?* Assigning the time and resources to address these and similar challenges was noted to be difficult.

Potential approaches

Potential partners and nodes may not be at the same phase of maturity, as such appropriate guidance and expertise may not be readily available. One way to address this could be to collate the different needs that have historically arisen, and the corresponding strategies used to address these from other FHDN nodes and share these with relevant partners.

An alternative but less inclusive approach could be to impose minimum joining criteria for partnership within the FHDN, only integrating partners with, for example, sufficiently mature technology and data interoperability. If this puts too many limitations on potential partners; one way to facilitate institutions with a lower maturity to join a FHDN could be through the building of pilot studies in small groups within an institution that then help expanding the infrastructure.

Regardless of approach, workshop participants unanimously agreed that data governance should remain locally managed.

The opportunities and benefits of the network should be clearly communicated to existing and future partners, this is covered in more detail in **4.8 Incentives.**

4.7 Scalability

Due to most FHDNs being established within research settings, the maintenance and scaling of FHDNs beyond the duration of the project is often poorly considered or addressed. Ensuring that the network developed is scalable and accessible to future partners presents opportunities for continued growth and knowledge generation, which prevents the FHDN from paradoxically turning into a data network silo itself. Furthermore, FHDNs designed within a specific healthcare domain may have a perceived incompatibility between the niche that it was developed in, and the environments it could be scaled into.

Potential approaches

Strategies should be incorporated during the establishment and operation phase of the FHDN to enable it to be scalable, including considerations integrated to lower the activation barrier to acceptance of new nodes. This could be through exploration with potential new partners to assess their needs and requirements and ensuring the resultant FHDN developed will be suitable for multiple purposes and use at different institutions.

Partners need to have resources assigned to support both technical and non-technical scalability issues. This could be in the form of clear decision-making processes, qualified IT personnel, funding for the acquisition and integration of required hardware and software, and training programs for users. Learnings and best practices from other institutions can be shared across disciplines.

Additionally, to effectively scale access to health data across healthcare institutions, research projects and jurisdictions, and to capitalise on the potential arising from the multiple FHDNs being established, it is also necessary that platform-platform federation between different FHDNs are encouraged and facilitated.

Alignment of the FHDN with relevant community standards and global initiatives, e.g., the European Health Data Space (24), increases its relevance for collaborative scaling. Communication and coordination between partners (e.g., pharmaceutical companies and healthcare institutions) on resources for furthering expansion and subsequent knowledge development could also be a catalyst to scalability.

4.8 Incentives

Incentives as a challenge cluster was ranked 5th in importance (alongside network establishment and scalability) in the workshop, however, its importance should not be underestimated. It has recurringly been named equally important for considering approaches that will increase the trust, sustainability, network establishment and scalability of a FHDN. Incentivisation has a key role to play in the establishment, operation and expansion of a FHDN, both at a node, with users, and with potential partners.

There are a number of factors that affect the nodes' or institutions' willingness to share, including fears about loss of control, data misuse and competing analysis, unclear credit and recognition strategies, differing principles and values, and a lack of resources (25). As such, some potential nodes may be possessive of their data sets and have limited interest in advocating for or accepting establishment of a FHDN that would effectively force them to share their data. Conventionally, pooled data storage which enables data access exclusively with selected collaborators is an easier route to recognition and publication.

Effective strategies to incentivise FHDNs with potential partners where resources are limited, assigned to higher priority considerations, or generally perceived as creating a conflict of interest may be inherently unfeasible. Methods to incentivise novel data access technologies are not prioritised and are time-intensive, and the following question remains: How can partners along the value chain be incentivised to change their status quo?

Potential approaches

Resources must be made available to advocate for and support the establishment and use of FHDNs. Convincing and clear explanations regarding the problems the FHDN is able to solve, and the opportunities and benefits it can bring, may be repeatedly required to existing and future partners, and pitched at different expertise levels targeting users and partners across IT, researchers and clinicians, as well as key decision makers. The incentive or reward for contributing data to a FHDN may be aligned with the advantage, be it recognition, access to additional data or commercial, that FHDN participation is expected to bring.

