Transnational health care research – carefully connecting the dots towards a European Health Data Space

Report on the workshop of the initiatives SPHN, Health-RI and MII on 26/27 June 2023 and the resulting follow-up process

• • • • legal framework • • data reuse • • informed consent • • interoperability • • data sharing • • sustainability • • fails and successes • • • •

Authors:
J. Beliën[1], M. Beusink[1], M. Boeker[2], J.-W. Boiten[1], N. Bolding[1], K. Buckow[3], K. Crameri[4], J. Drepper[3], P. Kleinert[3], T. Ganslandt[5], T. R. Geiger[6], H.-U. Prokosch[5], U. Sax[7], A. Scherag[8], S. H. Schug[9], S. C. Semler[3], M. Steiner[3], F. Wissing[10], S. Zenker[11]

(all authors contributed equally to the report and are listed in alphabetical order)

Contributors:
O. Appleton[4], A. Bernhardt[3], L. Boninodasilvasantos[12], P. Hirschi[13], S. Kapoor[4], T. Kirsten[14], J. Maurer[4], T. Khan[1], S. Österle[4], F. Prasser[15], P. Volkert[1]

1 Health-RI, Utrecht, The Netherlands (for Health-RI)
2 Technische Universität München, Munich, Germany (for MII)
3 Technology, Methods, and Infrastructure for Networked Medical Research (TMF), Berlin, Germany (for MII)
4 Personalized Health Informatics Group, SIB Swiss Institute of Bioinformatics, Basel, Switzerland (for SPHN)
5 Friedrich-Alexander-Universität, Erlangen, Germany (for MII)
6 Swiss Personalized Health Network, Swiss Academy of Medical Sciences, Bern, Switzerland (for SPHN)
7 Universitiätsmedizin Göttingen, Göttingen, Germany (for NFDI4health and NFDI section ELSA)
8 Universtitätsklinikum Jena, Jena, Germany (for MII)
9 IQmed® Beratung für Information und Qualität im Gesundheitswesen, Frankfurt am Main, Germany
10 Medizinischer Fakultätentag, Berlin, Germany (for MII)
11 Universitätsklinikum Bonn, Bonn, Germany (for MII)
12 Leids Universitair Medisch Centrum, Leids, The Netherlands (for Health-RI)
13 University Hospital Zurich, Zurich, Switzerland (for SPHN)
14 Universitätsklinikum Leipzig, Leipzig, Germany (for MII)
15 Berlin Institute of Health @ Charité, Berlin, Germany (for MII)

Funding acknowledgements:
The Swiss Personalized Health Network (SPHN) initiative is financed by the Swiss State Secretariat for Education, Research and Innovation.
The Medical Informatics Initiative (MII) is funded by the German Federal Ministry of Education and Research.

DOI: 10.5281/zenodo.12607200 July 2024

Content

DOI: 10.5281/zenodo.12607200 July 2024

Executive summary

On 26/27 June 2023, a cross-national exchange with representatives of the Swiss Personalized Health Network (SPHN), the Dutch national initiative for integrated health data infrastructure (Health-RI) and the German Medical Informatics Initiative (MII) took place in Berlin.

This Workshop report summarises the outcomes and actions from the presentations and discussions for the various sub-topics, such as consent and legal framework for data reuse, interoperability, methods of data sharing, and financing and sustainability.

The workshop resulted in the following joint vision of the three initiatives:

Making use of similarities in mandates/goals, conditions and solutions.

All three countries have implemented comparable initiatives. They work along similar approaches to make healthcare data available for scientific (re)use and face similar challenges in doing so.

Due to these commonalities, SPHN, Health-RI, and MII plan on joining forces wherever possible. The initiatives aim to strengthen their collaboration, exchange experiences and, if possible, jointly address common challenges.

National infrastructures for the reuse of healthcare data in the Netherlands, Germany and Switzerland

SPHN, Health-RI, and MII have established large-scale work programmes to build a national infrastructure that enables the retrieval and access to health care data. Guiding principles in all three initiatives are FAIR data management and data usage. This is enabled using international standards and terminologies for semantic and structural interoperability. While not all exchange formats in use are identical, bridging these seems viable. A data governance including use and access policies is either already established or corresponding developments are under way.

Potential joint contribution to research at European level

The representatives from the three initiatives agree on exploring the opportunities for expansion of the cooperation between the initiatives and envisage submitting joint funding applications for this aim.

