Scientific Data

https://doi.org/10.1038/s41597-026-08210-2

Article in Press

FAIR with FHIR: A Proposed Workflow for Retrospective FAIRification of FHIR-Compliant Health Data

Received: 10 October 2025

Accepted: 25 August 2026

Published online: 01 September 2026

Cite this article as: Heddema, F., Bernabé, C., Ihure, C. *et al.* FAIR with FHIR: A Proposed Workflow for Retrospective FAIRification of FHIR-Compliant Health Data. *Sci Data* (2026). https://doi.org/10.1038/s41597-026-08210-2

Femke Heddema, César Bernabé, Chris Ihure, Daniel Kapitan & Marco Roos

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

FAIR with FHIR: A Proposed Workflow for Retrospective FAIRification of FHIR-Compliant Health Data

Authors:
Femke Heddema[1], César Bernabé[2,3], Chris Ihure[4], Daniel Kapitan[1,5], Marco Roos[2]

[1]PharmAccess Foundation, Amsterdam, The Netherlands
[2]Human Genetics Department, Leiden University Medical Center, Leiden, The Netherlands
[3] School of Information Sciences, University of Illinois Urbana-Champaign, Champaign, IL, USA
[4]PharmAccess Kenya, Nairobi, Kenya
[5]Technical University Eindhoven, Eindhoven, The Netherlands

Correspondence:
Femke Heddema
f.heddema@pharmaccess.org

ARTICLE IN PRESS

Abstract

Despite increased digitalization, routinely collected healthcare data remains fragmented, affecting its use in research and policymaking. Although the Fast Healthcare Interoperability Resources (FHIR) standard is increasingly adopted to structure data for clinical exchange, its full potential for secondary use remains underutilized. A critical need exists to transform healthcare data into resources that can be easily discovered, accessed, combined, and reused for purposes outside of clinical exchange. Adhering to the FAIR principles enables this transformation, unlocking reuse opportunities for research, innovation, and evidence-based decision-making. Yet improving the Findability, Accessibility, Interoperability, and Reusability of FHIR-based health data requires structured workflows, a challenge that is particularly acute in resource-limited settings. To address this gap, we introduce FAIRwithFHIR: an agile, stakeholder-driven workflow for the retrospective FAIRification of FHIR-compliant resources, including the dynamic, transaction-oriented data typical of operational health systems. We demonstrate and evaluate this workflow by applying it to the MomCare dataset, a large maternal and child health dataset from Kenya. Our application illustrates a systematic process for assessing FAIRness, creating a prioritized action backlog, and implementing targeted improvements. Our findings highlight that FAIRwithFHIR can significantly enhance FAIR maturity, facilitating the ruse of routine implementation data for cross-disciplinary research. Critically, our stakeholder review showed that technical compliance alone is insufficient: realising reuse also requires a communication layer that makes the value of FAIR improvements evident to data (re)users. These findings underline the opportunity to apply retrospective FAIRification to FHIR-compliant datasets, as a workflow transferable across health domains, to drive data reuse between implementation and research communities.

ARTICLE IN PRESS

Introduction

Global health systems are under increasing pressure to manage rising care burden and growing care complexities with constrained resources. These pressures are particularly pronounced in maternal, neonatal- and child health (MNCH), where persistent gaps in outcomes continue to reflect structural inequities. Despite significant progress in the reduction of global maternal mortality since 1985, disparities remain especially persistent in Sub-Saharan Africa (SSA): Kenya recently reported a maternal mortality rate (MMR) of 594 per 100,000 live births, surpassing the rate reported four decades prior, with an estimated fifth of the reported maternal and child deaths in the country classified as preventable.[1] Evidence from data-driven interventions shows that well-structured health data can and has supported better clinical decision-making, resource allocation, and public health planning in this domain.[2–5] These factors together drive demand for resource-efficient access to health data that can be (re)used to identify inefficiencies, monitor outcomes, and guide care improvement decisions.[6–9]

The high digitization of the health systems worldwide has fundamentally transformed the scale and scope of available health data. Electronic health records, claim management systems, and digital clinical registries continuously generate detailed information on patient demographics, care utilization, diagnoses, treatments, and outcomes. There is an opportunity to (re)use this real-world data for health improvements. However, studies have consistently shown that health data is complex to access, link and interpret by stakeholders across the health system.[10] Without coordination at a national or network level, adoption of digital tools comes with inconsistent standardisation of data collection and variable infrastructure, factors that more frequently demonstrate in health systems with limited data governance capacity.[11] Widespread data implementations seem predominantly driven by primary use requirements, such as clinical workflows, billing, and regulatory compliance, with data reuse a secondary consideration. These challenges are particularly observed in MNCH care, where data is collected and used at multiple time points along the care continuum, often by different types of health workers operating within disconnected information systems.[12]

The recent adoption of interoperability standards is a response to this need for coordination across the health system. Kenya, like many other countries, has recently mandated the adoption of the HL7 Fast Healthcare Interoperability Resources (FHIR) standard for structuring and exchanging health data.[13] FHIR is designed for transactional data exchange in clinical workflows, facilitating real-time data sharing of electronic healthcare data across primary care processes by defining resources representing clinical concepts and specifying how these resources are encoded and transmitted.[14] The standard is also increasingly applied beyond operational workflows: initiatives such as SMART on FHIR, the Vulcan FHIR Accelerator, and the HL7 Clinical Research working group have developed tooling and implementation guides aimed at making FHIR-structured data available for secondary research purposes.[15–17] The FHIR standard is flexible in its design, mandating structural interoperability while positioning provenance, licensing and metadata as optional extensions. This flexibility also seems to create a dependency on the implementation guide or data model for adoption of supporting elements that are relevant for

ARTICLE IN PRESS

data reuse. Adoption of the FHIR standard may therefore still prove insufficient for the re(use) of this real-world data for health improvements.

A standardized framework for describing and evaluating whether data are ready for reuse has been extensively described in the scientific domain, and conceptualized as Findable, Accessible, Interoperable and Reusable (FAIR) data.[18] The FAIR principles were defined to maximize the benefit of generated data and allow long-term automated and controlled use of data across platforms and stakeholders, with an emphasis on machine-readability. Where FHIR mainly provides the how through a technical standard, FAIR provides the why, setting a normative standard for data reuse that includes that data should carry persistent identifiers, rich metadata, explicit licensing, and provenance information sufficient for reuse: Elements that remain optional when adopting FHIR. While FAIR is widely recognized as important guidance for optimizing data management, it has also been criticized for its lack of implementation guidance and inconsistent interpretation.[19]

Crucially, most work that assesses and examines FAIR data implementations in healthcare has focused on health data that is captured in biobanks, clinical trial outputs, and genomic repositories. As such, "FAIRification" workflows, structured approaches to improve FAIR adherence of a dataset, generally start with the prospective design of data collection processes and prioritize (meta)data standards relevant to the research community.[20,21] For data that is generated outside these scientific workflows, and is constrained by a specific interoperability standard, implementation guidance remains scarce. As a result, for those aiming to make FHIR-based data available for research purposes, it remains complex to estimate and outline the implementation effort required.[22,23]

This paper draws on a concrete implementation case to illustrate the challenge of (re)using real-world data for health system improvements. The MomCare maternal health program, operating across urban and rural areas of Kenya between 2017 and 2023, captured health-related data for approximately 30,000 participants to operationalise claims reimbursement and care quality improvement through bundled digital services, financial incentives, and structured risk assessments.[24] The operational data represents a valuable source of longitudinal health information for a population for whom such data are rarely available, with the program infrastructure tracking maternal and neonatal outcomes across the continuum of care. As such, the dataset was identified as a candidate for reuse in a meta-analysis describing health-related outcomes across the African continent (the HE2AT study), a secondary purpose for which it had not been prospectively designed. While the data was adapted to FHIR standards over the course of the program to meet Kenyan mandated health data exchange regulations, limited data management efforts had been made to enable research reuse. Despite the availability of established FAIRification frameworks[20,21,23,28] and an expanding literature on FHIR implementation,[11,15–17,25] we encountered a significant gap in practical guidance describing how FAIR principles are applied to datasets that have already been obtained and transformed to a mandated interoperability standard.