Additionally, once the FHDN is successfully established, strategies must shift to motivate, involve, encourage, and incentivise relevant users to adapt to using the new system.

5. The ideal federated health data network

In the final exercise of the workshop, participants were asked to reflect on their knowledge and experience of FHDNs, and answer the question: What would you need to do in the different phases of establishing, operating and expanding an ideal FHDN?

Comments were clustered by theme and summarised for the three phases of FHDNs, as shown in Figure 2, and described in detail in Table 2.

The critical topics for discussions were science, privacy, valorisation, intellectual property and sustainability, while the FHDN technology itself was seen more as a means to an end. Sufficient research and innovation resources should be made available to reach the establishment, operation, and expansion goals.

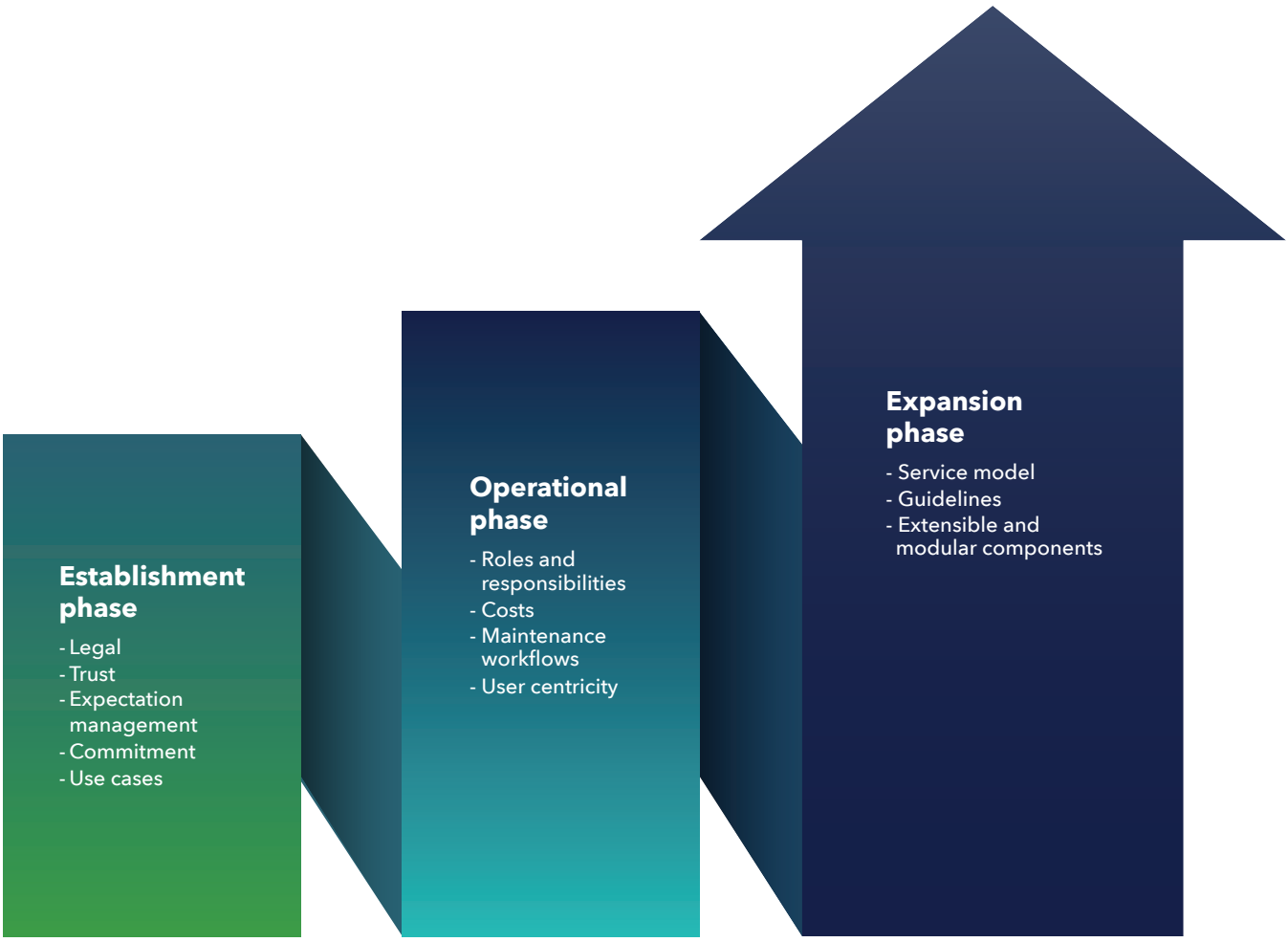


Figure 2: High-level overview of the requirements identified by the workshop participants, at the establishment, operation, and expansion phase of an ideal FHDN.

Table 2. Detailed descriptions of the requirements identified by the workshop participants, at the establishment, operation, and expansion phase of a FHDN.

Requirement	Description
Establishment phase	
Legal	The mapping and concurrent design of the requirements, frameworks and technological solutions to adhere to legal precedents.
Trust	Ensuring trust both offline and online across the FHDN ecosystem, promoted by working with known collaborators, and utilising reinforcement through continuous communication.
Expectation management	Clarifying the potentially non-linear relationship between investment from new technology deployment versus gain in scientific knowledge and improved clinical outcome.
Commitment	Assuring commitment at both the institutional level and from the leadership of all partners in the FHDN.
Use cases	The prioritisation and selection of high-profile, clear and explicit examples that appeal to all partners, as well as clear determination of functional and non-functional associated requirements.
Operational phase	
Roles and responsibilities	Defining the roles and responsibilities of the various partners within a FHDN to ensure that the setting and management of understandable and transparent guidelines are achieved through adequate staffing, software, hardware, data management and processes for maintaining requirements and fixing of bugs. This should be aided through early involvement of data management teams and ensuring recognition of clear ownership by the end user.