The initiatives target synergies with the upcoming EHDS regulation. This regulation will have the potential to stimulate the establishment of a common framework for health data, including a consistent, trusted, and efficient data governance, standards, data quality labels and practices. Moreover, the planned European Metadata Catalogue will help to identify, describe and compare health data sources and facilitate data sharing and interoperability between different stakeholders in the EHDS framework.

To make the presented visions a reality, activities were defined to intensify and strengthen cooperation between the initiatives. These will have to be addressed jointly in the upcoming months.

DOI: 10.5281/zenodo.12607200 July 2024

1. Workshop setting

On 26/27 June 2023, a cross-national exchange with representatives of the Swiss Personalized Health Network (SPHN), the Dutch national initiative for an integrated health data infrastructure (Health-RI) and the German Medical Informatics Initiative (MII) took place. The aim of the workshop was to gain a mutual understanding of the vision and implementation approaches of these initiatives. Commonalities should be emphasised, backgrounds for differences should be understood and key people should be connected. The workshop covered the following topics:

- Legal framework for data reuse,
- Patient Consent,
- Semantic Interoperability,
- Data Sharing,
- Failures and unsolved challenges,
- Financing and sustainability.

2. Introducing the three initiatives

The **Swiss Personalized Health Network (SPHN)** is a government-funded research infrastructure initiative with the aim of making data from routine health care as well as related biological data (i.e., omics data) better usable for research purposes. The focus of SPHN lies on standardisation of health data, infrastructure development for responsible and efficient sharing, processing and storage of sensitive data, as well as data FAIRification. The scope of action includes the primary and secondary use of data (also outside of research) with a view to the envisioned Swiss health data space and a digital public service.
Main infrastructure components and process innovations developed in the realm of SPHN encompass (i) well-connected and integrated data platforms at all five Swiss university hospitals, (ii) catalogues and tools for data findability, feasibility assessment as well as federated data exploration, (iii) a coherent semantic interoperability framework including a tool stack for enabling the semantic representation of health data within a knowledge graph, (iv) a trusted research environment for the processing and storage of sensitive data, (v) templates and blueprints for legal agreements for multiparty research projects using health care data and (vi) practical as well as ethical guidelines along the research workflow.

Health-RI is a Dutch non-profit foundation supporting a public private partnership of organisations with the aim to realize a national health-data infrastructure. In 2021, Health-RI received a major grant from the Dutch Ministry of Economic Affairs as part of the so-called “Growthfund”. The Health-RI mission is to build an integrated health data research infrastructure accessible for researchers, citizens, innovators, and care providers. Realising this ambition involves (i) optimizing the conditions (including ethical, societal, legal aspects) for building and maintaining a national health data infrastructure; (ii) implementing the required technical building blocks (data portal, FAIR data platforms, federated access)

DOI: 10.5281/zenodo.12607200 July 2024

of a health data infrastructure and (iii) providing mature services, such as supporting researchers and data managers by making infrastructure services, tools and data easy to locate, access and use. Health-RI is organised through a hub-and-node model with the local nodes usually centred around university medical centres. A specific focal point is the “obstacle removal trajectory”: Health-RI closely collaborates with responsible ministries (health, science, economic affairs) and a large range of stakeholders in the health data landscape to remove the legal, social and technical obstacles that currently still prevent us from sharing health data at scale.

The **German Medical Informatics Initiative (MII)** is a long-term project funded by the German Federal Ministry of Education and Research since 2018. The aim of the project is to establish a national infrastructure for the reuse of healthcare data. Data integration centres (DIC) established at all German university hospitals are responsible for the retrieval, harmonisation and provision of healthcare data. It’s main focus so far is data from university hospitals. The integration of further data sources is in progress. Data is provided for research purposes based on joint data sharing regulations. MII ‘Broad Consent’ is the prerequisite for the data use by researchers if raw data transfer is required for analysis. Alternatively, distributed analyses can be performed. As the central component of the MII infrastructure, the German Portal for Medical Research Data (FDPG) manages the application, authorisation and contract processes as well as the coordination of the data sharing process. It also offers functions for feasibility requests on the decentralised data and overviews of projects on MII data. All basic MII processes have been rolled out and are being expanded and improved in the current MII funding phase.

3. Consent and legal frameworks

All initiatives have implemented a consent-based solution as the basis, or one of the bases, for data usage (secondary use) but for different regulatory reasons.

