

The FAIR Humanitarian Data Space

Rens Kievit^{1,*†}, Joëlle Stocker^{2,3†} and Mirjam Van Reisen¹

¹Leiden University Medical Centre, Leiden University, Leiden, the Netherlands

²Department of Plant Sciences, Wageningen University & Research, Wageningen, the Netherlands

³Europe External Programme with Africa (EEPA), Brussels, Belgium

Abstract

We present a demonstration of the Humanitarian Data Space (HDS), a federated AI-ready architecture aligned with the FAIR principles that makes humanitarian datasets discoverable and queryable while keeping data ownership and control at the source in localised, federated repositories. The HDS was co-designed with humanitarian organizations and researchers working in the humanitarian domain to build trust and ensure it meets context-specific needs. The architecture of the HDS consists of three layers: a storage layer, in which each participant hosts its data in a local triplestore; a metadata layer of FAIR Data Points that publish structured metadata supporting Findability and Interoperability; and an integration layer that presents the connected datasets and can execute federated queries, given proper authorisation. Interoperability and data quality are enabled through a Common Data Model, and facilitated through six applications that let users structure and enter data through user interfaces backed by Extract, Transform, Load (ETL) pipelines. We demonstrate the HDS on a single use case involving heterogeneous humanitarian data, with two aims: (i) to test the integration of existing technologies into a complete architecture and (ii) to build trust among users in the HDS and in their ability to retain ownership and control over their sensitive data. From this demonstration we find that further research and planning is necessary to develop the Common Data Model to enable interoperability between real world datasets supported by the ETL pipelines, and to ensure algorithmic access, privacy control and security in a service layer to support the data space.

Keywords

FAIR Data, Data Space, Interoperability, Federated Querying

1. Introduction

Asymmetric power relations common in the humanitarian sector, such as between refugees and aid agencies, are often replicated in the digital sphere [1]. At the same time, usable data from the ground is scarce, especially in vulnerable regions and communities. Where data is present, its quality is low, limiting proper analysis, as well as interoperability and reusability. The lack of information leads to impunity for malevolent actors and impedes effective interventions from local, regional and international actors, leaving people in ever more vulnerable situations.

Issues with data quality and interoperability which were prominent at a large scale for research data, led to the publication of Wilkinson's reference paper on the FAIR principles a decade ago [2]. Similarly, the continued development of the semantic web has created a blueprint for how to reshape data and information to foster the extraction of insights for better decision making and EU Data Space developments provide blueprints for federated architectures. To address issues of humanitarian data, we collaborated with humanitarian organizations and researchers in the humanitarian context to design a federated, AI-ready Data Space architecture. This co-design approach is crucial in this sensitive context as it transfers agency towards domain experts and helps build trust among users. The architecture focuses on discoverability of sensitive datasets in the humanitarian domain and on enabling the reuse of that information under the conditions set by the controllers of that data. The federated structure ensures

Posters, Demos, Blue Sky, and Tutorials at SEMANTiCS 2026, Sep 2026, Ghent, Belgium

*Corresponding author.

†These authors contributed equally.

✉ r.kievit@lumc.nl (R. Kievit); joelle.stocker@wur.nl (J. Stocker); m.e.h.van_reisen@lumc.nl (M. Van Reisen)

ORCID 0000-0003-2328-4117 (R. Kievit); 0009-0000-4441-2435 (J. Stocker); 000https://www.overleaf.com/project0-0003-0627-8014 (M. Van Reisen)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

full control and ownership over the data, whilst allowing the exchange of knowledge on important humanitarian issues.

The development of federated architectures enables data sharing between many parties. However, where current data space implementations often still rely on providing direct authorizations to access or view data, we ultimately aim to implement FAIR-based approaches such as the Personal Health Train [3] or data visiting [4], ensuring sensitive data remains at its source as much as possible while still allowing it to be usable under well defined conditions.

In this demonstration, we present the work on the Humanitarian Data Space (HDS) performed in the period January to June 2026 for the European External Programme with Africa (EEPA). The HDS enables the reuse of heterogeneous, sensitive data, focusing on applications related to the protection of refugees, human trafficking along migration routes, and conflict-related sexual violence. In building the HDS, the most important principle that was followed was that data should be stored and controlled by the owners of the data and stored only locally to ensure regulatory compliance. We will demonstrate an engineering real-life experiment with data stored as RDF in localised repositories under the full control of the data owner, where computation is carried out over the sources and only aggregates are returned.