Costs	Ensuring that cost estimates are available (and payable) for both investment and maintenance, alongside the provision of incentives, where relevant.
Maintenance workflows	Setting of workflows for compliance with maintenance timelines and monitoring activities.
User centricity	Designing and implementing based on user support structures, easy access for maintenance and service, and good working interfaces across the user spectrum; from clinicians to data scientists and data managers.
Expansion phase	
Service model	Creation of a service model that can provide a roadmap to develop the fruits of academic output in FHDNs and federated learning into a full Software as a Service (SaaS) or otherwise.
Guidelines	Agreements and governance protocols between partners in the network (including sustainability and expansion opportunities), that are also available for review by parties interesting in joining the FHDN.
Extensible and modular components	The balancing of specific FHDN use case needs with generalisable technical features and APIs that together promote reusability and data availability for multiple applications.

6. Discussion and conclusion

This white paper summarises challenges and potential approaches to tackle these when establishing, operating, and expanding a FHDN. Eight challenges were identified by participants during a workshop hosted by DNV on 7th of December 2021. Participants were selected based on an initial FHDN landscape mapping exercise, where relevant descriptions, experience, use cases and infrastructures were captured. Participants represented different categories of FHDN stakeholders such as i) platform providers (both commercial and academic: Owkin, Vantage6, DataSHIELD, MedCO, Medical Informatics Platform); ii) European Commission (EC) funded projects (EUCANCan); iii) national cancer registries and institutions (Cancer Registry of Norway, Netherlands Comprehensive Cancer Organisation); and iv) an independent third-party assurance and risk management company (DNV). The workshop gathered pioneers at the forefront in the field of federated health data networks and created an arena for them to discuss common challenges, opportunities and needs. Through guided discussions, the main challenges were tabled, and potential approaches to these and ways forward were discussed.