ARTICLE IN PRESS

This paper addresses this gap by providing practical guidance to retrospectively identify FAIRification needs and FAIRify datasets that have adopted FHIR as an interoperability standard. Specifically, it investigates the FAIR adherence of a real-world MNCH FHIR-based dataset collected in a low-resource setting, documents the practical steps taken to improve its FAIR adherence, and generalizes these steps into a reusable FAIRification workflow: FAIRwithFHIR. With this work it aims to provide practical steps that facilitate the use of operational data sources for longitudinal health research purposes. Throughout this paper, we use *reuse case* for the concrete secondary-use scenario that guides FAIRification, following the goal-oriented FAIRification planning method GO-Plan, and *data reusers* for the researchers and systems that access the dataset for this secondary purpose, distinct from the primary users who record and use the data during care delivery and claims processing.[28]

Method

Our methodology was designed as a multi-stage process to systematically assess, enhance, and validate the FAIRness of the MomCare dataset. First, we applied an existing FAIR assessment framework to the dataset to define a FAIRification backlog. We then reviewed existing FAIRification workflows against this backlog and, where they fell short, designed a new workflow: FAIRwithFHIR. Finally, we applied FAIRwithFHIR to the MomCare dataset and evaluated the outcome through a FAIR re-assessment and a review with data reusers.

Data Collection

Data were obtained from the MomCare maternal health program, implemented across urban and rural areas of Kenya between 2017 and 2023. The dataset comprises health-related claims and survey data from approximately 30,000 enrolled mothers, collected to operationalise claims reimbursement and care quality improvement. Over the course of the program, data were structured according to HL7 FHIR standards to comply with Kenyan mandated health data exchange regulations, using an extended International Patient Summary profile as the foundational data model (IPS+ FHIR profile; the profile specification is available in the project repository).

Ethics Statement

All data used in this study were de-identified prior to analysis through a three-step process, redacting any personal identifier information in the patient resources, pseudonymizing the persistent identifiers to prevent reidentification of a person within the operational system and perturbing event dates to prevent identification based on gestational events. The pseudonyms were generated and assigned by the MomCare data controller (PharmAccess) as part of this de-identification process; the linkage between pseudonyms and operational identifiers is stored securely within the operational environment and was not accessible to the authors during analysis. Ethical approval for the (re)use of the MomCare data for this study was obtained from the Amref Health Africa Ethics and Scientific Review Committee (ESRC; IORG0007315,

ARTICLE IN PRESS

FWA00018713), approval number P1649-2024. Written informed consent was secured from all participants upon enrollment for the MomCare program.

FAIR assessment

FAIR assessment model

We used the Research Data Alliance (RDA) FAIR Data Maturity Model to assess the FAIRness of the FHIR-based MomCare dataset. Our selection of this specific framework was driven by two primary motivations. First, the RDA framework has been previously used to assess the use of HL7 FHIR for implementing the FAIR guiding principles.[25] Second, these indicators were designed to be used by diverse types of stakeholders (e.g. researchers, data stewards, policy makers, funders). Thus, by adopting a framework that standardizes expectations across different groups, we ensured that our assessment captured both technical gaps and broader requirements for data reuse, an important and desired aspect for the context of MomCare. The results of this initial assessment process aimed to form a significant contribution in themselves, as we hypothesize that the identified shortcomings are representative of many similar datasets built on the FHIR standard.

Assessment process

The RDA model uses a set of core assessment indicators to measure FAIRness and assigns priorities (essential, important, useful).[26] Three members of the study team (FH, CB and DK) independently evaluated the presence of each of the RDA-defined FAIR indicators in the dataset, scoring each indicator as not present, partially present or fully present and providing a supporting rationale for each score. An indicator was scored as fully present when it was satisfied and verifiable in the MomCare materials for all relevant (meta)data elements, as partially present when it was satisfied for only part of the relevant elements or was present but not documented in a findable or machine-actionable way, and as not present when no supporting evidence could be found. Given the ordinal nature of this scale, the individual scores were not averaged; discrepancies between assessors were resolved through consensus discussion, resulting in a single agreed rating per indicator. This consolidated assessment of the dataset's adherence to the FAIR principles was accompanied by a detailed "FAIRification backlog," consisting of a prioritized list of specific actions required to improve the dataset's FAIR compliance. The backlog provided a set of requirements that directly informed the workflow design and the FAIRification phases.

FAIRification workflow

Design process

To address the requirements outlined in the FAIRification backlog, we designed an agile workflow for iterative implementation. The design of this workflow was based on established frameworks from the literature, in particular the generic FAIRification workflow by Jacobsen *et al.*, the FAIR in Action framework by Welter *et al.* and the GO-Plan FAIRification planning method by Bernabé *et al.*[20,21,28] and further refined through close consultation with subject matter experts in FAIR

ARTICLE IN PRESS

implementation and data stewardship. The decision to use an agile model was driven by the nature of the assessment's outputs, which highlighted the need to address distinct FAIR aspects. Because the backlog items address largely distinct FAIR aspects, they can be scoped as independent modules and distributed across contributors with different expertise, allowing several modules to be worked on in parallel. The agile approach adds the iterative element: modules are delivered in short cycles, and the backlog is re-prioritized between cycles as resources, insights and reuse requirements evolve.

Workflow application

We first validated the workflow itself by applying it to the MomCare dataset. This practical application served to test its feasibility, efficiency, and ability to systematically guide FAIRification tasks.

Output assessment and user evaluation

To assess the added value of the resulting FAIRified MomCare dataset, we engaged key stakeholders. Identified data reusers from the selected study group within the African Open Data Science (DS-I) Consortium were provided with the enhanced dataset and asked to assess it using the RDA FAIR indicators, and review how the improvements in FAIR compliance directly supported and facilitated their research activities. These evaluators were deliberately not the same people who performed the baseline assessment and re-assessment, which were conducted by the study team as described under Assessment process. The study team assessments required detailed knowledge of the dataset’s internal structure, whereas the user evaluation was designed to test whether the implemented improvements were recognizable and valuable to external data reusers approaching the dataset without such prior knowledge.

ARTICLE IN PRESS

ARTICLE IN PRESS

Results

We present the initial FAIRness assessment of the FHIR-based MomCare dataset and the agile, iterative FAIRification workflow subsequently designed, applied and evaluated to address its findings.

FAIR maturity of the MomCare dataset

The outcomes of the RDA-model based FAIR assessment of the MomCare dataset are visualized in Table 1. The scores indicate the presence of gaps in FAIR maturity for essential FAIR aspects. For each FAIR principle, the supporting rationale of the scores is described in further detail. The assessment concerned the MomCare dataset as implemented (FHIR R4 resources conforming to the IPS+ profile, served from the program's FHIR server) rather than the abstract FHIR specification. Because several FHIR capabilities that support FAIRness are optional, we describe for each principle both what was observed in the MomCare dataset and, where relevant, the unused FHIR mechanism that explains the gap.