In **Switzerland**, the legal framework is set by fragmented cantonal data protection laws, a national data protection law and the Federal Act on Research involving Human Beings (Human Research Act, HRA). Under the HRA, which came into force in 2014, consent is mandatory for secondary use of data. A general consent has been approved by all seven Swiss regional ethics committees and is widely used by Swiss hospitals.

In **Germany**, the complex structure of healthcare organisations and research infrastructures is reflected in a highly fragmented legislative framework. Thus, various legal grounds for research exist, but no regulation fits all participating sites, especially for use cases going beyond local analysis of anonymized datasets. As a result, the MII developed a broad consent that has been approved by all federal state data protection authorities and the umbrella organisation of German ethical boards. The exact wording of the broad consent is fixed to address standardisation and transparency requirements. Project-specific additional requirements can be addressed via a modular extension process which ascertains that approved extension modules become available to the entire community.

DOI: 10.5281/zenodo.12607200 July 2024

In the **Netherlands**, a patient consent is often the legal basis for processing under the Medical Treatment Contracts Act and the GDPR Implementation Act. There is a possibility for implicit opt-out in case consent cannot reasonably be obtained but whether this is the case is decided on a case-by-case basis. Researchers and Ethical boards may use the national Code of Conduct for guidance related to consent and participant information.

3.1. Implementing consent

Implementation details for patient consent differ between the three initiatives when considering aspects like the extent consent is being used, the type of consent, the creation of the templates, the length of the respective document(s) and uniformity/nonuniformity of consent templates in use by the implementation sites:

SPHN relies on a harmonised general consent for the further use of data and samples, consisting of three pages with one single "yes or no" option.

The **MII** elaborated – in an extended iterative process that involved all German data protection authorities and all German medical ethics boards – a modular 12 pages broad consent that aims to prepare for a large range of contingencies; site customization is strictly constrained by a dedicated guidance document, which also defines required procedural measures to ensure adherence to the agreements reached with the data protection authorities[1].

Health-RI works towards a national consent/opt-out registry. Currently, consent forms being used by each site differ in many aspects[2]. Some hospitals use a broad consent procedure, which exists in parallel to options for an opt-out under certain conditions provided by the national law: The Medical Treatment Contracts Act (MTCA).

3.2. Some implications for further international collaboration and/or cross-border data sharing

Each of the initiatives endorses open and intensified communication to engage healthy citizens and patients on national and international levels. Respective information campaigns could be highly instrumental for underlining the importance of using and reusing health data for research purposes. The current finalisation of new legislation towards establishing the European Health Data Space (EHDS) presents a good opportunity to give this topic more (public) attention.

---

[1] Zenker S, Strech D, Ihrig K, Jahns R, Müller G, Schickhardt C, u. a. Data protection-compliant broad consent for secondary use of health care data and human biosamples for (bio)medical research: Towards a new German national standard. J Biomed Inform. Juli 2022;131:104096

[2] even upon the availability of an EU-wide domain consent template, which is the case for rare diseases (https://www.ejprarediseases.org/ern-registries-generic-icf/), each academic hospital may implement it differently according to its local legal and medical-ethical teams' requirements.

DOI: 10.5281/zenodo.12607200 July 2024

Next to the technical and semantic dimension, data sharing largely depends on policies and trust. In former times, datasets were either transferred to a requestor in its entirety or not at all, while nowadays more refined methods for granting access are feasible. As data holding organisations and citizens feel uneasy towards potentially uncontrolled use of their data we should/can design a careful and considerate roadmap from (1) making data discoverable by sharing the notion of sheer existence (check in a portal with a sparse metadata set), (2) share some more details about the data collection by providing an interoperable information model including data models and value sets, (3) offering the policy metadata together with the data to inform about the legal circumstances and ways to get to use the data and in some cases (4) offering an anonymized test data set to train the requester's infrastructure to (5) use federated analytics methods that enable the algorithm to be brought to the decentralised data and compute in a federated manner instead of bringing the data together in a central location. In all cases finally the data subject should be informed about implemented data use and the corresponding results/publications (transparency).

All those steps are currently provided in the national solutions, but we see no obstacle to implement each of the steps 1-5 in an international setting under the umbrella of European legislation, in particular GDPR and EHDS.