The challenges identified in the workshop were clustered in order of (shared) importance: (I) trust, (II) infrastructure interoperability, (III) sustainability, (III) legal and regulatory, (IV) data interoperability, (V) network establishment, (V) scalability and (V) incentives. These categories can be broadly characterised by two main types: the primarily technical challenges (i.e., infrastructure interoperability, data interoperability, scalability), and those of a more non-technical nature (i.e., trust, network establishment, legal issues, incentivisation and sustainability).

Further discussion around these challenges clearly revealed that establishing, operating and expanding FHDNs requires a broad spectrum of expertise. While the technical challenges can in most instances be solved by the platform providers, the cross-disciplinary, non-technical issues need to be addressed through more diverse expertise. At the same time, the requirements from one challenge may have consequences on another challenge, and as such are frequently not solvable in isolation from each other. For example, the requirement for trust emerged during the workshop as the most important enabler in all phases of a FHDN, and as such represents a critical component for addressing all challenges discussed.

As an independent third-party assurance and risk management company, DNV explores opportunities for continual value creation through the assurance and enabling of trust in emerging technologies. By investigating the challenges and approaches in FHDNs we can better understand how our independent role could support all phases of establishing, incentivising, and enabling trust to and between partners on both technical and non-technical fronts.

The identification of assurance needs for FHDNs, relies upon a collaborative process with partners and stakeholders, allowing an understanding of needs, and subsequent co-creation of tools and approaches that respond to these needs. In the context of FHDNs, DNV could enable trust using approaches that address trust gaps and other challenges highlighted in the workshop, through the development of governance frameworks (e.g., standards, guidelines, recommended practices, risk management and codes of conduct), cybersecurity, assurance, and data quality assessments among others, that could accelerate the uptake of FHDNs and subsequent data access for improved clinical outcomes within healthcare.

Finally, although this workshop focused on the healthcare sector where FHDNs are gaining recognition as the preferred solution to overcome the barriers in access to health data, it should be noted that the challenges, opportunities and potential approaches are common to and of relevance for other industries and sectors applying FNs to overcome issues of data access and distributed learning, especially for commercially sensitive data. It is hoped that papers such as this one can raise awareness of these challenges and potential solutions and inspire development in the next phases of FNs.