Findability. While the FHIR standard provides a foundation for findability, significant gaps existed in the MomCare dataset. Each MomCare resource carries a FHIR resource ID (F1) that is globally unique only in combination with the server URL; this identifier is not persistent, as any change to the server URL renders it obsolete. Our backlog therefore included assigning a globally unique, persistent URL to the MomCare FHIR server itself.

While all MomCare resources contain the FHIR metadata element (F2), its content was not rich enough for discovery: the MomCare metadata lacked essential information such as provenance, theme and ownership. The corrective action was to enrich the metadata, ideally by developing and hosting dedicated metadata.

Although FHIR supports explicit data–metadata links via the Reference datatype (F3), these are optional and were not used in the MomCare dataset, leaving data and descriptive information disconnected; we identified the need to create and maintain these links explicitly.

For search and indexing (F4), the MomCare metadata was not registered in any searchable catalogue, and critical (meta)data elements were not annotated with ontologies or controlled vocabularies. The native FHIR search framework enables queries on a known server but defines no registration workflow that would make the dataset discoverable elsewhere; the proposed action was therefore to register the dedicated metadata in a community catalogue and to enrich it with ontologies to increase semantic depth.

Accessibility. While the technical protocols for accessing the MomCare data are open, its specific access conditions were not explicitly defined. The FHIR RESTful API and standard HTTP protocols are open and universally implementable (A1).[27] However, the access conditions for the MomCare data were not described in its metadata: role-based access control was in place operationally, but a prospective data reuser could not learn the conditions for obtaining access from the (meta)data

ARTICLE IN PRESS

itself, as FHIR does not require access conditions to be embedded in metadata. A key action is therefore to make these conditions explicit in dedicated metadata records.

We also identified a critical issue with long-term metadata availability (A2): because the MomCare metadata was captured within the FHIR resources themselves, it ceases to exist if a resource is deleted. The required action is to make metadata available independently of the primary data resource; complementary adoption of a separate metadata standard such as DCAT can support this persistence.

Interoperability. In the MomCare dataset, semantic integration was limited because several optional FHIR capabilities had not been implemented. All data was captured in the machine-readable HL7 FHIR v4 JSON format, satisfying the knowledge-representation indicators (I1). In this format, coded values carry their vocabulary and code within a CodeableConcept, so the information required for semantic linkage is present in principle; however, these codes are literal values inside FHIR-specific structures, and interpreting them requires tooling that understands the FHIR information model. Moreover, of the serializations defined by the FHIR standard (XML, JSON and RDF), only JSON was exposed, and the optional annotation of coded values with IRIs from external ontologies had not been used, while our reuse case required linking MomCare records to resources outside the FHIR ecosystem. We therefore identified the need to additionally expose the FHIR-defined Resource Description Framework (RDF) serialization, in a form that data reusers can consume without FHIR-specific tooling. In RDF, coded values are expressed as resolvable IRIs of the vocabularies they are drawn from: a SNOMED CT-coded pre-eclampsia diagnosis in MomCare, for example, carries the same IRI as any external repository that annotates its data with SNOMED CT, so that records can be linked across datasets through these shared vocabulary identifiers. This linkage complements, rather than replaces, schema-level mappings between data models.

ARTICLE IN PRESS

While FHIR provides a formal mechanism for binding data elements to controlled vocabularies (I2), this is not mandated in the general specification or the IPS implementation guide, and several critical MomCare data elements were consequently captured without formal vocabulary bindings; the action was to enrich the FHIR resources by formally including ontologies in data annotation.

A significant limitation was cross-data referencing (I3). The FHIR Reference datatype links individual FHIR resources to one another, within or across FHIR servers, but it is not designed to express qualified, dataset-level references to related external resources — such as the relation between the MomCare dataset and the cohort studies or climate datasets it can be combined with — and no such references were present in the MomCare (meta)data. Our proposed action was to express these qualified references at the metadata level, using relation properties from established vocabularies (e.g., DCAT and Dublin Core) in the dedicated metadata record.

Reusability. Our assessment identified significant documentation gaps. For licensing (R1.1), FHIR technically enables licensing attributes but does not mandate them, so this information was missing from the MomCare distribution. Provenance (R1.2) showed the same pattern: FHIR defines a 'provenance' resource but does not require it, and it was absent. For domain-relevant community standards (R1.3), the MomCare data itself complies with community standards — the FHIR specification and the IPS-based profile — but no community standard was used for describing the dataset as a whole: neither the FHIR core specification nor the IPS implementation guide adopted for MomCare prescribes a dataset-level metadata standard, limiting consistent description and interpretation of the dataset across contexts.

The resulting backlog with actions derived from our assessments are summarized in Table 2.

Table 1. Summarized overview of the FHIR-FAIR adherence assessment following the RDA model.
Rows are marked with ~ Partial, ✗ None or ✓ Full to denote aspects of the FHIR-based dataset that fully align with FAIR principles. ✓Full: Indicator is readily available in the MomCare materials. ~ Partial: Available within the MomCare materials, but difficult to find or interpret. ✗None: Not available in the MomCare materials; extensions or alternative solutions are needed. Element type: M = Metadata-related indicator; D = Data-related indicator; M/D = Applies to both data and metadata.

PRESS

Indicator	FAIR maturity indicator description	FAIR adherence
Findable		
F1-01M/D	(Meta)data are identified by a persistent identifier.	✗ **None**
F1-02M/D	(Meta)data are identified by a globally unique identifier	~ **Partial**
F2-01M	Rich metadata are provided to allow discovery	~ **Partial**
F3-01M	Metadata include the identifier for the data	~ **Partial**
F4-01M	Metadata are offered in such a way that it can be harvested and indexed	~ **Partial**
Accessible		
A1-01M	Metadata contain information to enable the user to get access to the data	~ **Partial**
A1-02M	Metadata can be accessed manually (i.e., with human intervention)	✓ **Full**
A1-02D	Data can be accessed manually (i.e., with human intervention)	✓ **Full**
A1-03M	Metadata identifier resolves to a metadata record	✓ **Full**
A1-03D	Data identifier resolves to a digital object	✓ **Full**

A1-04M/D	(Meta)data are accessed through standardised protocol	✓ **Full**
A1-05D	Data can be accessed automatically (i.e., by a computer program)	✓ **Full**
A1.1-01M	Metadata are accessible through a free access protocol	✓ **Full**
A1.1-01D	Data are accessible through a free access protocol	✓ **Full**
A1.2-01D	Data are accessible through an access protocol that supports authentication and authorization	~ **Partial**
A2-01M	Metadata are guaranteed to remain available after data are no longer available	✗ **None**
Interoperable		
I1-01M	Metadata use knowledge representation expressed in standardized format	✓ **Full**
I1-01D	Data use knowledge representation expressed in standardized format	✓ **Full**
I1-02M	Metadata use machine-understandable knowledge representation	✓ **Full**
I1-02D	Data use machine-understandable knowledge representation	✓ **Full**
I2-01M	Metadata use FAIR-compliant vocabularies	~ **Partial**
I2-01D	Data use FAIR-compliant vocabularies	~ **Partial**
I3-01M	Metadata include references to other metadata	~ **Partial**
I3-01D	Data include references to other data	~ **Partial**
I3-02M	Metadata include (qualified) references to other metadata	✗ **None**
I3-02D	Data include (qualified) references to other data	✗ **None**
Reusable		
R1-01M	Plurality of accurate and relevant attributes are provided to facilitate reuse	~ **Partial**
R1.1-01M	Metadata include information about the license under which the data can be reused	✗ **None**
R1.1-02M	Metadata refer to a standard reuse license	✗ **None**
R1.1-03M	Metadata refer to a machine-understandable reuse license	✗ **None**

R1.2-01M	Metadata include provenance information according to community-specific standards	✗ **None**
R1.2-02M	Metadata include provenance information according to a cross-community language	~ **Partial**
R1.3-01M	Metadata comply with a community standard	✗ **None**
R1.3-01D	Data comply with a community standard	✓ **Full**
R1.3-02M	Metadata are expressed in compliance with a machine-understandable community standard	✗ **None**
R1.3-02D	Data are expressed in compliance with a machine-understandable community standard	✓ **Full**

Table 2. Overview of FAIRification backlog with regards to MomCare's FHIR-compliant data and the RDA-model assigned priority.

