The JARDIN Hackathon to Seek Solutions to Overcome Technical Barriers in Health Data Exchange: From the Point of Care to European Registry Networks

CÉSAR BERNABÉ
DAPHNE WIJNBERGEN
ALBERTO CÁMARA
KAROLIS CREMERS
MARGARIDA MAGALHÃES
DANIELA VICENTINI ALBRING
SERGI AGUILÓ-CASTILLO
KALIA ORPHANOU
STELLA TAMANA
MARIA XENOPHONTOS
LAURA MENOTTI
MIRCO CAZZARO
ORNELLA IRRERA
JOËLLE THONNARD
SANDER VAN BOOM
IRIS C. M. PELSMA
ANNIKA JACOBSEN
ANDREW GIBSON
VERONICA POPA
MARK WILKINSON
MARCO ROOS

*Author affiliations can be found in the back matter of this article

RESEARCH PAPERS

ubiquity press

ABSTRACT

Effective exchange of health data between healthcare providers, national registries, and European Reference Networks (ERNs) is crucial for advancing research and clinical care of rare diseases. Technical, organisational, and legal barriers persist in many situations that disrupt this process. The Joint Action on Integration of ERNs into National Healthcare Systems (JARDIN) tackles these barriers by identifying such existing solutions. At the technical level, there are existing standards, software, and infrastructures that could be adapted and reused if there were evidence of a method and value in doing so. These need to be identified, matched against known problems, and tested before subsequent work can begin to evaluate and deploy solutions. This paper describes the methodologies and outcomes of a hackathon conducted as a software-oriented workshop to collaboratively identify and implement solutions to technical barriers identified as relevant to rare diseases. The event brought together 47 experts from 10 European countries to combine their knowledge and experience to tackle challenges of data harmonisation, secure data access, and the Findable, Accessible, Interoperable and Reusable (FAIR) description of data services. Motivated by a simple use case, the participants proposed adapting existing infrastructures to create a realistic, secure data-sharing architecture based on existing standards. Solutions included federated querying of rare disease data integrated with a standard for data access conditions, the use of semantic models for rare disease data harmonisation across organisations, and extending a standard metadata model to better describe rare disease data being held by organisations. The process of expert-driven selection of the components yielded approaches relevant for the rare disease community and has provided important momentum. As an output of the JARDIN initiative, this will lead to further real-world evaluation and pilot testing together with rare disease data stakeholders and ultimately EU-wide recommendations that will benefit coordination and focus on improving data exchange.

CORRESPONDING AUTHOR:
Marco Roos
Biosemantics Group, Human Genetics Department, Leiden University Medical Center, Leiden, the Netherlands
m.roos@lumc.nl

KEYWORDS:
FAIR principles; rare diseases data; data exchange solutions

TO CITE THIS ARTICLE:
Bernabé, C., Wijnbergen, D., Cámara, A., Cremers, K., Magalhães, M., Albring, D.V., Aguiló-Castillo, S., Orphanou, K., Tamana, S., Xenophontos, M., Menotti, L., Cazzaro, M., Irrera, O., Thonnard, J., van Boom, S., Pelsma, I.C.M., Jacobsen, A., Gibson, A., Popa, V., Wilkinson, M. and Roos, M. 2026 The JARDIN Hackathon to Seek Solutions to Overcome Technical Barriers in Health Data Exchange: From the Point of Care to European Registry Networks. *Data Science Journal*, 25: 29, pp. 1–17. DOI: https://doi.org/10.5334/dsj-2026-029

INTRODUCTION

The European Reference Networks (ERNs) (European Commission, 2025a) represent a significant collaborative effort to improve the care for people with rare diseases (RDs), as well as complex and low-prevalence diseases in the European Union Member States and Norway. This is achieved through the exchange of patient information among healthcare providers (HCPs), national registries, and ERNs. Unfortunately, heterogeneity of data systems, formats, and standards presents a substantial barrier to achieving this vision. The Joint Action on the Integration of ERNs into National Healthcare Systems (JARDIN) (JARDIN Joint Action, 2025b) has been working to facilitate the seamless flow of health data from its point of capture to its ultimate use, thus improving accessibility to ERNs by integrating them into National Healthcare Systems.

To explore potential solutions to prominent technical and semantic interoperability challenges, JARDIN's Work Package 8 ('Data Management WP') (JARDIN Joint Action, 2025a) convened a 'Hackathon on Health Data Federated Querying' on 4 April 2025. The event gathered information technology (IT) specialists from 10 EU member states. Participants possessed relevant expertise in key areas such as the Findable, Accessible, Interoperable and Reusable (FAIR) principles, semantic web technologies, software development, database operations, and healthcare data standards. The hackathon design was inspired by established, software-focused collaborative workshop formats created to foster awareness of the FAIR principles.

The hackathon objectives were derived from the results of a broad survey that was also conducted as a task of the Data Management WP.[1] This survey evaluated the technical, legal, and organisational barriers (Hughes *et al.*, 2023) by collecting information from clinicians, hospital IT experts, and national authorities. The survey results revealed significant differences in how health data is managed in different European countries. While a few countries use fully digital systems with national registries and standardised codes, the majority still rely on paper-based documentation, or a combination of digital and paper (Henriques *et al.*, 2025). Sharing data manually within ERNs is a laborious process that is lacking (i) an infrastructure architecture for secure data exchange; (ii) the use of standardised diagnosis codes, and (iii) the capacity to make data sources more FAIR. Hence, the hackathon's primary goal was to collaboratively identify and experiment with potential solutions to automate the secure exchange of harmonised FAIR data. The event aimed at answering the following questions: '1. What components are needed for secure querying of RD data?', '2. How can heterogeneous HCP data be harmonised using semantic models?', and '3. What technologies enable machine-actionable discovery of data services?'

This paper provides a comprehensive review of the hackathon, detailing the challenges addressed, the solutions proposed by the working groups, and its role as a translation exercise informing the strategic next steps for the JARDIN initiative.

MATERIALS AND METHODS: THE HACKATHON DESIGN

The hackathon was designed to address three challenges selected from the many identified in the Data Management WP's broad survey[1,2]. These three were chosen in consultation with a panel of JARDIN advisors, prioritising those deemed most urgent and feasible within the scope and duration of a hackathon. The JARDIN advisory panel consisted of 16 professionals with diverse areas of expertise, including patient representation, RD clinical expertise, public health, hospital management, and technical expertise in (meta)data management, semantic web, systems architecture, and the FAIR principles. The selected challenges are presented in Section 'The Hackathon's Focus Challenges'.

A hackathon format was selected, as it provides an intensive, collaborative environment well suited to addressing complex technical challenges within a limited timeframe. The

1 The reader can refer to Henriques *et al.* (2025) for more information on the JARDIN Survey on Data Sharing and Interoperability.

2 Note that the Data Management WP (WP8) survey identified a higher number of challenges relating to technical, organisational, and legal barriers. The hackathon focused on a selection of technical challenges. Other challenges that were not addressed during the hackathon remain relevant and are planned to be addressed in future work.

organising team had prior experience with this format through the 'Bring Your Own Data' (BYOD) workshops (Bernabé *et al*., 2024), which had demonstrated its effectiveness in bringing together multidisciplinary participants to develop practical solutions. In addition, we anticipated that much of the target audience (particularly participants with technical and informatics backgrounds) would already be familiar with this format, facilitating rapid engagement and productive collaboration. Also, given the nature of the challenges selected, which required hands-on exploration and prototyping, the hackathon structure represented a natural and pragmatic choice.

To ground the hackathon's activities in a tangible scenario encompassing all three challenges, a common use case was established: a clinician responsible for clinical trials seeking to identify eligible RD patients across different institutions. In this scenario, a list of patients containing their age, diagnosis, and genotype is required from various hospitals to facilitate recruitment for the trial. This use case highlights the critical need for simplified, secure data flows between hospitals, national systems, and ERN registries.