In the real world, the complete architecture at the time of writing comprises six applications, with 10 users (five organizations and five individual users) and 13 localized repositories linked to six FAIR Data Points across five countries, spanning two continents (Africa and Europe) in 2026. In this demo, we demonstrate the use of the HDS among its users, after having trained engineers into the technical elements that constitute the HDS, to show purpose and build trust that data visiting is relevant while ownership and control over sensitive data is fully retained. We focus on three of the applications with structurally valid, expert informed AI-generated (Claude Opus 4.8) mock data attributed to fictitious organizations set up in a sandbox for this experiment.¹ The goal of the demo is twofold: we aim to test the integration of separate components into the full HDS stack, and by performing the demo we aim to demonstrate the purpose of entering data in the HDS to allow knowledge generated from data visiting, while ownership and control over the data is retained in local repositories.

2. Architecture of the Humanitarian Data Space

In order to enable the reuse of data stored in local repositories under the control of their holders we identified three components that needed to be set up for a first demonstration of a FAIR Data Space in the humanitarian context based on distinct applications that create FAIR data in federated repositories:

1. Standardized and interoperable description of metadata of datasets for discoverability.
2. Interoperable semantics to enable linking of data between organizations and applications.
3. Interface for selecting desired datasets and executing federated queries.

The complete architecture of the HDS is shown in Figure 1. Below we describe the technological structure in more detail.

As mentioned above, the HDS consists of six distinct applications (these are named: RPDS, SafeLine, SafeVoice, Emergency Diary, ThinX, FAIR Data Sphere)² that each serve a specific use case in the context of human data collection. Each application contains an ETL-Pipeline (Extract, Transform, Load) to FAIRify structured data and store it as RDF in a local triplestore under the control of the deployer of the application. To enable full machine actionability, robust semantics must be set up in a federated setting. The HDS provides standard semantics centered on Incidents occurring to certain Victims by certain Perpetrators at certain Locations. Where possible, connections are made to other ontologies to enhance interoperability, including FOAF³ and Schema⁴. Each application defines a Data Input Structure and a CDM specific to the use case it was designed for.

¹The demo is accessible at <https://sandbox.humanitariandataspace.com>

²<https://github.com/vodan-development> contains all repositories with application code.

³<https://xmlns.com/foaf/spec/>

⁴<https://schema.org/>

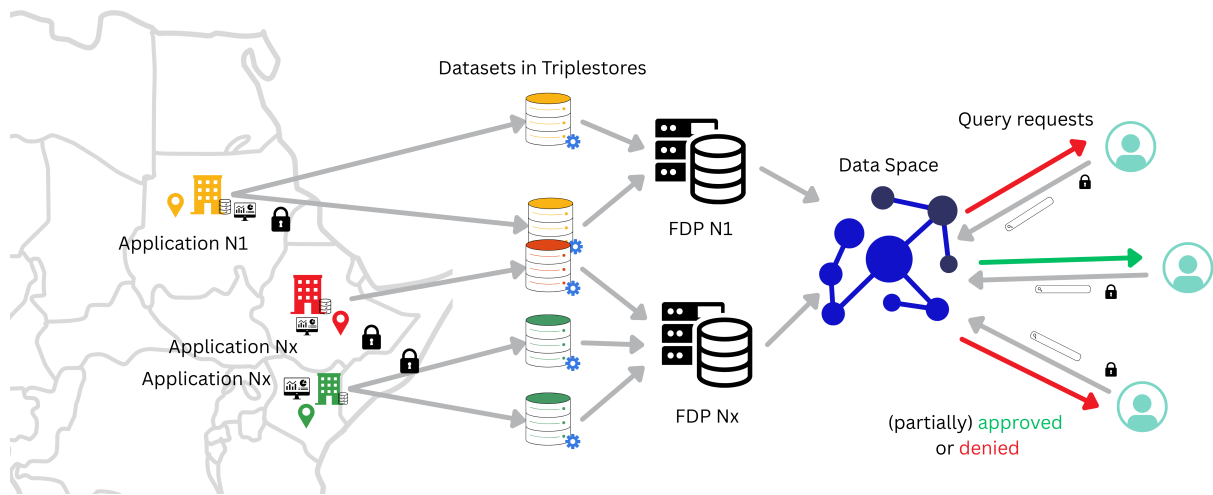


Figure 1: Visual representation of the architecture of the HDS. Data is generated through multiple applications hosted by multiple organizations. This data is stored in federated repositories under the control of the data holders. The data is made discoverable through FAIR Data Points (FDPs) that describe relevant metadata regarding content and context. The Data Space presents metadata from the connected FDPs and will allow users to send access requests and, if authorized, send federated queries to underlying data.