FAIR indicator	*FAIRification backlog*	*RDA priority*
F1	Link identifiers to persistent identifiers (e.g., W3ID)	Essential
F2	Include additional metadata based on community needs (e.g. transformation provenance, identifier logic)	Essential
F3	Define persistent link between metadata and data	Essential
F4	Define indexing workflows and enrich metadata with controlled vocabularies and ontologies	Essential
A1	Specify and embed explicit and standardized access protocol in metadata	Essential
A2	Ensure persistence of metadata beyond FHIR server by publishing independent metadata	Essential
I2	Extend mapping to relevant community standards and formats	Important
I3	Define minimum included references to other (meta)data	Important
R1/R2	Define in profiles and implementation guide what provenance and licensing attributes are required	Essential
R3	Ensure compliance with metadata community standards	Essential

FAIRwithFHIR workflow design

We propose FAIRwithFHIR, a tailored FAIRification workflow to improve FAIR maturity for FHIR-compliant health data (Figure 1). It builds on the phases of the FAIRification framework defined by Jacobsen (2020) and Welter (2023), and incorporates parts the FAIRification planning method proposed by Bernabe (2023).[21,28] It shows FAIRification as an iterative and cyclical process, designed to align FAIRification efforts with evolving resource constraints and reuse priorities.

ARTICLE IN PRESS

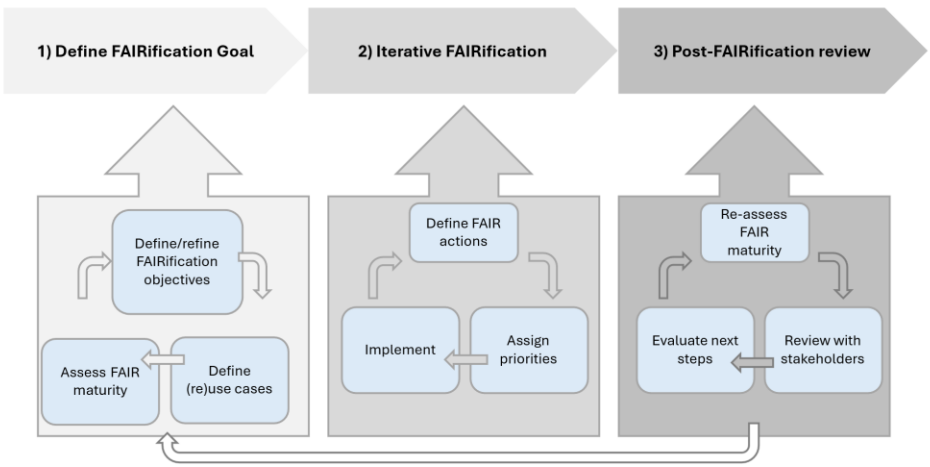

Figure 1. FAIRwithFHIR framework. The framework comprises three high-level phases of FAIRification: Define Goals, Implement Iteratively and Review, with each containing internal sub-stages and implementation workstreams.

The FAIRwithFHIR workflow consists of three iterative phases, each of which is further decomposed into specific steps. The first phase focuses on defining FAIRification goals and setting clear, actionable objectives for secondary data reuse. The second phase, iterative FAIRification implementation, translates the objectives and assessments into a prioritized list of FAIR actions. The third phase, Post-FAIRification Review, validates improved FAIR adherence through stakeholder evaluations. Feedback gathered during this review is then used to refine future goals and guide subsequent workflow cycles. The specific steps of these phases are described in the next subsections.

Although Phase 1 and Phase 3 both combine a FAIR assessment with stakeholder consultation, they serve distinct purposes and are therefore kept separate. Phase 1 is formative: it establishes a baseline in the absence of any FAIRification work and translates the reuse needs of stakeholders into objectives, answering the question of what would make the data reusable. Phase 3 is summative: it verifies against that baseline whether the implemented actions achieved their intended effect, and asks stakeholders whether the delivered changes are recognizable and useful. The two phases consequently differ in their inputs (no baseline versus a baseline plus implemented actions), in the questions posed to stakeholders, and in their outputs (a FAIRification backlog versus a validated improvement and refined goals). In subsequent cycles, the review output of Phase 3 feeds the goal definition of the next Phase 1, which is where the two phases connect; in our application, this distinction is visible in the shift from the reuse-driven objectives of the first cycle (Table 3) to the communication-focused objectives that emerged from the stakeholder review (Table 5).

ARTICLE IN PRESS

Phase 1: Define FAIRification Goals

Phase 1 establishes the direction for the FAIRification effort. It begins by defining an initial set of objectives that articulate the motivation for making the data FAIR; these are preliminary and refined through the two subsequent steps.

The second step is to define reuse cases, which is conducted in close collaboration with potential data reusers. This step is essential as it grounds the FAIRification process in practical, real-world applications. By understanding precisely how stakeholders intend to (re-)use the data, it is possible to make informed implementation decisions, such as selecting the most relevant community standards or prioritizing specific controlled vocabularies.

Subsequently, a formal assessment of the FAIR maturity level of the target data is performed. This provides a quantitative baseline of the dataset's current state, identifying specific technical gaps that need to be addressed.

These activities form a feedback loop: the needs from the reuse cases and the gaps revealed by the assessment iteratively refine the initial objectives, until stakeholders agree the resulting goals are technically sound and aligned with the community's reuse requirements.

Phase 2: Iterative FAIRification Implementation (FHIR-to-FAIR)

The second phase focuses on the technical implementation of the objectives defined in Phase 1. This phase operates as an agile, cyclical process designed for continuous improvement. The cycle begins with defining FAIR actions, where the objectives from the FAIRification backlog are broken down into specific, manageable tasks.

The defined actions are then prioritized to sequence the work — based on task dependencies, potential impact on data usability, and resource availability — after which the team proceeds with implementation.

Phase 3: Post-FAIRification Review

The final phase of the workflow is dedicated to evaluation and feedback. After implementation is completed, the resulting dataset undergoes a FAIRness re-assessment. This provides an understanding of the improvements made and verifies that the technical actions had their intended effect.

This assessment is subsequently augmented by a review with stakeholders. In this step, the FAIRified dataset is presented to the identified data reusers. Their feedback is essential for understanding the practical impact of the changes. They evaluate whether the improved FAIR level enhances the data's utility and ease of reuse for their specific research questions.

Insights from the re-assessment and the stakeholder review are synthesized to evaluate next steps, directly informing the refinement of FAIRification goals for the next workflow iteration.

ARTICLE IN PRESS

Workflow application

This section details the application of our proposed FAIRification workflow to the FHIR-based MomCare dataset. The results that follow represent the cumulative output of several iterations through the workflow's phases.

Phase 1: Define FAIRification Goals

Initial FAIRification objective. The initial FAIRification objective for the MomCare dataset was to enhance its adherence to the FAIR principles, thereby unlocking its potential for secondary research use, such as for maternal and child health improvements. This was motivated by the fact that the longitudinal data collected by the program held significant potential for addressing broader public health questions beyond its initial purpose of claims management.