The four-hour hackathon agenda was divided into three main parts with the following structure and content:

- Introduction (20 minutes): This session provided participants with all necessary context, including an overview of the JARDIN project as well as the Data Management WP, challenges in health data exchange, a review of the three selected hackathon challenges, and organisational instructions.
- Working Sessions & Discussion (3 hours 15 minutes): This central block was divided as follows:
    - Working Session 1 (1 hour 45 minutes): The participants chose which challenge to work on and joined a corresponding breakout group.
    - Group Discussion (20 minutes): All participants reconvened to discuss progress. This also served as an opportunity for individuals to switch working groups if they desired.
    - Working Session 2 (1 hour 10 minutes): The participants continued their collaborative work.
- Wrap-up (15 minutes): The event concluded with a final session to summarise key outcomes and outline potential future directions.

The participants were given resources to help them document their proposed solution and recommend an implementation strategy. These included a virtual whiteboard for brainstorming, a shared storage space for transferring files, and a documentation template. These challenge-specific templates are available as Supplementary Material (Bernabé *et al*., 2025). The template first provided participants with a detailed description of their challenges and the use case. It then had specific sections for the teams to complete, where they could describe their proposed solution, its proof of concept or demonstration, the implementation strategy, and any potential scalability risks. The templates were developed by two hackathon organisers and iteratively refined by contributors from the Data Management WP team to ensure alignment with the JARDIN initiative's broader strategic goals.

PARTICIPANTS OVERVIEW

The hackathon was attended by 47 participants from 10 European countries: Belgium, Cyprus, Denmark, Estonia, Finland, France, Germany, Italy, Spain, and The Netherlands. Participants were invited through JARDIN's mailing list, social media channels, and internal newsletter and were encouraged to share the call with their peers. Given the broad scope of the JARDIN initiative and its engagement with multiple institutions across Europe, these dissemination channels were used to ensure wide visibility and encourage participation from a diverse community of stakeholders. Participation was open to all registrants regardless of their specific expertise, as the objective of the hackathon was to foster multidisciplinary collaboration and gather a broad range of perspectives rather than to select participants based on predefined criteria. A report detailing the hackathon's participation and outcomes is publicly available in Bernabé *et al*. (2025).

To better understand the collective technical proficiency of the group, participants were asked to self-assess their expertise across several key domains. Forty of the 47 participants (85.1%) responded to the assessment. This assessment utilised a five-point rating scale, where 1 represented 'none or limited knowledge' and 5 represented 'expert'. The areas evaluated included FAIR principles, semantic web tools, ontologies, and software development.

As shown in Table 1, self-assessed expertise varied considerably across topics, which may reflect the diverse backgrounds of hackathon participants and their different levels of domain-specific knowledge. Participants who completed the self-assessment reported the highest level of expertise in FAIR principles, whereas HCP software and ontologies were associated with the lowest self-assessed knowledge. Nevertheless, except for the HCP software, at least 20% of respondents reported 'high' (4) or 'expert' (5) level in specific domains. Additionally, apart from this topic and the Knowledge/Data topic, between 10% and 20% reported an 'expert' level (5). These results indicate a considerable level of expertise across most topics and, from a group perspective, an overall strong expertise. Furthermore, each working group included at least one participant with a high to very high level of expertise (data not shown), which supported the execution of some technically complex tasks within the challenges.

TOPIC (n = 40)	MEDIAN (IQR)	SELF-ASSESSED EXPERTISE, n (%)
FAIR principles	3 (2)	**1:** 8 (20%); **2:** 5 (12.5%); **3:** 12 (30%); **4:** 7 (17.5%); **5:** 8 (20%)
Semantic web	2 (2)	**1:** 18 (45%); **2:** 9 (22.5%) **3:** 5 (12.5%); **4:** 3 (7.5%); **5:** 5 (12.5%)
Ontologies	2 (1.5)	**1:** 10 (25%); **2:** 13 (32.5%); **3:** 7 (17.5%); **4:** 6 (15%); **5:** 4 (10%)
Knowledge/data	3 (2)	**1:** 11 (27.5%); **2:** 8 (20%); **3:** 13 (32.5%); **4:** 8 (20%); **5:** 0 (0%)
(Meta)Data Standards	3 (2.25)	**1:** 12 (30.0%); **2:** 7 (17.5%); **3:** 11 (27.5%); **4:** 5 (12.5%); **5:** 5 (12.5%)
Software/scripting	3 (3)	**1:** 11 (27.5%); **2:** 6 (15.0%); **3:** 9 (22.5%); **4:** 8 (20.0%); **5:** 6 (15.0%)
HCP software	1 (2)	**1:** 23 (57.5%); **2:** 6 (15.0%); **3:** 9 (22.5%); **4:** 1 (2.5%); **5:** 1 (2.5%)

Table 1 Median, interquartile range (IQR) and percentage distribution of self-assessed expertise, based on responses collected using a five-point rating scale.

Significance is measured on a Likert scale ranging from '1 – Low expertise on topic' to '5 – Expert on topic'.

THE HACKATHON'S FOCUS CHALLENGES

The data collected when patients visit an HCP is usually stored locally (digitally). With the patient's consent, (anonymised) data can be shared with larger networks, such as national registries or ERNs. The sharing process involves exporting the data from the local HCP system to the network, where it is imported by other providers and merged with additional information. These larger knowledge bases enable reuse by various stakeholders – including researchers, patient organisations, and government bodies – under specific data access and use conditions.

This data exchange process introduces two fundamental challenges that were addressed at the hackathon. First, because medical data are sensitive, they must be shared within a secure environment that protects patient privacy and ensures they are used only under approved conditions (Challenge #1: Querying Service). Second, for the patient data to be meaningfully combined, it must be harmonised into a consistent format and structure that accounts for the different sources (Challenge #2: HCP Data Harmonisation).

However, the data flow is not just one-way. HCPs also receive information and feedback regarding diagnosis and/or treatment from other providers and ERNs. Therefore, the different systems must be able to find and understand each other automatically. This creates a third challenge: each data source must describe itself, its data, and its access rules in a clear, machine-actionable format. This problem was addressed as *Challenge #3: Metadata Description*. These three challenges are described in the next subsections.

CHALLENGE #1: WHAT COMPONENTS ARE NEEDED FOR SECURE QUERYING OF RD DATA?

The first challenge focused on designing a secure service architecture to allow data querying from multiple sources (i.e., HCPs, national registries, and ERNs) without exposing sensitive information. The overall goal was to develop a service that could receive a request, execute a query across several harmonised datasets, and return aggregated, non-identifiable results.

Therefore, the main tasks for the participants were (i) to define a clear workflow and architecture for the querying process; (ii) to create a proof of concept or a demonstration for the specific patient cohort use case; and (iii) to discuss methods to allow data to be used under controlled conditions.

To support this work, participants were provided with sample data (available as Supplementary Material; Bernabé *et al.*, 2025). These were AI-generated CSV files that had been annotated with an ontology to simulate the fully harmonised data needed for this challenge. Consequently, the group was informed that they should focus on the data exchange itself, as the harmonisation and metadata exposure aspects were to be addressed by other working groups.

The choice of using CSV files was based on their simplicity and ease of use, which was a practical consideration given the hackathon's limited timeframe. Furthermore, it is expected that any HCP system is minimally capable of exporting data in at least a CSV format. Additionally, the example files were synthetically generated using AI because the objective of the challenges was to identify solutions that could be generalised across different systems rather than tailored to a specific existing dataset or format. Using synthetic examples allowed participants to focus on the underlying methodological and technical issues while avoiding dependencies on particular data structures or governance constraints associated with real-world datasets.

CHALLENGE #2: HOW CAN HETEROGENEOUS HCP DATA BE HARMONISED USING SEMANTIC MODELS?

The second challenge focused on investigating methods to generate harmonised data from exports of HCP systems. As in Challenge #1, the exercise assumed that all systems are able to export their data into a simple CSV file. This allowed participants to focus on the task of harmonising the data itself, rather than on the complexities of extracting it in different ways from different systems. While the ideal solution would use secure Application Programming Interfaces (APIs), it is argued that the methods developed for CSV files could be adapted for such systems.