4. Interoperability

Each initiative has developed information models facilitating the harmonised provision of data for generic research purposes. While having been developed independently of each other, each of these models are referring to internationally established code systems and terminologies such as ICD-10/11, SNOMED CT, LOINC and UCUM. Regarding data exchange formats, Health-RI and MII use HL7 FHIR, while SPHN uses RDF as an exchange format where the schema builds on controlled vocabularies of RDFS, OWL and SKOS. RDF is also in some use by Health-RI, e.g. for parts of its metadata catalogue (FAIR Data Point). All three initiatives use supplemental standards and technology, e.g. to map data or to make it easier to analyse. All three initiatives agree on using a decentralized storage topology with option for federated and centralized analyses.

The **SPHN** data schema relies on the SPHN Concepts, which can be used as modular building blocks. These SPHN Concepts are described in the SPHN Dataset and formalised in an RDF Schema[3]. The SPHN Dataset and RDF Schema are developed by the SPHN Data Coordination Centre with support of the SPHN Semantics Working Group based on requirements and data availability by the data providing sites and SPHN projects. The graph-based representation enables automated validation and data quality checks as well as annotation of datasets according to the FAIR principles. Data storage remains decentralised on the data platforms of the hospitals. To facilitate data transformation to the common SPHN format, SPHN offers the SPHN Connector, a tool which transforms, de-identifies and validates the data.

[3] https://www.biomedit.ch/rdf/sphn-ontology/sphn

DOI: 10.5281/zenodo.12607200 July 2024

The **MII** modular Core Dataset (MII CDS) describes the structure, format and semantic annotation of the MII data based on HL7 FHIR. The MII CDS is divided into basic modules, which rely on broadly available data (e.g. ICD codes, lab data) and are mandatory, and extension modules covering a broader range of the clinical care and research process (e.g. molecular diagnostic information, biospecimens), which may be provided by MII sites depending on availability and use case requirements.
Within a governance framework, interdisciplinary teams collect requirements and design FHIR profiles with community participation through the HL7 balloting process, resulting in publicly available implementation guides. The use of HL7 FHIR profiles aligns the MII CDS with the ongoing implementation of open standards in the German healthcare system coordinated by the Interoperability Council[4] and is moreover in line with the European Electronic Health Record Exchange (EEHRxF) format pathway. The EEHRxF will become mandatory once the EHDS regulation enters into force (given the assumed to be final compromise between the European Parliament and the Member States – Consilium).
The FHIR profiles also enable automated validation of instance data against their respective specifications. Data storage remains similarly decentralised in the data integration centres (DIC) established by participating hospitals. DIC are responsible for extracting data from healthcare IT systems, transforming it according to the FHIR-based MII core dataset and making it accessible.

The Health-RI schemas are organised according to the metaphor of a sunflower with a hybrid governance where the core is defined centrally, and the subdomains responsible for the different petals define their own schemas. The core consists of common metadata about resources (e.g., datasets) in the ecosystem. The domain-specific semantic models ("petals") should be applicable for primary use (health professionals) as well as for secondary use (researchers) of health data. When multiple petals have elements in common, these will be embedded in the core. The core team provides guidance for the petals from their onset, and mappings among different standards will be done where needed, thus preventing quality loss between user groups. For Health-RI, a decentralised data storage model is foreseen with a coordinating role for each of the regional Health-RI nodes which become responsible for unlocking the health data in their region.

In terms of the data sources made available by the initiatives for secondary use, clearly distinct approaches are pursued by MII on the one hand and Health-RI and SPHN on the other:

For the **MII**, at this time primarily data from inpatient healthcare[5] is made available in a harmonised way, as the initiative has so far focused on university hospitals, which have limited outpatient care mandates within the German health system. However, the ongoing roll-out of a personal electronic health record (pEHR) covering publicly insured patients, i.e. approx. 90% of the population, provides explicit options for research use of pEHR and hospital data. The MII is aiming to connect further data sources. Preliminary work on this has already begun in the MII through the Digital Hub funding initiative

[4] https://www.ina.gematik.de/mitwirken/expertengremium

[5] Outpatient healthdata is available in MII's DICs as well. The amount is just less given reduced documentation requirements for ambulatory cases and given the division of sectors in German healthcare with the majority of cases being treated in ambulatory practices.

DOI: 10.5281/zenodo.12607200 July 2024

(providing outreach into the ambulatory sector) and the ongoing use case projects of the 3rd MII funding phase, which will integrate new datasets from several clinical domains.