Reuse case. To ground our FAIRification efforts in a concrete application, we selected a specific reuse case: a meta-analysis on the impact of heat on pregnancy outcomes, coordinated by the Heat and Health African Transdisciplinary (HE2AT) center.[29] This research consortium project aims to inform wider health policy by reusing "data from cohorts and trials among pregnant women and neonates conducted across sub-Saharan Africa since the year 2000, analysing relationships between heat exposure and outcomes (preterm birth, birth weight and pre- eclampsia) to inform quantification of heat-related disease burden and pilot a district-level climate change indicator."[29]

The HE2AT research group, composed of both domain and technical experts, closely collaborates with the eLwazi Open Data Science Group, both of which are part of the DS-I Africa consortium.[30] This consortium supports the FAIR use of research datasets for meta-analyses and makes the metadata of these datasets available by registering them on the eLwazi platform.[31] This research team set-up provides an opportunity to test the effects of the FAIRification of the MomCare dataset and evaluate its applicability in a setting that is transdisciplinary yet relevant to MNCH.

FAIRness assessment. The FAIRness assessment was conducted using the indicators from the RDA FAIR Data Maturity Model (Table 1). The findings were then analysed in the context of the stakeholder-defined reuse objectives to formulate a targeted set of FAIRification objectives, as summarized in Table 3.

The resulting FAIRification objectives were designed to directly translate the high-level reuse goals into concrete technical actions based on the assessment's findings. The objectives deliberately mirror the structure of the FAIR principles: each objective operationalizes one principle for this specific dataset and reuse case, replacing the principle's generic formulation with a verifiable target and the concrete standards and registries agreed with the stakeholders (e.g., W3ID for identifiers, eLwazi for registration, DCAT for the metadata model). Table 3 makes this translation traceable from reuse objective, via the assessment finding, to FAIRification objective. For Findability, the reuse objective to make the dataset "easily found" led to the FAIRification objective of assigning persistent identifiers and registering the metadata in a

community-relevant catalogue. To meet the Accessibility goal of having "clearly described" access conditions, the objectives focused on specifying a standardized access protocol and publishing metadata in a separate, persistent repository. For Interoperability, the objective was to make the data available in additional formats that are relevant for our reuse stakeholders. Finally, to enhance Reusability, the objectives centred on specifying the data license and reuse conditions explicitly within the metadata, and on annotating the data with FAIR ontologies and vocabularies used by the research community.

Table 3. Reuse objectives and related FAIRification goals, organized per FAIR aspect. A summary of the FAIR assessment is presented in this table to facilitate readability. The FAIRification objectives have IDs (FO#) assigned as they are further referenced in this paper.

FAIR aspect	***Reuse objective***	***FAIR assessment (summary)***	***FAIRification objective***
Find.	The MomCare dataset is easily found by those conducting meta-analysis, especially those from MNCH research groups. Ideally, the metadata would be found on the eLwazi Open Data Science Platform.	Globally, unique and persistent identifiers are not used. The metadata does not contain sufficient information to enable findability.	*FO1:* (Meta) data are assigned globally unique and persistent identifiers *FO2:* Metadata are registered on a research data catalogue that is relevant for the research community
Acces.	Conditions and governance mechanisms for data access are clearly described so that data can be reused by other research groups, such as the HE2AT group.	Information on how to request access to the data is insufficient. The longevity of metadata is not guaranteed.	*FO3:* Standardized Access Protocol is specified in the metadata *FO4:* Metadata is published in a separate repository that can guarantee its longevity
Interop.	The MomCare dataset is made available in formats used in MNCH research groups, including the HE2AT group.	The (meta)data does not use FAIR vocabularies, and it does not reference other FAIR (meta)data.	*FO5:* The (meta)data is annotated with FAIR ontologies and vocabularies that facilitate merging with additional (meta)data. *FO6:* The (meta)data is made available in different formats that are relevant for the research community.
Reus.	The MomCare (meta)data concepts are aligned with ontologies and vocabularies used by the research community, including the HE2AT research group. There is clear information on the data's license and use conditions.	The metadata does not include license or provenance information, nor is it standardized using community-relevant standards.	*FO7:* (Meta)Data use schemas that are relevant for the research community. *FO8:* Data reuse conditions are specified in the metadata *FO9:* Data license is specified in the metadata.

Phase 2: Iterative FAIRification Implementation (FHIR-to-FAIR)

FAIR actions backlog and prioritization. From the FAIRification objectives, FAIRification actions were specified. As visualized in Table 4, these actions outline the activities necessary for optimal adherence to the FAIR principles. These activities include ensuring the persistence of identifiers, registering the metadata in a research data catalogue, aligning the metadata model with research ontologies, and enhancing sharing and analytics properties. Additionally, activities that address gaps within the FHIR-based environment must be performed. These activities include enforcing the use of references to community-compliant metadata, ontologies, and access and reuse conditions.

Table 4. FAIRification backlog and RDA-model assigned priority.

Related FAIRification objective	*FAIR action ID*	*FAIR actions backlog*	*FAIR action priority*
FO1	*FA1*	Link (meta)data identifiers to persistent identifiers using W3ID.	Essential
FO2	*FA2*	Register metadata in the eLwazi catalogue.	Essential
FO2, FO4	*FA3*	Develop and publish machine-understandable metadata.	Essential
FO2	*FA4*	Define persistent link between metadata and data.	Essential
FO5	*FA5*	Annotate metadata with DCAT terms and use additional ontologies to enrich the metadata model.	Essential
FO6, FO7	*FA6*	Annotate the metadata with community relevant ontologies and make it available in RDF.	Important
FO6, FO7	*FA7*	Annotate the data with ontologies that are also used in HE2AT and make it available in other formats including RDF.	Important
FO8	*FA8*	Make the data use conditions explicit in the machine-understandable metadata model	Essential
FO9	*FA9*	Make the data license explicit in the machine-understandable metadata model	Essential

FAIR implementation

The implementation of the action backlog was a collaborative effort, bringing together experts in FAIR technologies and semantic web, software development, and the MNCH domain. In brief, each (meta)data element was assigned a persistent W3ID identifier and the MomCare metadata was registered in the eLwazi catalogue to improve persistence and findability; the resulting machine-actionable metadata record (metadata.ttl) is available in the project repository.[32] To enhance interoperability and reusability, we developed a standardized, machine-understandable metadata model that extends the DCAT standard with descriptors relevant to data reusers

ARTICLE IN PRESS

(illustrated in the metadata-model diagram in the project repository), published it as RDF, and exposed the MomCare data into an ontology-annotated RDF version using the Ontop virtual knowledge-graph platform. Access and reuse conditions were made explicit in the metadata using the Data Use Conditions ontology (DUO) and a CC BY-SA 4.0 license. The full action-by-action implementation, with worked examples, is provided in the project's public repository, together with the complete documentation, models and code.[33]

Phase 3: Post-FAIRification Review

Re-assess FAIRness. A structured re-assessment verified measurable improvements across each FAIR dimension. Findability gains included a persistent W3ID identifier (F1) and rich, use-case-relevant metadata via a standardized DCAT model linking to related resources (F2, F3); registration in the eLwazi catalogue and an RDF version on Zenodo and GitHub ensured discoverability and machine-actionability (F4).

With respect to Accessibility, access conditions are now explicitly defined via DUO references for both human and machine agents (A1), and publication across three repositories (eLwazi, Zenodo, and GitHub) ensures metadata persistence independent of the primary data source (A2).

For Interoperability and Reusability, the re-assessment confirmed adherence to community standards through the integration of SNOMED CT, LOINC, and ICD-10 (I1, I2). The cross-vocabulary alignment itself is implemented in the semantic mapping layer, which maps the source data elements to the IRIs of these vocabularies; the RDF serialization exposed through these mappings is the output through which the alignment becomes available to data reusers. Reusability was verified through the explicit inclusion of a machine-readable license (CC BY-SA 4.0) and comprehensive documentation supporting future use (R1, R2).