With this simplified setup, the goal was to find ways to convert different CSV files into a single, harmonised format that uses semantic web standards. The main tasks for the participants were (i) to design or reuse a semantic data model for the common use case and (ii) to develop scripts to convert various source CSVs into the standard target template. Also, a critical requirement of this challenge was mapping different diagnosis codes (e.g., ICD-10 or free text) to Orphanet codes (Orphanet, 2025), a process identified as a best practice for coding RD diagnoses by the European Steering Group on Health Promotion, Disease Prevention and Management of Non-Communicable Diseases (European Commission, 2025b).

CHALLENGE #3: WHAT TECHNOLOGIES ENABLE MACHINE-ACTIONABLE DISCOVERY OF DATA SERVICES?

The third challenge focused on a fundamental aspect of the FAIR principles: making data services easy to find and use within ERNs. The main goal was to improve how these services are described using the FAIR Data Point (FDP) (Bonino da Silva Santos *et al.*, 2025), which is a software specification for publishing standardised metadata. For the purpose of this challenge, a 'data service' was defined as the technical means by which one data source (such as an ERN registry) allows another party (such as an HCP) to access its data.

To make the task manageable, two key assumptions were made. The first was that a secure network (the output of Challenge #1) was already in place. Second, to allow participants to focus on the metadata itself rather than on software development, the FDP was chosen as the standard tool for publishing the service descriptions. Given this setup, the participants' specific tasks were to (i) review the existing FDP metadata model; (ii) identify any gaps in its ability to describe data services; and (iii) suggest improvements to make these descriptions more useful for both people and automated systems.

DOI: 10.5334/dsj-2026-029

RESULTS: PROPOSED SOLUTIONS AND TECHNICAL FRAMEWORKS

The hackathon's working groups proposed a series of complementary solutions, demonstrating a strong preference for adapting and integrating existing, tested technologies rather than building entirely new systems from scratch. Overall, the solutions are closely interrelated: the high-level architecture proposed in Challenge #1 serves as the overarching framework, while the solutions to Challenge #2 and Challenge #3 provide detailed specifications for key components within that framework, as described below.

A summarised overview of the proposed solutions and their relationship with the research questions that motivated this work is provided in Table 2, with further details presented in the following subsections. It should be noted that the following subsections refer to a variety of technologies, including software tools, scripting languages, and semantic models. As a comprehensive introduction to all of these falls outside the scope of this paper, Table 2 includes references for further exploration. Some technologies, however, are introduced briefly, as they are fundamental to understanding the proposed solutions as a whole.

Table 2 Summarisation of the main solutions proposed by the hackathon participants.

EJP RD: European Joint Programme on Rare Diseases; NCIT: National Cancer Institute Thesaurus; ODRL: Open Digital Rights Language; RML: RDF Mapping Language; SIO: Semanticscience Integrated Ontology.

CHALLENGE	SUMMARY OF KEY FINDINGS	RELATED REUSED TECHNOLOGIES
Challenge #1: What components are needed for secure querying of RD data?	A hybrid querying architecture was proposed consisting of a query trigger, central query processor, and secure data querying interfaces at data-holding institutions. Queries are authorised based on machine-readable policies, executed locally against harmonised datasets, and return only aggregated results, ensuring privacy by keeping sensitive patient data within institutional boundaries.	ODRL (Iannella, 2004) for machine-readable query policies; FAIR Data Point (Bonino da Silva Santos *et al.*, 2025) for query policies description; GA4GH Beacon Protocol (Rambla *et al.*, 2022) for query interface standardisation; FAIR Data Station (Bonino da Silva Santos, Burger and Kaliyaperumal, 2021) for query execution; Dremio (Dremio, 2025) and Ontop (Bagosi *et al.*, 2014) for querying virtualisation and translation.
Challenge #2: How can heterogeneous HCP data be harmonised using semantic models?	Heterogeneous HCP data can be harmonised through a semantic data model, supported by a transformation pipeline that ingests raw exports, performs semantic enrichment and code mapping to Orphanet identifiers, and converts the data into a standardised representation.	CARE-SM (Alarcón-Moreno and Wilkinson, 2024), SIO (Dumontier *et al.*, 2014) and Orphanet codes (Orphanet, 2025) for semantic enrichment and interoperability; RML (Dimou *et al.*, 2014) or Ontop (Bagosi *et al.*, 2014) for syntactic harmonisation and mapping; DuckDB (Raasveldt and Mühleisen, 2019), SQL and Python functions for data ingestion.
Challenge #3: What technologies enable machine-actionable discovery of data services?	Machine-actionable discovery of data services is enhanced with domain-specific resource descriptions and formalised usage conditions so that data services, access rules, and query capabilities can be discovered and interpreted automatically by both humans and software agents.	FAIR Data Point (Bonino da Silva Santos *et al.*, 2025) for resource description; DCAT (Albertoni *et al.*, 2024); and the EJP RD Metadata Model (EJP RD, 2025b) for metadata standardisation; NCIT (de Coronado, Remennik, and Elkin, 2023) and Orphanet codes (Orphanet, 2025) for machine-readable domain description in rare diseases.

SOLUTION TO CHALLENGE #1: A HYBRID QUERYING ARCHITECTURE

The group working on Challenge #1 proposed an architecture diagram to detail and organise the components necessary to enable secure querying of RD data. The diagram, depicted in Figure 1 as a simplified version for the sake of readability, was originally designed using the Archimate modelling language (Josey *et al.*, 2016) (available in the Supplementary Material; Bernabé *et al.*, 2025).

The proposed architecture consists of four main components that work together to process a query. The *query trigger* is the user's entry point to the network. A researcher uses the trigger to submit their query. Along with the query itself, they must provide the intended conditions for data reuse and information about their user profile.

The *query processor* acts as the central hub of the network. It receives the request from the trigger and first checks the requester's profile and access conditions to authorise the query against the network's policies. If authorised, the processor then finds relevant data sources

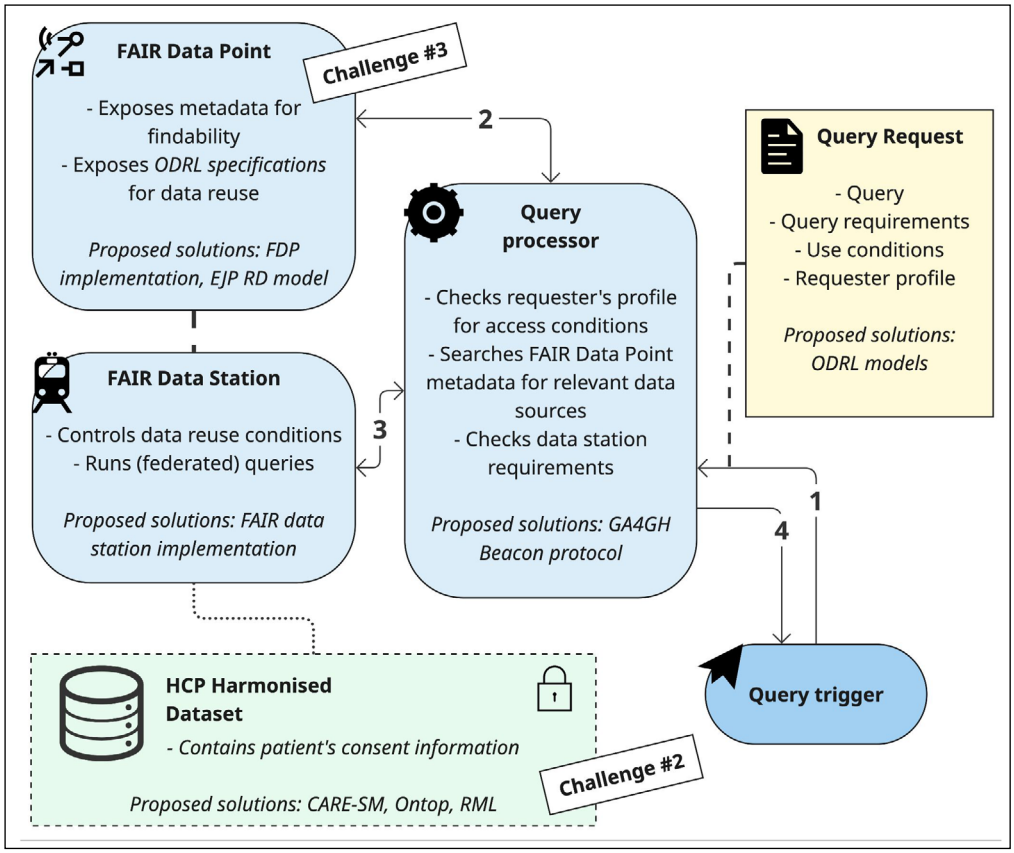

Figure 1 Proposed architecture for the network of resources containing four main components: the query trigger, the query processor, the FAIR Data Station, and the FAIR Data Point. The query request and the HCP harmonised dataset are also represented in the figure. The harmonised dataset and FAIR Data Point components are further detailed in challenges #2 and #3, respectively. The numbers on each connection show the order in which the requests are executed. Connections with arrows on both sides indicate that a response to the request is required before the flow can continue. Dotted lines show protected data access, while dashed lines show data exchange.

by searching the metadata available in their FDPs (further specified in Challenge #3). Once matching sources are found, the processor forwards the query to their FAIR Data Stations. Note that the query processor checks the allowed use conditions by comparing the information provided in the query request with that exposed by the FDPs. The group suggested making this information machine-actionable using the Open Digital Rights Language (ODRL) standard. The ODRL model is explained further in Subsection 'The ODRL model'.