For **Health-RI** and **SPHN**, data from outpatient care are also in scope next to data from inpatient care, although this is still largely on the drawing board. Furthermore, additional data types from research, such as genomic data, study data and register data are linked by both initiatives. Health-RI also aims to access data from pEHRs in the future.

Making structured and harmonised data available is a challenging task for all three initiatives. The process requires time and qualified staff.

5. Data sharing and analysis

5.1. Data sharing and data access principles

The principles for data sharing in **SPHN** are defined by a National Steering Board. Data are controlled by data producers, i.e. hospitals and research consortia. Data requests from researchers must be approved by institutional governance boards complying with local regulations and national and cantonal laws as well as responsible ethics committees' decisions. Data are analysed within the SPHN Trusted Research Environment "BioMedIT"[6]. SPHN enables project-specific data sharing within research consortia via BioMedIT. More than 35 SPHN-funded projects and four large "National Data Streams" have been supported in this way until now. Furthermore, selected cancer and laboratory datasets have been assembled for federated analysis spanning a limited number of hospitals.

For the **MII**, local, similarly structured Use & Access Committees grant data access in each particular case in close coordination with the local data protection officers on the basis of the state or federal regulations. In addition, approval by an ethics committee is required for each project as a whole.
The process of data sharing and the related working instructions follow along the guidance developed by the respective MII working group. Non-consented datasets can be used for federated analysis, where aggregated (anonymized) analysis results are provided to the researcher and, additionally, consented datasets can be analysed centrally. MII provides datasets to researchers for either federated analysis (consented and non-consented data is analysed locally and results are provided) or central analysis (consented data are consolidated centrally at data management units and provided to the researchers directly).

Health-RI has not yet implemented an overarching data governance; specific implementations exist for biobank & cohort collaborations, and for the national COVID-19 data repository. A more generic data access procedure will probably be implemented as part of the Dutch Health Data Access Body for the EHDS (with participation of Health-RI). Health-RI uses Data Catalogue vocabulary (DCAT[7]), as a standard

[6] https://www.biomedit.ch/home/biomed-it-infrastructure.html

[7] In particular the DCAT application profile (DCAT-AP) as established within Europe.

DOI: 10.5281/zenodo.12607200 July 2024

for describing datasets, to enable cross-data portal search for datasets or to make decentralised data more searchable centrally.

All initiatives see FAIR data sharing as a crucial principle for improving data quality and data handling processes by data providers ("FAIR at the source"). Concerning data discoverability, the GDPR guides the modalities available for data use in the three initiatives, leading then to e.g. central analysis of individual data and/or federated analysis of anonymous data. SPHN is restricted to consented datasets and non-consented datasets if exempted according to Art. 34 HRA and authorised by the ethics committees.

Difference between the implementations of central vs. federated approaches can be observed: SPHN predominantly utilises data transfers in a central, trusted research environment (BioMedIT) for central analyses (while federated analyses are also in use); MII stronger emphasises federated analyses with the data staying on premise, i.e. employing its well-defined process for feasibility, data request, use & access committee and data use agreement. While SPHN and MII already share datasets with users, Health-RI had not yet started this practice at the time of the workshop except for a national COVID-19 data set being available for user requests.

5.2. Processes and tools for requesting data

SPHN provides a feasibility query tool, the SPHN Federated Query System, which allows queries across all five Swiss university hospitals to identify patients with data in the clinical data platforms. In the current setup data of 0.6 Mio patients, who all signed the general consent, are included. The following data elements can be currently queried: age, gender, diagnosis, procedure, lab test and results; and medications. For supporting the users of the search tools, SPHN provides focused video tutorials and requires researchers to be trained in various aspects of data analysis, FAIR datasets, privacy and data protection.

MII interconnects projects with existing expertise in the data integration centres (DIC) and intends to provide test datasets to be studied by applicants for training. Further projects are initiated within the MII to provide training materials, tutorials and workshops depending on the needs of researchers and staff. The MII supports the optional execution of feasibility queries that return the number of patients matching the requests on specific datasets. The MII provides a web interface to researchers for feasibility queries which are provided to the sites as a standardised cohort selection. Furthermore, the MII envisions providing a similar web interface to be used for the data selection (this functionality is currently provisionally implemented using Excel sheets).