Review with stakeholders. As described in the methods, the review with stakeholders focused on assessing whether they could perceive the implemented FAIR aspects and their added benefits. However, stakeholders demonstrated considerable variation in their assessments of the MomCare data's adherence to the FAIR principles. It was observed that the ease with which participants could identify FAIR attributes depended heavily on their technical nature. While high-level, user-facing enhancements were consistently recognized, more abstract or deeply embedded technical features proved difficult for many to locate or interpret, highlighting a gap in communication rather than implementation. Table 5 summarizes these findings, categorizing the main themes from the stakeholder review into three groups based on their perceived visibility: attributes that were overall positively identified, those with high variation in perception, and those that were complex for most to identify.

Table 5. Stakeholder perceptions with regards to identifying FAIR aspects within the (meta)data.

Stakeholder perception	FAIR aspect

ARTICLE IN PRESS

Overall positive	Metadata accessibility (A1) Reuse information (R1) in metadata Metadata interoperability (I1, I3) Findability of metadata (F2)
High variation in ease of identification	Findability of data (F1) Interoperability of data (I1) Referencing of metadata (I3) Rich reuse attributes (R1)
Complex for most	Data (automated) access protocol and provenance information (A1) Data knowledge presentation, FAIR-compliance vocabularies and references (I1-3) Uncertainties about data compliance with community standards or vocabularies (R2)

Evaluate next steps. Based on our application, the evaluation of next steps was highly informed by the preceding assessment phases. In our application, this evaluation revealed a crucial insight: While the technical re-assessment confirmed that the dataset was demonstrably more FAIR, the stakeholder feedback showed that the perceived value of these improvements was not uniform, underscoring a gap between technical implementation and user perception.

This feedback drove a refined set of objectives: future cycles must prioritize the communication and usability of FAIR features alongside their technical implementation. Key next steps include user-focused documentation—clear guides and diagrams for concepts like the data access protocol and provenance information that non-technical users struggled to identify—and practical examples demonstrating how to leverage advanced FAIR attributes such as persistent identifiers and querying with the integrated community vocabularies. Incorporating these communication-focused tasks into the next iteration's backlog ensures the practical benefits of FAIRification become as tangible to stakeholders as the underlying technical compliance.

Discussion

As seen in the MomCare dataset, the majority of routinely collected health data remains constrained by its primary purpose (claims management and operational tracking) rather than structured for secondary research use, reflecting a broader gap in health informatics between operational data collection and research usability.[34,35]

The workflow presented in this paper shows how incremental but essential modifications can enhance the reusability of routinely captured healthcare data for research. FHIR is designed primarily for data exchange within healthcare systems and, while it supports many practices that align with FAIR, its role remains exchange-focused rather than research-centric.[11] FAIRwithFHIR leverages FHIR-based data structures to bridge this gap, moving from interoperability toward broader reusability. In doing so, FAIRwithFHIR builds upon existing work directed at metadata and

ARTICLE IN PRESS

data transformations between the two communities of practice involved: the healthcare interoperability community organized around HL7 FHIR, and the research data community organized around the FAIR principles.[23,25] It adds to this by not only identifying limitations within the FHIR ecosystem but utilizing existing elements within the FHIR standard to specify FAIR data and metadata requirements and proposing actionable extensions for increased FAIR maturity. Our findings indicate that these FAIRification actions (including improved metadata discoverability, expanded ontological mappings, and optimized data access methods) are enablers to transform FHIR-compliant datasets into machine-actionable, research-ready resources.

The application in the HE2AT study supports the value of FAIRwithFHIR in enabling cross-community data reuse, highlighting how the iterative workflow ensures that the FHIR-based dataset is not only interoperable within health systems but also structured for broader, FAIR-compliant research applications. While the FAIRification process was demonstrated using the MomCare dataset, several steps in the workflow, such as the publication of a dataset-level DCAT metadata record, the use of virtual RDF mappings through Ontop, and semantic alignment using existing FHIR value sets, are transferable to other FHIR-based implementations. While the transferability remains to be tested, it is expected that these approaches can be adapted in other health domains that share the defining characteristics of the MomCare case (longitudinal, routinely collected operational data that is increasingly captured in FHIR under national interoperability mandates, yet underutilized for research) such as chronic disease surveillance, immunization tracking, or primary care analytics.[11,13,16]

Strengths and limitations

The strengths of FAIRwithFHIR lie in its modularity and pragmatic implementation. By building upon the established FHIR data standard, its retrospective, stepwise approach facilitates FAIR implementation across operational data use contexts, extending FAIRification work beyond exclusively scientific workflows. Unlike existing generic FAIRification workflows, FAIRwithFHIR explicitly operationalizes FAIR principles within the specific technical and semantic structures of the FHIR standard. The workflow integrates FHIR-native mechanisms, such as resource profiling, extensions, and implementation guides, with practices that implement specific FAIR principles: capture of machine-actionable metadata (F2, R1), exposure of the data in the FHIR-defined RDF serialization (I1, I3), and alignment with FAIR-compliant vocabularies and ontologies (I2). In the MomCare application, each of these is traceable to a concrete implementation step: the DCAT-based metadata model with its registration in the eLwazi catalogue (FA2–FA5, Table 4), the virtual RDF exposure of the FHIR data through Ontop (FA6–FA7), and the vocabulary annotations building on the SNOMED CT, LOINC and ICD-10 bindings of the IPS+ profile. Because the RDF exposure is generated virtually from the operational FHIR data rather than through a converted copy, FAIRification occurs within the existing FHIR implementation rather than as a separate post-processing step, offering a practical blueprint for embedding FAIR practices directly into operational health data systems.

ARTICLE IN PRESS

The workflow also has clear limitations. Its success depends heavily on the availability and completeness of high-quality metadata, a persistent weakness in many real-world FHIR implementations.[36,37] Additionally, while FAIRwithFHIR strengthens FAIR maturity, its impact on user adoption remains to be further explored. Our stakeholder review reveals a challenge that has also been reported for FAIRification efforts in other domains; a disconnect between successful technical implementation and the practical perception of FAIRness by data reusers (see Results, Table 5). While data reusers readily identified visible improvements like accessible metadata and clear licensing, they struggled to appreciate more abstract attributes. This finding suggests that technical compliance alone is insufficient, in line with other studies reporting that the adoption of FAIR data also depends strongly on organizational factors such as stakeholder communication, as well as available resources and capacity in the clinical informatics domain.[38,39] To achieve true reusability, FAIRification workflows must extend beyond technical tasks to include a communication layer, with user-focused documentation that makes the benefits of 'under-the-hood' features tangible.

Stakeholder involvement

The stakeholder involvement that FAIRwithFHIR relies on proved both feasible and demanding in our application. Identifying stakeholders was comparatively easy because the reuse case came with an organized community: the HE2AT project and the eLwazi platform within the DS-I Africa consortium provided data reusers with an interest in the dataset. Sustaining involvement across the workflow phases was more difficult: engagement was largely confined to structured scoring of the FAIR indicators and depended on stakeholders investing time in material that was, for some, outside their role or area of expertise. Within the engagement we observed that questions shifted from user-facing aspects, such as documentation and access conditions, toward vocabulary bindings and provenance. As stakeholder involvement remained constrained because of time and capacity limits, we could not measure whether repeated exposure of this kind builds durable FAIR capacity within a reuse community. Future research should evaluate this learning effect using formal assessment tools that measure how stakeholder participation affects FAIR literacy and technical engagement.