The *FAIR Data Station* resides at the data-holding institution (e.g., an HCP). It receives the query from the query processor and performs a final check, comparing the query's reuse conditions against the local data's specific permissions. If the request is compliant, the Data Station executes the query on the HCP harmonised dataset and returns its results to the query processor.

The *HCP harmonised dataset*, which is the focus of Challenge #2, resides locally and is queried by the FAIR Data Station. It must also contain the necessary information regarding patient informed consent. After a query is run, the FAIR Data Station sends the aggregated results back to the central query processor. At the end of the process, the query processor gathers the results from all responding Data Stations, aggregates them into a single response, and returns it to the user via the query trigger.

This entire process, outlined by the proposed architecture, is an example of federated querying. This approach significantly enhances data protection because the sensitive raw data never leaves the secure environment of the local institution. Instead, only the query travels to the data, and only aggregated, non-sensitive results are returned to the query processor. For the implementation of the components of the architecture, the group suggested reusing existing resources from other initiatives, as further detailed in Subsection 'Leveraging existing initiatives'.

The ODRL model

ODRL is a World Wide Web Consortium (W3C) standard for creating machine-readable policies that express rights, permissions, and obligations over digital assets like data. These policies are used to automatically determine if a data access request should be granted by matching the request's parameters against the data's usage rules. When a request is received, a system can automatically compare the requester's profile and intent against the Assignee and Constraint definitions in the policy to authorise or deny the action.

To demonstrate its practical application, the group working on Challenge #1 also drafted an initial ODRL policy model for the patient cohort use case as a suggestion for how access conditions could be managed within their proposed architecture. This is presented in Figure 2.

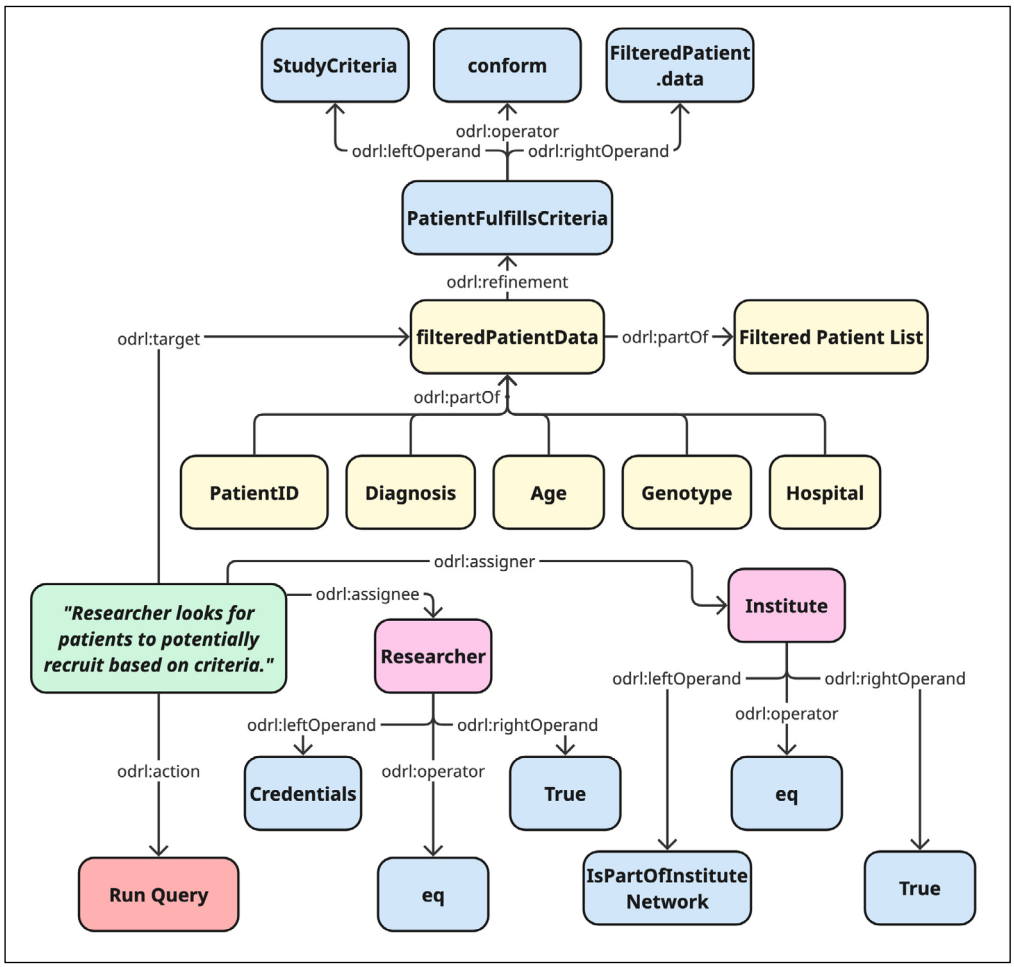

Figure 2 Example of ODRL model to be used in the proposed architecture, designed for the patient cohort use case.

Figure 2 provides a specific example of an ODRL policy model designed to govern access to a filtered patient dataset for research. The model grants a *Permission* for the *Action* 'Run Query' to an *Assignee* defined as a 'Researcher'. This permission, however, is only valid if two constraints on the researcher are met: first, that their 'Credentials' are valid, and second, that they are confirmed to be part of the 'Institute Network', which acts as the *Assigner* of the rights.

Furthermore, the policy defines constraints on the data itself. The *Target* of the query is specified as 'Filtered Patient Data'. The model shows that this data has already been created through a refinement process where only patients fulfilling the 'Study Criteria' (e.g., having provided informed consent) are included. This ensures the query only runs on data from the appropriate patient cohort.

By formally defining these conditions for both the user and the data, the ODRL model provides an automated method for enforcing complex data access rules, ensuring that sensitive information is protected while enabling responsible reuse.

Leveraging existing initiatives

To ensure better convergence and accelerate development, the working group recommended reusing components from other initiatives. This approach also increases collaboration and helps solutions mature more quickly. Based on the participants' experience, two specific projects were highlighted as key sources for reusable solutions: the Heterogeneous Semantic Data Integration for the Gut-brain Interplay (HEREDITARY) project (European Commission, 2024b) and the European Joint Programme on Rare Diseases (EJP RD) (European Commission, 2019).

The HEREDITARY project is a European initiative focused on integrating data to study the relationship between gut health and neurodegenerative diseases. Its federated analytics

framework is designed around a virtual data lakehouse architecture (Oreščanin and Hlupić, 2021), which allows complex analyses to be run across distributed sources without centralising sensitive data. Preliminary results on genomic data show the HEREDITARY project's Semantic Data Integration platform can facilitate federated querying, even across multiple, heterogeneous datasets. This capability is pivotal for cross-institutional studies, where accessing comprehensive, interoperable data while satisfying legal constraints is critical (Menotti *et al.*, 2025).