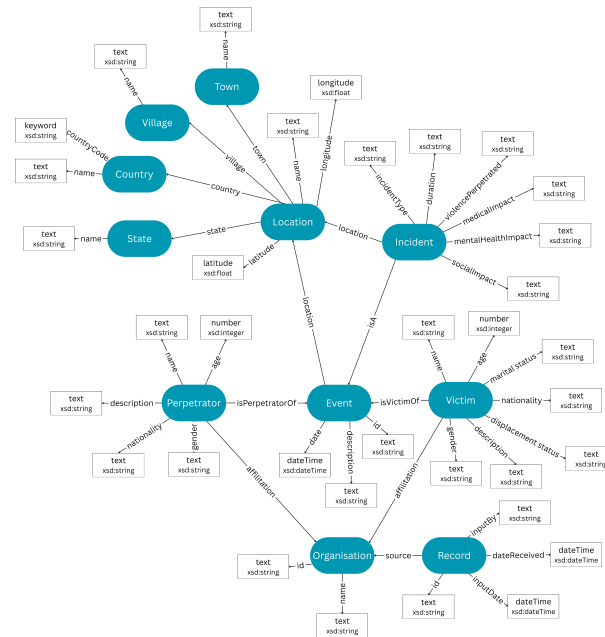


Figure 2: A visualization of part of the CDM containing the most central concepts and relationships, specifically focusing on use cases of the reporting of incidents of sexual violence.

Each organization hosting one or more applications describes these in a FAIR Data Point (FDP) [5]. The FDP provides a structured manner of describing data through standardized FAIR metadata such as authors, usage licenses and human readable descriptions. The default FDP consists of four layers to categorize and structure data, these are the FAIR Data Point, Catalogs, Datasets and Distributions. To ensure technical and semantic interoperability between deployments of FDPs between participating organizations we agreed upon a standardized implementation of these resource layers and the interpretation of the metadata fields. Specifically:

- FAIR Data Point - Describes the organization deploying applications in the HDS.
- Catalogs - Describes the specific application generating data.

- Dataset - Describes a collection of data within which data is homogeneous, this corresponds to a specific user.
- Distribution - Describes a subset of the user's data, e.g. bound by time and links to the SPARQL endpoint of the distribution for querying.

Because the metadata described in the FDPs is machine actionable and agreed upon⁵ we can extract descriptions from the federated FDPs as described by [6] and present them in a central location to allow for discovery of data. This is the core of the HDS. From the HDS, users can discover potentially relevant datasets, send access requests to this data through the provided email addresses, or if authorized send out federated queries to the remote repositories.

2.1. Privacy and security mechanisms

A crucial part in building trust in the architecture and in what happens with sensitive information are the components necessary for ensuring the safety and privacy of underlying data. In the sandbox demonstration presented here we only employ username and password authentication and defer user based authorization with repository-level read/write access to the triplestore layer. However in the real world deployment of the HDS we currently employ the following process for access control on the usage of real world data:

1. Through the HDS, users identify relevant datasets and corresponding contact points provided on the FDP (currently primarily an email address);
2. User designs a SPARQL query and describes it with purpose, scope, privacy strategies etc.;
3. The package of query and description is sent to each repository contact point for approval;
4. Queries are executed on approved datasets by a legally authorized person;
5. If necessary, privacy strategies are applied to ensure no sensitive data is shared;
6. Results are communicated back to the user and data providers.

Under privacy strategies we understand methods such as aggregation, deidentification or only sharing statistical trends, to ensure sensitive data is not made visible to external data consumers. This process is in part supported on the HDS itself by assisting in the description of the query and enabling the federated execution with authentication per repository. In the future we aim to enhance these mechanisms by integrating identity providers and federated authentication tools such as OAuth instead of per-endpoint credentials, and by integrating access- and usage control mechanisms such as those based on ODRL [7] in a connector between the data space and the datasets.