Directions for future research

FHIR's inherent focus on transactional exchange highlights additional considerations needed when FAIRifying dynamic datasets, particularly those related to data provenance, versioning, and persistent identifiers. While the FAIR principles can be applied to both static datasets (datasets that are fixed at the time of publication, such as the export of a completed trial) and dynamic datasets (datasets that continue to change while being reused, such as the transaction-oriented records on an operational FHIR server, to which resources are added and updated as care delivery continues), most FAIRification processes to date have been framed around static datasets.[20,21,23] For a dynamic dataset, descriptive metadata that was accurate at the moment of FAIRification (such as record counts, coverage period, and provenance information) becomes outdated as the data change. In the MomCare application, the dataset-level metadata record was curated manually, so keeping it current depends on the cyclical review phase, which periodically re-verifies whether the metadata still describes the dataset correctly. Future applications of

ARTICLE IN PRESS

FAIRwithFHIR that involve dynamic datasets should therefore define programmatic metadata generation as an explicit FAIRification objective: because FHIR data are already structured, pipelines can derive descriptive metadata directly from the FHIR source and regenerate it whenever the data change. Without such automation, metadata maintenance requires ongoing manual curation effort, which is resource-intensive and may not scale well in low-resource settings.

Finally, several low-effort but high-impact modifications could strengthen FAIR alignment within the HL7/FHIR ecosystem itself. Our results show, however, that FHIR mechanisms which remain optional tend to go unused; proposals at the ecosystem level should take this into account. Defining FAIR metadata profiles within FHIR Implementation Guides (e.g., IPS or the Research Study IG), with the relevant metadata elements marked as required or must-support rather than optional, would give implementers a common, testable target for the publication of machine-actionable metadata. Promoting the FHIR-defined RDF representation as a first-class output, rather than an optional serialization, would enhance interoperability with linked data infrastructures. Moreover, community-driven mappings between FHIR resources and common FAIR ontologies (e.g., DCAT ontologies) could accelerate cross-domain discoverability and reuse. Because FHIR prescribes a standardized information model, such mappings need to be defined only once: community-maintained versions of the virtual mappings used for MomCare would allow any FHIR-compliant implementation to expose its data as ontology-annotated RDF that data reusers can consume without FHIR-specific tooling. Such measures would help ensure that future FHIR datasets are ‘born’ FAIR, reducing the need for post-hoc FAIRification and supporting a truly interoperable research data ecosystem.

ARTICLE IN PRESS

Conclusion

This paper introduced and evaluated FAIRwithFHIR, an agile, stakeholder-driven workflow to address the gap between FHIR-compliant operational health data and the needs of the research community. By applying this iterative process to the MomCare dataset, we demonstrated a structured methodology for transforming a real-world clinical resource into a more FAIR asset, guided by the specific reuse case of the HE2AT study. Our findings confirm that while leveraging on a FHIR-based dataset provides a strong foundation for interoperability, significant enhancements are required to achieve true reusability. Critically, our stakeholder review revealed that technical FAIR compliance alone is insufficient, as the value of foundational technical features is often not self-evident to data reusers. Therefore, we conclude that successful FAIRification requires an approach that not only implements technical solutions but also integrates a communication layer, ensuring that the resulting data is not just technically FAIR, but also practically understandable and usable by its target research community.

To support future FHIR-to-FAIR initiatives, we translate the findings of the MomCare application into four lessons for improving the FAIRness of FHIR-compliant implementation data:

ARTICLE IN PRESS

1. Anchor FAIRification choices in a concrete reuse case and its community.
    - The implementation choices that shaped our FAIRification all followed directly from the reuse case defined in Phase 1 (Table 3). Defining this reuse case together with the target community turns the generic FAIR principles into verifiable, dataset-specific objectives and prevents FAIRification from optimizing indicators without serving actual data reusers.
2. Treat optional FHIR capabilities as pathways to FAIRification.
    - Most gaps identified in our assessment traced back to FHIR mechanisms that exist in the standard but are optional and remained unused in the MomCare implementation: terminology bindings, the Provenance resource, explicit references, and the RDF serialization (Tables 1 and 2). Retrospective FAIRification of FHIR-compliant data is therefore largely a matter of activating capabilities already defined in the ecosystem, rather than of introducing new infrastructure.
3. Publish dataset-level metadata independently of the FHIR server.
    - An operational FHIR server can guarantee neither the persistence of metadata beyond the lifetime of its resources (A2) nor discoverability by the research community (F4). The largest observed gains in FAIR maturity came from the separate, machine-actionable DCAT metadata record, published in the eLwazi catalogue and archived on Zenodo, which persists and can be indexed independently of the primary data source.
4. Pair every technical improvement with a communication layer.
    - Our stakeholder review showed that technically successful FAIRification can remain partly invisible to data reusers (Table 5). User-focused documentation and worked examples that make ‘under-the-hood’ improvements tangible are needed to convert FAIR maturity into actual reuse, and should be planned as FAIRification actions in their own right.

By applying FAIRification to routine FHIR-based health data in successive cycles of assessment, implementation and review, we demonstrate a structured approach to overcoming existing limitations in MNCH data accessibility. Future work should focus on machine-actionable, resource-efficient FAIRification workflows, integrating best practices from both the healthcare and research communities to facilitate improvements in maternal and child health outcomes across the global health system.

ARTICLE IN PRESS

Acknowledgements

The authors thank the DS-I Africa community for their feedback, especially Lyndon Zass for his work regarding the metadata model review and the engagement of the data user community. Additionally, the authors thank John Grimes for providing the example code for creating the patient timeline table with the SQL-on-FHIR specification and Pathling as a runner. The authors also extend their extensive gratitude and appreciation to Philip van Damme and Annika Jacobsen for their expert advise and review in shaping the FAIR with FHIR workflow.

Funding

PharmAccess is supported by a structural grant from the Netherlands Ministry of Foreign Affairs.

Contributions

FH coordinated and led the writing, conceptualization of the framework, use-case application and analysis. CB managed the conceptualization and extensively contributed to the application to the use-case as well as review & editing. CI contributed to the technical aspects of the FAIR implementation methods. DK contributed to technical conceptualization, application and review & editing. MR contributed to the review & editing.

Data and Code Availability

The data and analytical materials supporting the findings of this study are openly available at the PharmAccess Foundation repository, “research-momcare” (https://doi.org/10.5281/zenodo.18663937). The repository contains all non-identifiable code, scripts, and documentation required to reproduce the analyses presented in the manuscript.

Please cite the repository as:

Heddema, F., Bernabé, C., Chris, I., Kapitan, D., & Roos, M. (2026). Data and code for ‘FAIR with FHIR: A Proposed Workflow for Retrospective FAIRification of FHIR-Compliant Health Data’ [Data set]. Zenodo. https://doi.org/10.5281/zenodo.18663937

Ethical Clearance

Ethical approval for the use of the data in this study was obtained from the Amref Health Africa Ethics and Scientific Review Committee (ESRC) under reference number P1649-2024. The ESRC is accredited by the Kenyan National Commission for Science, Technology and Innovation (NACOSTI), registration number NCST/NBC/AC/0912, and is registered with the US Office for Human Research Protections (OHRP) under IRB organization number IORG0007315 and Federalwide Assurance FWA00018713. All data used in this study were de-identified prior to analysis.

ARTICLE IN PRESS

References

1. UNICEF. Levels and trends in child mortality. UNICEF Data https://data.unicef.org/resources/levels-and-trends-in-child-mortality/ (2021).