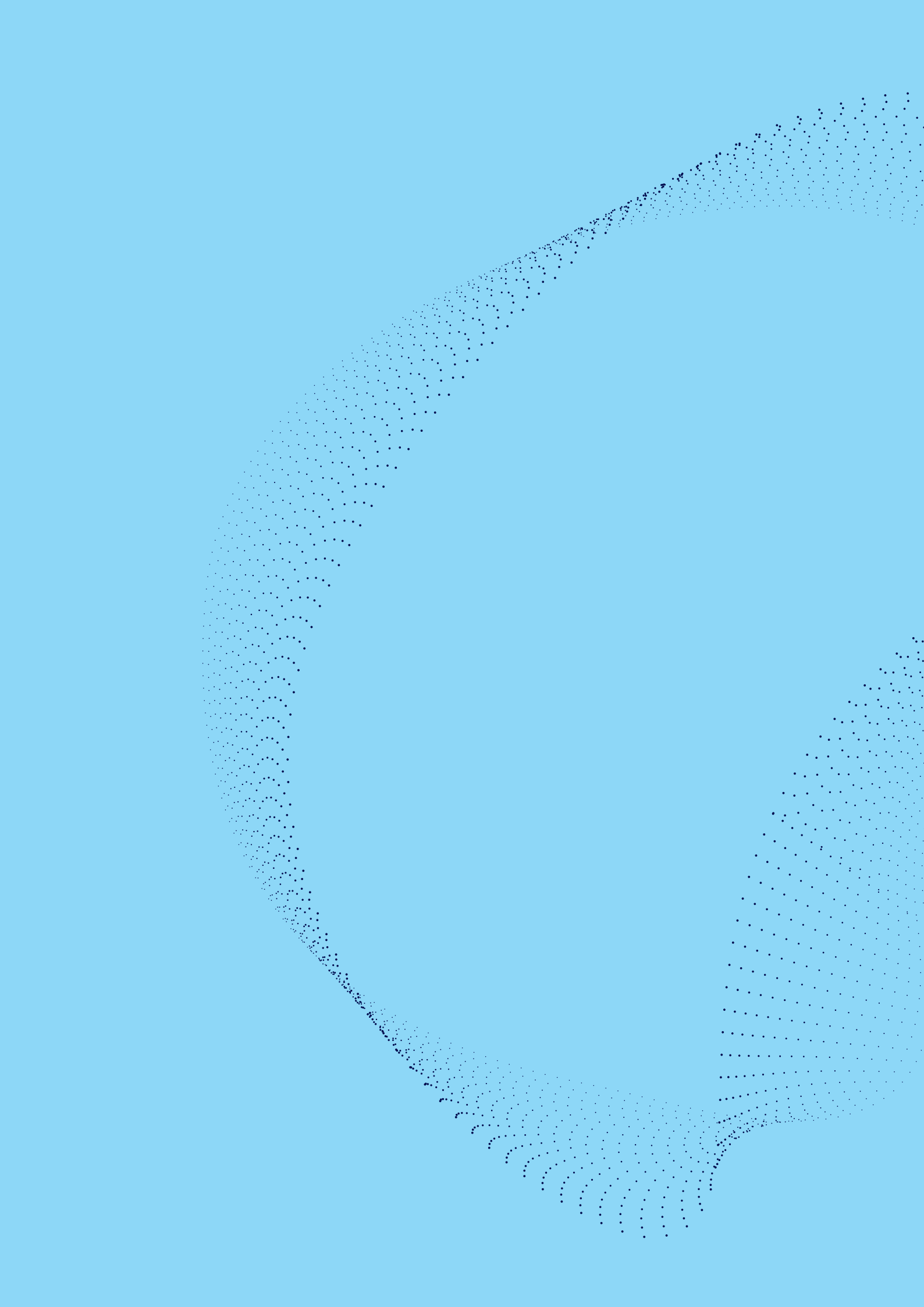
Acknowledgement

This workshop was sponsored by BigMed (26), Norway's largest precision medicine initiative.

We would like to acknowledge and thank the workshop participants for sharing their expert opinion and knowledge about FHDNs for the creation of this white paper.