EJP RD was a large-scale initiative created to build a comprehensive ecosystem for RD research, bringing together a wide range of stakeholders to accelerate diagnosis and therapy development. An important result of this programme is the EJP RD Virtual Platform (EJP RD, 2025c), a federated infrastructure that allows researchers to discover and query RD resources (e.g., registries, biobanks) under certain conditions. The EJP RD concluded in 2024 and is now being followed up by the European Rare Disease Research Alliance (ERDERA) (European Commission, 2024a).

Building on these established projects, the group made specific recommendations for the components of their proposed architecture.

Harmonised dataset. To create a powerful and scalable data harmonisation solution, the group recommended adopting the data virtualisation architecture pioneered by the HEREDITARY project. This approach is powered by a strategic orchestration that requires two main components: a data lake platform and a query translator. The former enables data access to multiple relational sources, harmonising different schemas with virtual views and providing a single viewpoint on top of it; the latter is capable of exploiting ontology mappings to rewrite and unfold graph-based queries in relational format. A possible implementation for the data virtualisation leveraging open-source components uses Dremio as the data lake platform and Ontop as the ontology-based query translator. This setup exposes a virtual knowledge graph, enabling complex and semantic-enriched queries across all connected sources. This recommendation also aligns with the suggestions made by the group for Challenge #2 (as described in the next subsection), in which the proposed semantic model can act as the central model for this virtual layer.

FAIR Data Station and FDP. The group recommended connecting with existing development efforts that are already creating a formal specification and implementations for the FAIR Data Station concept. Additionally, the group proposed extending the existing FDP metadata model following the EJP RD Metadata Model to describe RD resources, which is consistent with the work presented by the Challenge #3 group.

Query processor. For this central component, the group suggested a hybrid approach using tested solutions from both HEREDITARY and EJP RD. The key recommendation was to use the GA4GH Beacon v2 protocol as the standardised interface for querying resources within the network. The GA4GH Beacon protocol is an international API specification standard for discovering genomic and clinical data in a privacy-protecting manner. In its latest version (Beacon v2), it allows a user to ask complex questions of a data source (e.g., 'how many female patients over 40 with a specific diagnosis do you have?') and receive a granulated response depending on the access level of the user (boolean, count, or record response) while maintaining the anonymity of the patients. In the proposed federated network, Beacon would serve as the uniform API, allowing the central query processor to communicate with all connected Data Stations using a single, shared language.

SOLUTION TO CHALLENGE #2: DATA HARMONISATION VIA CARE-SM

The working group for Challenge #2 proposed a solution centred on the Clinical and Registry Entries Semantic Model (CARE-SM) as a means to harmonise heterogeneous HCP data. This model was chosen as the standard for harmonising data because it uses the machine-readable Resource Description Framework (RDF) to represent patient information in a structured, graph-based format. The foundational structure of CARE-SM relies on the Semanticscience Integrated Ontology (SIO), which serves as its core schema and defines all concepts within the data model through its upper-class classes and properties. CARE-SM follows a design pattern in which

individual roles are realised in processes, which in turn have outputs that follow controlled attribute types (e.g., range). This pattern facilitates data integration and querying. Figure 3 shows a simplified excerpt of CARE-SM, including an example of how the clinical trial use case data is instantiated.

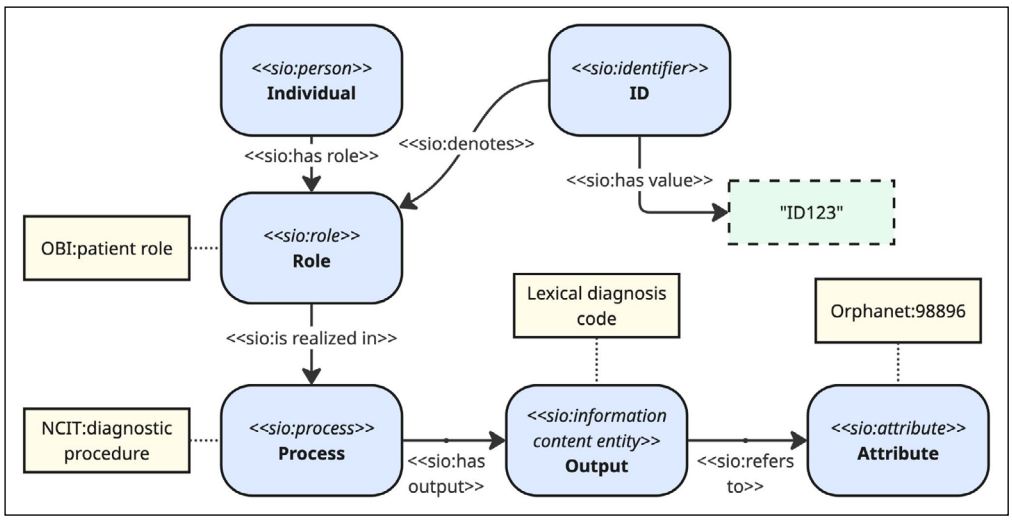

Figure 3 A simplified excerpt of CARE-SM. Elements in light blue rounded rectangles represent classes of CARE-SM, with SIO's superclasses described in *<<italic>>*. The specific ontological type for each node is shown in a yellow box, and examples of data instances are shown in green boxes with dashed borders.

To demonstrate their approach, the group developed a proof of concept of a data transformation and harmonisation pipeline with three main steps. A complete example of this pipeline was provided in a Python notebook (available in the Supplementary Material; Bernabé *et al.*, 2025). An illustration of this pipeline is presented in Figure 4:

1. **Ingestion (focus on data format):** Raw data from source CSV files is imported into a standard tabular structure using an automated tool. The group suggested DuckDB as an example of such a tool.
2. **Semantic enrichment:** The imported data is then mapped to a standard template following controlled vocabularies. This step uses a combination of approaches, including SQL functions, Python scripts for more complex logic, and even Large Language Models to identify patterns in unstructured text. Examples of these transformation approaches are available in the Python notebook.
3. **Syntactic harmonisation:** Finally, the standardised tabular data is converted into the semantically rich RDF format following CARE-SM. This is achieved using tools like Ontop or the RDF Mapping Language (RML). For instance, with Ontop, the OWL implementation of CARE-SM could be used, and mappings retrieved via SQL queries on the tabular data produced in the previous step. The Ontop endpoint would then be provided in the semantic layer of the proposed architecture from Challenge #1.

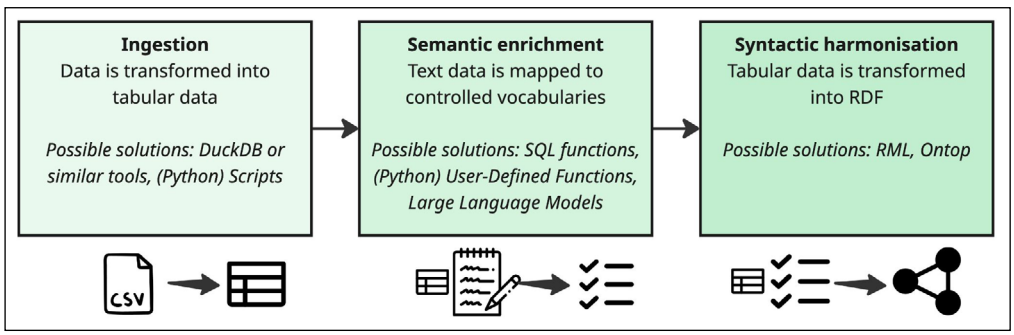

Figure 4 Illustration of the proposed pipeline for data harmonisation in Challenge #2.

The group emphasised that while automation is helpful, human expertise is essential for creating the initial mappings from a local system to the common template. Furthermore, the group suggested a two-phase, iterative implementation strategy for the solution proposed. In phase 1 (Pilot), institutions should begin by adapting their systems to harmonise a small, core set of data elements required to answer specific 'simple' use cases (e.g., the elements

listed in the clinical trial use case). This allows them to build expertise while navigating the challenging initial setup. Subsequently, in phase 2 (Expansion), initiated after the initial hurdles are overcome, the systems can be expanded to export more complex and comprehensive patient information.