Health-RI is developing a data request application as part of its meta data catalogue. A further integration with the national data request application developed in the context of the EHDS is foreseen for the future.

DOI: 10.5281/zenodo.12607200 July 2024

5.3. Research infrastructure, processes and tooling for data analyses:

SPHN and MII use middleware and server infrastructure that is designed to fulfil specific tasks within the initiatives.

SPHN uses BioMedIT as trusted research environment, links data using the SPHN Interoperability Framework, uses end-to-end encrypted data transfer using a tool called “sett”.

The **MII** relies on the DataSharingFramework (DSF) for data encryption, documentation and transfer and developed additional tools for feasibility queries. MII embraces federated analyses - after a well-defined process for feasibility, data request, use & access committee and data use agreement - with the data staying on premise.

Health-RI has not formally standardised on a solution, but each of the Health-RI regional centres are supporting a Trusted Research Environment. When it comes to providing data and services to researchers, Health-RI is still in a conceptual phase in many aspects, although various predecessor projects already have developed core services on which Health-RI is building.

5.4. Research culture and understanding the benefits

The participation of data holders in national data sharing could still be improved for each of the three initiatives. The incentives and benefits of national data infrastructure initiatives need to be spelled out more clearly, as these are often more long-term and generic in comparison to (local) project-centred initiatives which tend to be prioritized for that reason.

6. Financing and Sustainability

The **SPHN** initiative (2017-2024) has been financed by the Swiss State Secretariat for Education, Research and Innovation (SERI) with CHF 135 million plus matching contributions from hospitals and universities. After the completion of the initiative, the SPHN Data Coordination Centre (SPHN-DCC) will be sustained with core financing from SERI, with a minimum financing of the network’s decentralized data infrastructure components at universities and hospitals s. In the long-term, high-quality health data needs to be incentivised and rewarded by the healthcare system, not research funding. Going forward, the SPHN-DCC will keep a multi-stakeholder governance model in order to act as an independent broker offering high-value services to all stakeholders of the health data ecosystem.

The **MII** is currently in its third funding phase. The development and networking phase (2018-2022) was funded with 200 million euros by the Federal Ministry for Education and Research. The subsequent consolidation and extension phase (2023 - 2026) will again provide around 200 million euros with a further 80 million euros for junior research groups and regional projects. Data sharing infrastructure and processes should be established and consolidated by the end of the current funding phase. However, ensuring the sustainable operation of the data integration centres and the central components of the

DOI: 10.5281/zenodo.12607200 July 2024

MII infrastructure as well as the continuous further development of the components requires basic permanent funding that is envisaged through the German “Network University Medicine” structure.

For the period 2022-2028 **Health-RI** is mainly financed by the economic stimulation funds (“Growth Fund”) of 69 Mio EUR of the Ministry of Economic Affairs with support of the Ministry of Health and the Ministry of Science & Education. In the future, the financing will likely be a mixture of permanent funding or utility funding for e.g. EHDS services, membership fees, fee for services and project-based innovation funding.

All initiatives see (further) needs for ensuring their long-term operation.

7. Joint vision and key aspects of the intended cooperation

The Netherlands, Germany and Switzerland have implemented comparable initiatives to establish national infrastructures for the reuse of healthcare data. They work on similar solutions to make healthcare data available for scientific (re)use and thus face similar challenges.

In view of these commonalities SPHN, Health-RI, and MII plan on joining forces. The authors of this workshop report aim to strengthen their collaboration, exchange experiences and address common challenges jointly.

Building on the momentum of the upcoming EHDS regulation, the three initiatives could become a focal point for further international collaboration, with the ambition to overcome existing differences and to enable cross-border data sharing for secondary use. This joint action should offer our researchers capabilities for meaningful research on larger patient populations.

The representatives from all three initiatives agree on expanding the cooperation between the initiatives, especially by jointly responding to the identified common challenges.

7.1. Drivers and (clinical) business models for data quality related efforts

Data quality us a key requirement for effective health research. Further investments are therefore needed, both in innovative, e.g. AI-supported, solutions, and the necessary staff and financial resources. Better integration with clinical processes can improve data quality and healthcare delivery by aligning the research infrastructure with the workflows and goals of healthcare providers. For example, data quality can be improved by using standardised definitions, formats, and codes for medical data; implementing data validation and verification mechanisms; and facilitating data sharing, data-reuse and interoperability among different systems and stakeholders.