3. Demonstration of federated querying

For the purpose of the demonstration we have set up a sandbox environment describing structurally valid, expert informed AI-generated (Claude Opus 4.8) mock data of three applications: SafeVoice and Safeline hosted by five mock organizations for ten deployments, and FAIR Data Sphere hosted by a sixth corresponding to eleven repositories in total. The FAIR Data Sphere data contains situation reports describing broad events occurring in a specific region. SafeVoice and SafeLine describe detailed information on human trafficking and sexual violence incidents respectively. The data that is generated only contains identifiers to reflect a real life situation where actual names and other details might be hidden in an initial exploration like the use case presented here. This does assume an ideal scenario where all organizations share the same identifiers, this would probably not be the case in real applications.

The use case we have set out for this demonstration follows a researcher that notices a specific group (identifier GRP-A17) as being active in a situation report, they try to gather more information on this

⁵Instructions for deploying and configuring the FDPs for the HDS are documented on https://docs.superhuman.com/d/How-to-Setup-a-FAIR-Data-Point-on-Personal-Computer_dwb97Bxf0Do/Installing-FDP-on-Personal-Computer_suhKtish#_lub5aReH

group from their dataset but find a limited view (left side of Figure 3). To gain more insight into the activity of this group they use the HDS to identify other organizations active in the same region and set up a query to gather pseudonymized information from these external datasets. The query is not included in this work but can be seen in the demonstration. Using the HDS we can then execute the query (here assuming it has gone through the complete approval process as described in Section 2.1) and see the results (right side of Figure 3). In this figure, each node corresponds to an object as identified in any of the repositories, gold colored and emboldened nodes appear in at least two repositories. Each arrow pointing between two nodes represent a predicate connecting them in one repository. Areas with many connecting arrows therefore appear in many repositories. In that way, this federation not only acts as a method of enlarging the picture regarding the activities of GRP-A17 but can also act as a way of validating claims using different data sources.

For ease of demonstration the eleven repositories are hosted on two servers with the same authorization and are made discoverable through one FAIR Data Point⁶, however on a technical level this can easily be transformed to a federated setup. Currently in the HDS per-organization FDP⁷ and AllegroGraph servers are already live and can be discovered on the HDS. The AllegroGraph servers are hosted on two Hetzner "Cost-Optimized" servers with 2 vCPU and 4 GB RAM, the split is made because both servers are at capacity of RAM usage, which might influence execution times. Each repository in the demo holds 689 to 1640 triples and execution of the query on each repository took 300-1500 ms, this query execution occurs in parallel with results being concatenated for presentation in the HDS.

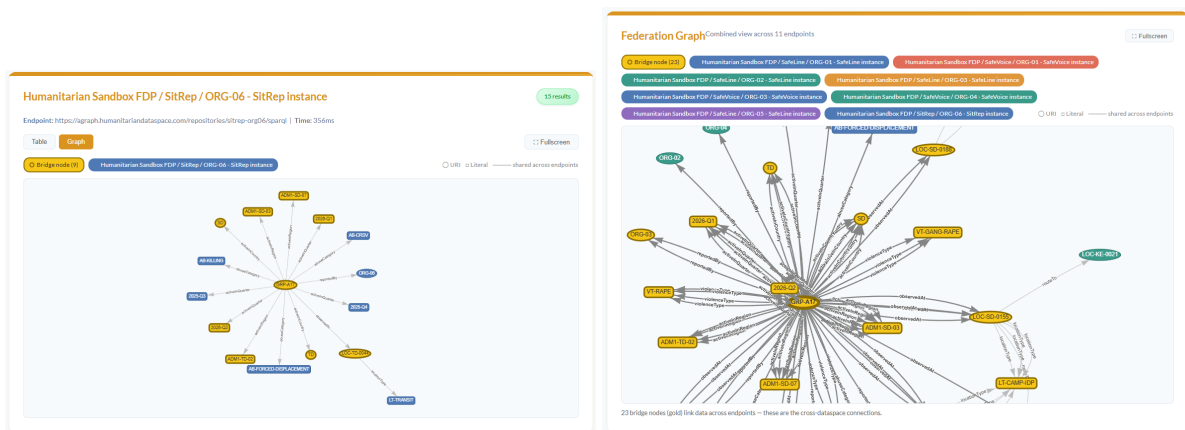


Figure 3: Results from the demonstration query as shown through the interface of the Humanitarian Data Space query view. *left:* The Knowledge Graph of information pertaining to group with ID GRP-A17 as known by Organization 6. *right:* Zoomed in portion of the Knowledge Graph of all information known on GRP-A17 to all data providers for the query.