SOLUTION TO CHALLENGE #3: EXTENDING THE FDP METADATA MODEL

The working group for Challenge #3 focused on describing a data service using an FDP as a means to enable machine-actionable discovery of data services. For their work, the group utilised FDP version 1.16 and developed a proof-of-concept demonstration, in which they successfully configured a test instance of FDP, demonstrating how a data service for the RD domain can be described in practice.

The group's recommendations are based on the FDP's internal structure, which uses the W3C standard Data Catalog Vocabulary (DCAT). The FDP is built on DCAT because it is the international standard for describing data catalogs, which promotes widespread interoperability. Maintaining compatibility with DCAT is therefore essential to ensure that the data services described in this project can be discovered and used by other major data initiatives. The EJP RD Metadata Model, suggested by group #2, is also aligned with DCAT.

DCAT organises metadata using a logical hierarchy. At the highest level is the *Catalog*, which acts as a container for a collection of datasets. Each *Catalog* contains one or more *Datasets*, which represent a conceptual collection of data (e.g., 'A registry of rare disease patients'). A single *Dataset* can be made available in different forms, each of which is called a *Distribution* (e.g., a CSV file and an API could both be distributions of the RD patients' data). Finally, a *Data Service* describes the specific operation that provides access to the data, such as a query endpoint or a direct download link.

First, the group provided recommendations on which ontological terms and keywords to use at each level of the DCAT hierarchy. To illustrate, examples of suggested keywords and ontology terms are described in Table 3. For instance, at the Catalog and Dataset level, they suggested using terms from the National Cancer Institute Thesaurus (NCIT) and Orphanet codes to describe key concepts such as 'Patient outcomes registry', 'Diagnosis', 'Genotype', and specific diseases (e.g., Duchenne muscular dystrophy).

LABEL	ONTOLOGY ID
Patient outcomes registry	NCIT_C119669
Patient identifier	NCIT_C164337
Diagnosis	NCIT_C154625
Age	NCIT_C25150
Genotype	NCIT_C16631
Healthcare provider	NCIT_C16666
Diseases	Orpha_98896 (example for 'Duchenne muscular dystrophy')

Table 3 Examples of ontology terms to be used to describe rare disease resources in the FDP.

FDP: FAIR Data Point.

Second, the group provided two key technical recommendations on how to describe a data service within the DCAT vocabulary:

- **Properly link the data service to the dataset:** The group recommended using the *dcat:servesDataset* property to explicitly connect the data service record to the parent dataset record. This is additional to the existing association of *Data Service* with *Distribution*, because it ensures the service is linked to the overall descriptive metadata (like dcat:theme and keywords) that are defined at the Dataset level.
- **Describe service parameters in the endpoint description:** To explain what parameters a data service accepts (e.g., for a query), the group recommended using the *dcat:endpointDescription* property. This property is the designated place to provide documentation on how other systems can interact with the service. An endpoint description is shown in Figure 5.

Figure 5 An example of a data service description as displayed in the FAIR Data Point (FDP) user interface. The 'Ontological Description' and 'Keywords' fields use standard terms to define the topic of the data being served, while the 'Endpoint URL' provides the direct machine-readable address to access the service. Finally, the 'Endpoint Description' offers human-readable instructions, such as a link to external documentation or a direct explanation of how to use the service.

Finally, the group emphasised a critical requirement for this approach: the creation of detailed documentation and guidelines. They emphasised that these guides must not only explain the technical steps for installing an FDP but, more importantly, how to populate it correctly using high-quality, well-chosen ontological terms.

Risk assessment

An additional, crucial recommendation from the hackathon conclusion session was that the implementation of any solution must be preceded by a thorough risk assessment. For instance, the working group for Challenge #2 identified several risks and their potential mitigations, which are broadly applicable to all the solutions presented in this paper:

- *Data privacy and compliance:* A primary risk is non-compliance with data protection regulations, such as the General Data Protection Regulation (GDPR), which may have significant legal and ethical implications. This risk can be addressed by enforcing strict local data processing rules and using secure aggregation protocols to ensure full compliance.
- *System performance:* A service that communicates with multiple networks may become inefficient for practical use, hindering its adoption by clinicians and researchers. To mitigate this, performance can be improved by optimising query execution, implementing data caching, and using parallel processing where feasible.
- *User-friendliness:* A system's usability is crucial for adoption. This can be improved by focusing on clear communication and accessible design. This involves providing technical documentation and usage instructions in clear, accessible language and offering this support in multiple languages to cater to a diverse user base. Furthermore, it is important to follow established user interface (UI) design principles wherever applicable to ensure the service is intuitive and predictable for all users.
- *Technical integration:* Connecting a new service to the diverse and often outdated IT systems at different institutions poses a major technical challenge. This requires designing the system in a modular and adaptable way, in close collaboration with local IT teams to create customised integration solutions.
- *Cybersecurity vulnerabilities:* Any service providing access to distributed health data can be a target for cyberattacks. Protecting against this risk involves implementing robust security measures, including end-to-end encryption, regular security audits, and strict

access control protocols. It is of note that one of the tools developed for the creation and publication of FAIR data and metadata (including FDPs) in the EJP RD project, FAIR-in-a-box (EJP RD, 2025a), is under an ongoing assessment and peer review of the cybersecurity of its components. This tool uses components from different publishers, originally posing a challenge when it comes to ensuring its complete security. Now, a pipeline has been created that automatically implements the latest security patches for all components, re-tests each patched container, and publishes the results in GitHub to ensure full transparency. An independent working group has been established to deeply monitor the security reports for at least one of the components and take action if deemed necessary and plausible.

DISCUSSION

The JARDIN hackathon presented in this paper brought together a multidisciplinary community of clinicians, patient representatives, public health experts, managers, FAIR data stewards, semantic web, and software engineering experts to explore interoperable health data exchange solutions grounded in real-world use cases. This effort underscored that building a functional data exchange ecosystem is not solely a technical problem but also a socio-technical one that requires continuous collaboration. It also demonstrated that hackathon-style collaborations are an effective way to align technical and semantic practices across institutions and to converge on implementable frameworks, such as the one outlined in Figure 1. Overall, these initiatives position JARDIN efforts as methodological models for identifying future community-driven solutions to interoperability challenges in health data exchange.

Although developed by separate groups, the solutions proposed for the three challenges are complementary and can operate collectively as an integrated system. In fact, while individual technical interventions can resolve specific bottlenecks in health data exchange, the true strength of the hackathon's proposed solutions lies in their synergistic integration. The data harmonisation process from Challenge #2 provides the standardised data that is then shared through the secure network architecture from Challenge #1. In turn, the metadata standards from Challenge #3 are used to describe the services within this network, making them discoverable and usable. Additionally, this interconnected design also allows for a coordinated implementation strategy. Individual institutions can incrementally adopt the data harmonisation and service description methods to build local expertise. Simultaneously, a dedicated group can work in parallel to establish the core network infrastructure (Challenge #1) in close collaboration with these institutions. Ultimately, this interconnected design does not just solve three separate problems; it creates a scalable, reproducible blueprint for automated data exchange.

The outcomes of the hackathon converge on key messages that are critical for the future of automated and secure exchange of health data in Europe. Beyond producing actionable solutions, the event strengthened the collective capacity of the European RD community to operationalise the FAIR principles in practice. A prominent theme was the consensus to leverage existing, validated frameworks instead of developing novel ones. The proposals to adapt existing solutions reflect a pragmatic approach that can accelerate development and ensure greater stability and interoperability of health data.

The next subsection provides recommendations for the RD community based on the key findings of the hackathon, and the following subsection lists limitations of this approach.

RECOMMENDATIONS FOR THE RARE DISEASES COMMUNITY

Prioritise incremental and practical steps. The hackathon highlighted the value of starting with lower-complexity tasks, such as harmonising CSV files. These provide tangible goals that deliver immediate value while laying the groundwork for more complex solutions such as live, API-based data exchange. This iterative approach is essential for managing complexity and encouraging adoption among diverse institutions with varying technical capabilities.

Define common data elements collaboratively. A key recommendation emerging from the hackathon is the need to rely on a comprehensive and standardised set of core data elements for RD as a foundation for data harmonisation. The definition of these elements should be

carried out collaboratively and accompanied by the selection of appropriate standards and coding systems for each variable. Within JARDIN, this effort is being addressed through a dedicated task in the Data Management WP, whose outcomes will provide the essential basis for implementing the harmonisation pipeline proposed during the hackathon.