11. References

1. Alter G, Gonzalez R. *Responsible practices for data sharing*. Am Psychol. 2018;73(2):146–56.
2. DNV. *Dynamic consent in clinical genetics: Implementation barriers.* Oslo; 2021.
3. Peloquin D, DiMaio M, Bierer B, Barnes M. *Disruptive and avoidable: GDPR challenges to secondary research uses of data.* Eur J Hum Genet. 2020;28(6): 697–705.
4. Andanda P. *Towards a Paradigm Shift in Governing Data Access and Related Intellectual Property Rights in Big Data and Health-Related Research.* IIC - Int Rev Intellect Prop Compet Law. 2019;50(9):1052–81.
5. Schwalbe N, Wahl B, Song J, Lehtimaki S. *Data Sharing and Global Public Health: Defining What We Mean by Data.* Vol. 2, Frontiers in Digital Health. 2020.
6. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. *A systematic review of barriers to data sharing in public health.* BMC Public Health. 2014;14(1):1144.
7. Hallock H, Marshall SE, 't Hoen PAC, Nygård JF, Hoorne B, Fox C, et al. *Federated Networks for Distributed Analysis of Health Data.* Vol. 9, Frontiers in Public Health. 2021.
8. Perelman school of Medicine. *The Federated Tumor Segmentation (FeTS) initiative [Internet].* Available from: **https://www.med.upenn.edu/cbica/fets/**
9. ELIXIR. *Beyond 1 Million Genomes [Internet].* Available from: **https://b1mg-project.eu/**
10. Naz S, Phan KT, Chen Y-PP. *A comprehensive review of federated learning for COVID-19 detection.* Int J Intell Syst. 2022 Mar 1;37(3):2371–92.
11. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, et al. *Federated learning for predicting clinical outcomes in patients with COVID-19*. Nat Med. 2021;27(10):1735–43.
12. European Medicines Agency. *Data Analysis and Real World Interrogation Network (DARWIN EU) [Internet].* 2021. Available from: **https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-world-interrogation-network-darwin-eu**
13. World Economic Forum. *Sharing Sensitive Health Data in a Federated Data Consortium Model: An Eight-Step Guide.* 2020.
14. Raisaro JL, Marino F, Troncoso-Pastoriza J, Beau-Lejdstrom R, Bellazzi R, Murphy R, et al. SCOR: *A secure international informatics infrastructure to investigate COVID-19*. J Am Med Informatics Assoc. 2020 Nov;27(11):1721–6.
15. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. *The future of digital health with federated learning*. npj Digit Med. 2020;3(1):119.
16. Mandl KD, Glauser T, Krantz ID, Avillach P, Bartels A, Beggs AH, et al. *The Genomics Research and Innovation Network: creating an interoperable, federated, genomics learning system*. Genet Med. 2020;22(2):371–80.
17. Veeraragavan NR, Nygård JF. DeCanSec: *A Decentralized Architecture for Secure Statistical Computations on Distributed Health Registry Data.* In: ACM International Conference Proceeding Series. 2021.
18. Becker C. Sustainability and Longevity: *Two Sides of the Same Quality?* 2014.
19. OECD. *Blockchain technologies as a digital enabler for sustainable infrastructure.* OECD Environ Policy Pap. 2019;16.
20. Mukta R, Paik H, Lu Q, Kanhere SS. *A survey of data minimisation techniques in blockchain-based healthcare.* Comput Networks. 2022;205:108766.
21. *Informatics for Integrating Biology and the Bedside (i2b2-transSMART) [Internet].* 2022. Available from: **https://i2b2transmart.org/**
22. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data. 2016;3(1):160018.
23. *The Observational Health Data Sciences and Informatics (OHDSI) [Internet].* 2022. Available from: **https://www.ohdsi.org/**
24. European Commission: *European Health Data Space [Internet].* 2019. Available from: **https://ec.europa.eu/health/ehealth/dataspace_en**
25. Devriendt T, Borry P, Shabani M. *Factors that influence data sharing through data sharing platforms: A qualitative study on the views and experiences of cohort holders and platform developers.* PLoS One. 2021 Jul;16(7):e0254202.
26. *BigMed [Internet].* Available from: **https://bigmed.no/**



ABOUT DNV

We are the independent expert in risk management and quality assurance. Driven by our purpose, to safeguard life, property and the environment, we empower our customers and their stakeholders with facts and reliable insights so that critical decisions can be made with confidence. As a trusted voice for many of the worlds most successful organisations, we use our knowledge to advance safety and performance, set industry benchmarks, and inspire and invent solutions to tackle global transformations.

Healthcare Programme, Group Research and Development
Group Research and Development focuses on knowledge development and provides scientific legitimacy, thought leadership and forecasts on the world's major transformations. DNV's healthcare programme works with partners through large-scale public-private research and innovation projects to address the data-sharing and infrastructure needs, legal and regulatory challenges, and gaps in trust that prevent the clinical implementation of precision medicine, AI, and new digital technologies.

Contact: Sharmini.Alagaratnam@dnv.com

DNV AS
NO-1322 Hovik,
Norway
Tel 47 67 57 00
www.dnv.com