4. Conclusion & Future work

We have demonstrated the HDS in action with structurally valid, expert informed AI-generated mock data that reflects a real-life scenario. We have found that the integration of components into a full stack for a FAIR, federated Data Space is possible, however more work is necessary to develop these into tools that are usable in real world settings. This is not only an engineering challenge, as further research and planning by experts with users is necessary to ensure that these tools align with their specific contexts. The co-design approach taken for the development of the HDS has been important in ensuring this alignment with requirements, and in collaboratively building trust in the infrastructure.

Based on the experience from the demonstrations we have identified three main areas that require further research and development: (i) the Common Data Model for shared semantics between pipelines

⁶<https://fdp.renskievit.com>

⁷See all active FDPs on <https://aun.mu.edu.et/ahds/fair-data-points/>

and applications (ii), the structured transformation of raw data into semantically rich, machine actionable data in the form of the ETL pipeline, and (iii) a service layer between the data and the outside world where techniques to enable privacy and security can be incorporated. Finally, we are testing the use of the same underlying principles in a Africa Health Data ⁸Space which promotes data sovereignty of sensitive health data, including patient data.

Acknowledgments

We would like to thank the Digital Governance in Africa programme supported by Enabel for funding under grant number BEL22006-10001. We would also like to thank the EEPA and VODAN teams who supported the development of the HDS at various levels.

Thanks to the developers of ACM consolidated LaTeX styles <https://github.com/borisveytsman/acmart> and to the developers of Elsevier updated L^AT_EX templates <https://www.ctan.org/tex-archive/macros/latex/contrib/els-cas-templates>.

Declaration on Generative AI

During the preparation of this work, the authors used Claude Opus in order to: Paraphrase and reword, Improve writing style. The authors also used Claude Opus to assist in developing software components of the HDS demonstration and in generating the mock data used in it. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] M. Madianou, Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises, *Social Media + Society* 5 (2019) 2056305119863146. doi:10.1177/2056305119863146.
- [2] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., The FAIR guiding principles for scientific data management and stewardship, *Scientific Data* 3 (2016) 160018. doi:10.1038/sdata.2016.18.
- [3] L. O. Bonino da Silva Santos, L. Ferreira Pires, V. Graciano Martinez, J. L. Rebelo Moreira, R. Silva Souza Guizzardi, Personal health train architecture with dynamic cloud staging, *SN Computer Science* 4 (2022) 14. doi:10.1007/s42979-022-01422-4.
- [4] R. Plug, Y. Liang, M. Basajja, A. Aktau, P. H. P. Jati, S. Y. Amare, G. Taye, M. Mpeza-mihigo, F. Oladipo, M. van Reisen, FAIR and GDPR compliant population health data generation, processing and analytics, volume 3127 of *CEUR Workshop Proceedings*, CEUR, 2022. URL: https://www.researchgate.net/publication/361857299_FAIR_and_GDPR_Compliant_Population_Health_Data_Generation_Processing_and_Analytics.
- [5] L. O. Bonino da Silva Santos, K. Burger, R. Kaliyaperumal, M. D. Wilkinson, Fair data point: A fair-oriented approach for metadata publication, *Data Intelligence* 5 (2023) 163–183. doi:10.1162/dint_a_00160.
- [6] R. Kievit, S. Y. Amare, M. van Reisen, J. Stocker, P. Gotlieb, Wireless technologies for disaster management and protection, in: *Handbook of Sustainable Wireless Communication Systems*, CRC Press, 2026, pp. 260–282. doi:10.1201/9781003641605-9.
- [7] W. Slabbinck, J. Rojas Meléndez, B. Esteves, R. Verborgh, P. Colpaert, May the FORCE be with you? a framework for ODRL rule compliance through evaluation, in: D. Chaves-Fraga, I. Heibi, D. Garijo, D. Collarana, A. Salatino, S. Vahdati (Eds.), *Proceedings of the Workshop on NeXt-generation Data Governance*, volume 4064 of *CEUR Workshop Proceedings*, CEUR, 2025. URL: <https://ceur-ws.org/Vol-4064/NXDG25-paper6.pdf>.

⁸<https://africahealthdataspace.com>