Let the FAIR principles guide your strategy. It is observed that the FAIR principles and semantic web artefacts provide a clear roadmap for the development of all proposed solutions. For instance, using FDPs makes services findable, employing semantic models ensures data is interoperable, and defining clear access conditions makes data accessible. These principles offer the foundational guidelines for building a scalable and responsible data ecosystem, while semantic web technologies such as RDF provide proven, efficient methods for data harmonisation and exchange.

Build capacity via training. To further support and accelerate deployment, additional initiatives should be undertaken. For instance, targeted training programmes can be provided to institutions to help their staff prepare for implementing the new data transformation and access protocols. Indeed, as described in Bernabé *et al*. (2024), similar initiatives leveraged BYOD workshops to increase awareness of, expertise in, and research on the FAIR principles and FAIR-enabling software and standards.

LIMITATIONS OF THIS WORK

While the solutions proposed during the hackathon represent a significant strategic step forward, they were developed within a constrained setting and therefore relied on simplified representations of complex real-world challenges. To ensure the tasks were achievable within the limited duration of the event, several aspects of the selected challenges were intentionally scoped down. For example, the task of converting unstructured clinical information (such as free-text clinical notes) into structured formats was excluded from the harmonisation challenge, even though it is a crucial step for integrating real-world healthcare data.

In addition, the proposed architectures have not yet been extensively validated in operational environments. However, to assess their practical feasibility, an ongoing task within the JARDIN Data Management WP is focusing on testing these conceptual frameworks through pilot implementations with partner institutions across Europe. These pilot activities aim to evaluate how the proposed solutions perform in realistic institutional contexts and to collect implementation feedback that can inform further refinements. Nevertheless, it is important to emphasise that the solutions are not purely theoretical: they were designed by domain experts and build upon technologies and approaches that have already been explored and validated in previous initiatives.

Another relevant aspect only partially considered during the hackathon focuses on the scalability of the proposed solutions. For instance, in Challenge #2, the data harmonisation process relies partially on human expertise to verify the mapping of local codes to standard Orphanet codes. While this reliance on manual verification may temporarily impact the overall scalability of the solution, we maintain that human oversight remains necessary at this stage to guarantee semantic accuracy within the highly specific RD domain. Future work should investigate how other approaches (e.g., Large Language Models) can be more effectively integrated to semi-automate this mapping without compromising clinical precision.

Finally, the hackathon's design and execution inherently present certain limitations. First, although we convened a diverse group of European experts, the participant pool was not exhaustive, leaving open the possibility that a different cohort might have proposed alternative solutions. This risk is substantially mitigated, however, by grounding the proposed solutions in established initiatives, which themselves are built upon extensive input from a broad network of experts. Second, we selected the software-oriented BYOD workshop format due to the existing familiarity of both the organisers and many participants. We did not primarily evaluate alternative collaborative formats to determine if another structure might better serve these specific interoperability challenges. Lastly, the supporting materials provided (such as example CSVs and documentation templates) may have inadvertently biased the proposed solutions or caused working groups to overlook certain implementation nuances. To address this, we rely on upcoming institutional pilot projects to expose and resolve these real-world friction points.

CONCLUSION AND FUTURE STEPS

The exchange of health data across European institutions remains hindered by persistent barriers such as heterogeneous data formats, fragmented infrastructures, and limited use of common standards. While several initiatives have addressed parts of this challenge, there remains a gap in integrating these approaches into a coherent and operational architecture. The JARDIN Hackathon on Health Data Federated focused on addressing this gap by convening a multidisciplinary group of experts to experiment with concrete, complementary solutions that build upon these established efforts.

From this work, a clear strategic direction emerged: to build on existing successful solutions, proceed with practical, incremental steps, and let the FAIR principles guide all stages of development. By aligning with ongoing European initiatives and FAIR data standards, the proposed architecture provides a scalable and reproducible blueprint for automating the secure exchange of harmonised data from the point of care to research networks.

The JARDIN project will continue to collaborate with the wider community to expand the inventory of potential solutions addressing these and other related challenges. In parallel with testing the solutions identified during the hackathon, further work within the project will focus on exploring approaches for additional challenges previously catalogued. The strategies adopted will depend on the nature of each challenge (technical, organisational, or legal) and will involve the relevant experts and stakeholder groups required to address them effectively.

ACKNOWLEDGEMENTS

We are grateful to Emiliano Reynares (IQVIA) and Dennis van Gerwen (LUMC) for their invaluable insights, and to all other hackathon participants for their collaborative efforts.

FUNDING INFORMATION

This work was supported by the JARDIN – Joint Action on integration of ERNs into National Healthcare Systems project (101129863), which is co-funded by the European Union under the grant number EU4H-2022-JA2-IBA-05.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

César Bernabé orcid.org/0000-0003-1795-5930
Biosemantics Group, Human Genetics Department, Leiden University Medical Center, Leiden, the Netherlands; University of Illinois Urbana-Champaign, Urbana-Champaign, IL, USA

Daphne Wijnbergen orcid.org/0000-0002-7449-6657
Biosemantics Group, Human Genetics Department, Leiden University Medical Center, Leiden, the Netherlands

Alberto Cámara orcid.org/0000-0001-5613-9704
Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Centro de Biotecnología y Genómica de Plantas (CBGP) UPM – INIA/CSIC, Universidad Politécnica de Madrid (UPM) – Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)/Consejo Superior de Investigaciones Científicas (CSIC), Madrid, ES 28223, Spain

Karolis Cremers orcid.org/0000-0002-1756-3905
Biosemantics Group, Human Genetics Department, Leiden University Medical Center, Leiden, the Netherlands

Margarida Magalhães orcid.org/0000-0003-3253-7622
Direção-Geral da Saúde, Lisbon, Portugal

Daniela Vicentini Albring orcid.org/0000-0001-5081-9477
Bioinformatics Group, Department of Medical Biosciences, Radboud University Medical Center, Nijmegen, the Netherlands

Sergi Aguiló-Castillo orcid.org/0000-0003-0830-5733
Bioinformatics Group, Department of Medical Biosciences, Radboud University Medical Center, Nijmegen, the Netherlands

Kalia Orphanou orcid.org/0000-0003-1351-2489
Molecular Genetics Thalassaemia Department, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus

Stella Tamana orcid.org/0000-0002-3414-4972
Molecular Genetics Thalassaemia Department, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus

Maria Xenophontos orcid.org/0000-0001-5978-0193
Molecular Genetics Thalassaemia Department, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus

Laura Menotti orcid.org/0000-0002-0676-682X
Department of Information Engineering, University of Padova, Padova, Italy

Mirco Cazzaro orcid.org/0009-0006-3856-7207
Department of Information Engineering, University of Padova, Padova, Italy

Ornella Irrera orcid.org/0000-0003-2284-5699
Department of Information Engineering, University of Padova, Padova, Italy

Joëlle Thonnard
Cliniques Universitaires Saint-Luc, Brussels, Belgium

Sander van Boom orcid.org/0009-0004-1020-475X
Biosemantics Group, Human Genetics Department, Leiden University Medical Center, Leiden, the Netherlands; 4MedBox Nederland B.V., Leiden, the Netherlands

Iris C. M. Pelsma orcid.org/0000-0003-4956-4037
Biosemantics Group, Human Genetics Department, Leiden University Medical Center, Leiden, the Netherlands; Department of Medicine, Division of Endocrinology, and Center for Endocrine Tumors Leiden, Leiden University Medical Center, Leiden, the Netherlands; Rare Disease Office, Department of Quality and Patient Safety, Rare Disease, Leiden University Medical Center, Leiden, the Netherlands

Annika Jacobsen orcid.org/0000-0003-4818-2360
Biosemantics Group, Human Genetics Department, Leiden University Medical Center, Leiden, the Netherlands

Andrew Gibson orcid.org/0009-0007-8330-1990
Biosemantics Group, Human Genetics Department, Leiden University Medical Center, Leiden, the Netherlands

Veronica Popa
EURORDIS – Rare Diseases Europe, Paris, France

Mark Wilkinson orcid.org/0000-0001-6960-357X
Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Centro de Biotecnología y Genómica de Plantas (CBGP) UPM – INIA/CSIC, Universidad Politécnica de Madrid (UPM) – Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)/Consejo Superior de Investigaciones Científicas (CSIC), Madrid, ES 28223, Spain

Marco Roos orcid.org/0000-0002-8691-772X
Biosemantics Group, Human Genetics Department, Leiden University Medical Center, Leiden, the Netherlands

REFERENCES

Alarcón-Moreno, P. and **Wilkinson, M.D.** (2024) 'Take CARE of your patient data: clinical and registry entries (CARE) semantic model', *Proceedings of the 15th international conference on semantic web applications and tools for health care and life sciences (SWAT4HCLS)*. Leiden, Netherlands, February 26–29, 2024. CEUR Workshop Proceedings, pp. 116–123.