In addition, the EHDS will stimulate the improvement of data quality by providing a consistent, trusted, and efficient framework for data governance, standards, data quality labels and practices. In this context, data quality is the degree to which data meets the expectations and requirements of its users. It is important for making informed and intelligent decisions, increasing efficiency, and maximizing benefits. By creating a common framework for health data, the EHDS will facilitate data sharing and

DOI: 10.5281/zenodo.12607200 July 2024

interoperability, improve data accuracy and completeness, and foster data innovation and value creation.

Governed by the EHDS regulation, the European Metadata Catalogue will help to identify, describe, and compare health data sources and facilitate data sharing and interoperability between different stakeholders.

They (the initiatives) should therefore monitor the progress of legislation at the European and global levels keeping track of the latest developments and respond in a timely manner. This would allow them to align their policies and standards with the emerging European Health Data Space (EHDS) and other international initiatives, and where possible to take joint action.

7.2. Planning and implementing targeted public outreach

Patients and the general public have different needs, expectations, and preferences when it comes to health data and research. Therefore, the three initiatives are each designing and implementing public programs that take into account these differences and address their specific interests and concerns. Distinct public programs for patients and the general public are critical for the three initiatives as they can inform and empower these key stakeholders to participate in and benefit from health data sharing and research.

Outreach to researchers is another key aspect to consider as researchers can facilitate the exchange of knowledge, data and methods among the scientific community and foster interdisciplinary collaboration and innovation. The MII already offers several public programs for researchers, such as symposia, workshops, webinars, newsletters and publications, to showcase the results and achievements of the initiative and its consortia, and to share best practices and lessons learned. These could be extended to the other initiatives.

Joint publications are valuable tools for the three initiatives to communicate and collaborate with the wider scientific and health community. Joint publications, such as reports, articles, and white papers can showcase collaborative achievements and challenges, and share insights and recommendations for building and using health data infrastructures for research and care. They can also highlight synergies and complementarities and identify potential areas for future cooperation and alignment.

Further detailing of the collaboration could be organised through (targeted) follow-up meetings, in order to share and discuss successes and challenges, and ultimately to refine strategies and actions accordingly.

7.3. Qualified staff: Recruiting, training and qualifying staff

The MII, SPHN and Health-RI are national initiatives that aim to build infrastructures for the use of health data for research and innovation in Germany, Switzerland and the Netherlands respectively. They face similar challenges in recruiting skilled staff, especially in IT and data science, due to the high

DOI: 10.5281/zenodo.12607200 July 2024

demand and low supply of skilled workers in these fields across Europe. To address these skills shortages, the three initiatives may seek for opportunities to jointly address capacity and capability building e.g. promoting STEM education and careers among young people, with an emphasis on gender balance and diversity, through awareness campaigns, role models, mentoring and extracurricular activities.

One of these opportunities is to design and facilitate effective learning events that combine participants who are physically present in the same location with those who participate remotely. This hybrid format requires careful planning and coordination of the technical, logistical, and pedagogical aspects of the workshops, as well as the engagement and interaction of both groups of learners.

The three initiatives aim to provide researchers with the knowledge and skills necessary to use and contribute to the data platforms and services they provide. They do this by (1) building curricula that cover the relevant topics and competencies for researchers, such as data management, data analysis, data protection, data sharing, data governance, (2) expanding curricula that incorporate the latest developments and innovations in health data and research infrastructures, such as new methods, tools, standards, policies, and (3) translating curricula that adapt to the specific contexts and needs of researchers from different disciplines, sectors, regions, and languages, and that foster interdisciplinary and cross-sectoral collaboration.

The three initiatives aim to foster a culture of collaboration and learning among researchers who use and contribute to the health data platforms and services they provide. They do this by creating networks and partnerships that enable researchers to share information, ideas, and best practices, and to access resources and opportunities from other organisations, countries, or disciplines. They therefore wish to support mobility and secondment schemes that allow researchers to visit and work temporarily in the partner organisations.
Short-term (or even longer term) staff exchanges can be a powerful instrument to share best practices between the three initiatives, qualify the candidates and strengthen the communication between the experts. Next to the existing project budgets of the three initiatives it could be considered to request specific funding for these exchanges, e.g. through the European Marie-Curie funding scheme.

DOI: 10.5281/zenodo.12607200 July 2024