Albertoni, R. *et al*. (2024) *Data Catalog Vocabulary (DCAT) – Version 3*. Available at: https://www.w3.org/TR/vocab-dcat-3/ (Accessed: 12 November 2025).

Bagosi, T. *et al.* (2014) 'The Ontop framework for ontology based data access', in D. Zhao, J. Du, H. Wang, P. Wang, D. Ji, J.Z. Pan (eds.) *Communications in computer and information science*. Berlin, Heidelberg: Springer, pp. 67–77.

Bernabé, C. *et al.* (2025) 'The JARDIN hackathon to seek solutions to overcome technical barriers in health data exchange – Paper's supplementary material', *Zenodo*. Available at: https://doi.org/10.5281/zenodo.17599957 (Accessed: 13 November 2025).

Bernabé, C.H. *et al*. (2024) 'Building expertise on FAIR through evolving bring your own data (BYOD) workshops: describing the data, software, and management-focused approaches and their evolution', *Data Intelligence*, 6(2), pp. 429–456. Available at: https://doi.org/10.1162/dint_a_00236

Bonino da Silva Santos, L. *et al*. (2025) 'FAIR data point: a FAIR-oriented approach for metadata publication', *Data Intelligence*, 5(1), pp. 163–183. Available at: https://doi.org/10.1162/dint_a_00160

Bonino da Silva Santos, L., Burger, K. and **Kaliyaperumal, R.** (2021) *FAIR Data Train Specifications*. Available at: https://specs.fairdatatrain.org/ (Accessed: 12 November 2025).

de Coronado, S., Remennik, L. and **Elkin, P.L.** (2023) 'National Cancer Institute Thesaurus (NCIt)', *Terminology, Ontology and their Implementations*, 2, pp. 395–441. Available at: https://doi.org/10.1007/978-3-031-11039-9_17

Dimou, A. *et al.* (2014) 'RML: A generic language for integrated RDF mappings of heterogeneous data', *LDOW*, 1184.

Dremio (2025) *Intelligent Lakehouse Platform for Unified Data Access*. Available at: https://www.dremio.com/ (Accessed: 12 November 2025).

Dumontier, M. *et al.* (2014) 'The Semanticscience Integrated Ontology (SIO) for biomedical research and knowledge discovery', *Journal of Biomedical Semantics*, 5(1), p. 14. Available at: https://doi.org/10.1186/2041-1480-5-14

EJP RD (2025a) *GitHub – ejp-rd-vp/FiaB: FAIR IN A BOX*. Available at: https://github.com/ejp-rd-vp/FiaB (Accessed: 18 November 2025).

EJP RD (2025b) *GitHub – ejp-rd-vp/resource-metadata-schema: Metadata Model and Schemas for the EJP Virtual Platform*. Available at: https://github.com/ejp-rd-vp/resource-metadata-schema (Accessed: 12 November 2025).

EJP RD (2025c) *EJP-RD Resource Discovery Portal*. Available at: https://vp.erdera.org/ (Accessed: 12 November 2025).

European Commission (2019) *European Joint Programme on Rare Diseases (EJP RD) [Research project]*. CORDIS. Available at: https://cordis.europa.eu/project/id/825575 (Accessed: 29 May 2026).

European Commission (2024a) *European Rare Diseases Research Alliance (ERDERA) [Research project]*. CORDIS. Available at: https://doi.org/10.3030/101156595 (Accessed: 29 May 2026).

European Commission (2024b) *HetERogeneous sEmanData integratIon for the guT-bRain interplaY (HEREDITARY) [Research project]*. CORDIS. Available at: https://cordis.europa.eu/project/id/101137074 (Accessed: 29 May 2026).

European Commission (2025a) *European Reference Networks*. Available at: https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/european-reference-networks_en (Accessed: 18 November 2025).

European Commission (2025b) *Rare Diseases*. Available at: https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/rare-diseases_en (Accessed: 12 November 2025).

Henriques, I. de O.C., Sand, V., Albring, D.V. *et al.* (2025) 'Rare Disease Data Management across Europe: Insights from the JARDIN Survey on Data Sharing and Interoperability', *Research Square* [Preprint]. Available at: https://doi.org/10.21203/rs.3.rs-8164392/v1 (Accessed: 29 May 2026).

Hughes, L.D. *et al.* (2023) 'Addressing barriers in FAIR data practices for biomedical data', *Scintific Data*, 10(1), p. 98. Available at: https://doi.org/10.1038/s41597-023-01969-8

Iannella, R. (2004) 'The Open Digital Rights Language: XML for digital rights management', *Information Security Technical Report*, 9(3), pp. 47–55. Available at: https://doi.org/10.1016/S1363-4127(04)00031-7 (Accessed: 29 May 2026).

JARDIN Joint Action (2025a) *Data Management*. Available at: https://jardin-ern.eu/work-package/data-management/ (Accessed: 12 November 2025).

JARDIN Joint Action (2025b) *JARDIN Joint Action European Reference Networks*. Available at: https://jardin-ern.eu/ (Accessed: 12 November 2025).

Josey, A. *et al.* (2016) *An introduction to the ArchiMate 3.0 specification*. White Paper from The Open Group, 35.

Menotti, L. *et al.* (2025) 'HERO-Genomics: an ontology for integration and access of multicenter genomic data', *Zenodo*. Available at: https://doi.org/10.5281/zenodo.14936065 (Accessed: 29 May 2026).

Oreščanin, D. and **Hlupić, T.** (2021) 'Data lakehouse-a novel step in analytics architecture', *2021 44th international convention on information, communication and electronic technology (MIPRO)*. IEEE, September. Opatija, Croatia. pp. 1242–1246. Available at: https://doi.org/10.23919/MIPRO52101.2021.9597091

Orphanet (2025) *ORPHAcodes*. Available at: https://www.orphacode.org/ (Accessed: 12 November 2025).

Raasveldt, M. and **Mühleisen, H.** (2019) 'Duckdb: an embeddable analytical database', *Proceedings of the 2019 international conference on management of data*. 30 June 2019–5 July 2019. Amsterdam, NL. Association for Computing Machinery, New York, NY, United States. pp. 1981–1984. Available at: https://doi.org/10.1145/3299869.3320212

Rambla, J. *et al.* (2022) 'Beacon v2 and Beacon networks: A "lingua franca" for federated data discovery in biomedical genomics, and beyond', *Hum Mutat*, 43(6), pp. 791–799. Available at: https://doi.org/10.1002/humu.24369

TO CITE THIS ARTICLE:
Bernabé, C., Wijnbergen, D., Cámara, A., Cremers, K., Magalhães, M., Albring, D.V., Aguiló-Castillo, S., Orphanou, K., Tamana, S., Xenophontos, M., Menotti, L., Cazzaro, M., Irrera, O., Thonnard, J., van Boom, S., Pelsma, I.C.M., Jacobsen, A., Gibson, A., Popa, V., Wilkinson, M. and Roos, M. 2026 The JARDIN Hackathon to Seek Solutions to Overcome Technical Barriers in Health Data Exchange: From the Point of Care to European Registry Networks. *Data Science Journal*, 25: 29, pp. 1–17. DOI: https://doi.org/10.5334/dsj-2026-029

Submitted: 21 November 2025
Accepted: 09 June 2026
Published: 12 August 2026

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.