2. Alhumaidi, N. H., Dermawan, D., Kamaruzaman, H. F. & Alotaiq, N. The use of machine learning for analyzing real-world data in disease prediction and management: systematic review. JMIR Med. Inform. 13, e68898 (2025). https://doi.org/10.2196/68898

3. Al-khayat, Z., Franzen, N., Retèl, V. P. & Van Harten, W. H. Hospital-based real-world evidence in health technology assessment: insights from a scoping review of European, Australian and North American guidance and expert interviews. BMJ Open 15, e100304 (2025). https://doi.org/10.1136/bmjopen-2025-100304

4. Patel, S. Y., Akileswaran, C. & Basu, S. Early detection of high risk pregnancies using clinical and social data to improve health outcomes. npj Digit. Public Health 1, 3 (2026). https://doi.org/10.1038/s44482-025-00003-5

5. Hu, Y., Slavin, V., Enticott, J. & Callander, E. Beyond evidence hierarchies: leveraging randomized controlled trials and real-world data to advance the value of maternity care. Acta Obstet. Gynecol. Scand. 105, 395–402 (2026). https://doi.org/10.1111/aogs.70094

6. Liu, F. & Panagiotakos, D. Real-world data: a brief review of the methods, applications, challenges and opportunities. BMC Med. Res. Methodol. 22, 287 (2022). https://doi.org/10.1186/s12874-022-01768-6

7. Blonde, L., Khunti, K., Harris, S. B., Meizinger, C. & Skolnik, N. S. Interpretation and impact of real-world clinical data for the practicing clinician. Adv. Ther. 35, 1763–1774 (2018). https://doi.org/10.1007/s12325-018-0805-y

8. Nordo, A. H. et al. Use of EHRs data for clinical research: historical progress and current applications. Learn. Health Syst. 3, e10076 (2019). https://doi.org/10.1002/lrh2.10076

9. Shah, A. Using data for improvement. BMJ 364, l189 (2019). https://doi.org/10.1136/bmj.l189

10. Gehrmann, J., Herczog, E., Decker, S. & Beyan, O. What prevents us from reusing medical real-world data in research. Sci. Data 10, 459 (2023). https://doi.org/10.1038/s41597-023-02361-2

11. Kapitan, D. et al. Data interoperability in context: the importance of open-source implementations when choosing open standards. J. Med. Internet Res. 27, e66616 (2025). https://doi.org/10.2196/66616

12. Molenaar, J., Beňová, L., Christou, A., Lange, I. L. & van Olmen, J. Travelling numbers and broken loops: a qualitative systematic review on collecting and reporting maternal and

ARTICLE IN PRESS

neonatal health data in low- and lower-middle income countries. SSM Popul. Health 26, 101668 (2024). https://doi.org/10.1016/j.ssmph.2024.101668

13. Waiyaiya, E. et al. Digital health convergence workshop: the Kenyan experience. Oxf. Open Digit. Health 2, oqae014 (2024). https://doi.org/10.1093/oodh/oqae014

14. HL7 International. Overview — FHIR v5.0.0. https://hl7.org/fhir/R5/overview.html (2025).

15. Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S. & Ramoni, R. B. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J. Am. Med. Inform. Assoc. 23, 899–908 (2016). https://doi.org/10.1093/jamia/ocv189

16. Duda, S. N. et al. HL7 FHIR-based tools and initiatives to support clinical research: a scoping review. J. Am. Med. Inform. Assoc. 29, 1642–1653 (2022). https://doi.org/10.1093/jamia/ocac105

17. HL7. Clinical Study Protocol v1.0.0-ballot2. http://hl7.org/fhir/uv/clinical-study-protocol/2026May (2026)

18. Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

19. Jacobsen, A. et al. FAIR principles: interpretations and implementation considerations. Data Intell. 2, 10–29 (2020). https://doi.org/10.1162/dint_r_00024

20. Jacobsen, A. et al. A generic workflow for the data FAIRification process. Data Intell. 2, 56–65 (2020). https://doi.org/10.1162/dint_a_00028

21. Welter, D. et al. FAIR in action — a flexible framework to guide FAIRification. Sci. Data 10, 291 (2023). https://doi.org/10.1038/s41597-023-02167-2

22. Martínez-García, A., Alvarez-Romero, C., Román-Villarán, E., Bernabeu-Wittel, M. & Luis Parra-Calderón, C. FAIR principles to improve the impact on health research management outcomes. Heliyon 9, e15733 (2023). https://doi.org/10.1016/j.heliyon.2023.e15733

23. Sinaci, A. A. et al. From raw data to FAIR data: the FAIRification workflow for health research. Methods Inf. Med. 59, e21–e32 (2020). https://doi.org/10.1055/s-0040-1713684

24. Dohmen, P. et al. Implementing value-based healthcare using a digital health exchange platform to improve pregnancy and childbirth outcomes in urban and rural Kenya. Front. Public Health 10, 1040094 (2022). https://doi.org/10.3389/fpubh.2022.1040094

25. van Damme, P., Löbe, M., Benis, N., de Keizer, N. F. & Cornet, R. Assessing the use of HL7 FHIR for implementing the FAIR guiding principles: a case study of the MIMIC-IV

ARTICLE IN PRESS

Emergency Department module. JAMIA Open 7, ooae002 (2024). https://doi.org/10.1093/jamiaopen/ooae002

26. FAIR Data Maturity Model Working Group. FAIR data maturity model: specification and guidelines. Research Data Alliance https://doi.org/10.15497/rda00050 (2020).

27. HL7. HTTP — FHIR v6.0.0-ballot3. https://hl7.org/fhir/6.0.0-ballot3/http.html (2026).

28. Bernabé, C. et al. GO-Plan: a goal-oriented method for FAIRification planning. Inf. Serv. Use 45, 100–124 (2025). https://doi.org/10.1177/18758789251324008

29. NIH REPORTER. https://reporter.nih.gov/project-details/10314149 (2025).

30. NIH Common Fund. Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa). https://commonfund.nih.gov/AfricaData (2025).

31. eLwazi. eLwazi data catalogue. https://catalog.elwazi.org/#/datasets (2025).

32. Heddema, F., Bernabé, C., Ihure, C. Kapitan, D. & Roos, M. Data and code for 'FAIR with FHIR: a proposed workflow for retrospective FAIRification of FHIR-compliant health data'. Zenodo https://doi.org/10.5281/zenodo.21890900 (2026).

33. PharmAccess Foundation. PharmAccess/research-momcare. GitHub https://github.com/PharmAccess/research-momcare (2025).

34. Shah, S. M. & Khan, R. A. Secondary use of electronic health record: opportunities and challenges. IEEE Access 8, 136947–136165 (2020). https://doi.org/10.1109/ACCESS.2020.3011099

35. Emerson, S. D. et al. Secondary use of routinely collected administrative health data for epidemiologic research: answering research questions using data collected for a different purpose. Int. J. Popul. Data Sci. 9, 2407 (2024). https://doi.org/10.23889/ijpds.v9i1.2407

36. Marfoglia, A. et al. Towards real-world clinical data standardization: a modular FHIR-driven transformation pipeline to enhance semantic interoperability in healthcare. Comput. Biol. Med. 187, 109745 (2025). https://doi.org/10.1016/j.compbiomed.2025.109745

37. van Essen, M. H. J. et al. Implications of data extraction and processing of electronic health records for epidemiological research: observational study. J. Med. Internet Res. 27, e64628 (2025). https://doi.org/10.2196/64628

38. Wise, J. et al. Implementation and relevance of FAIR data principles in biopharmaceutical R&D. Drug Discov. Today 24, 933–938 (2019). https://doi.org/10.1016/j.drudis.2019.01.008

ARTICLE IN PRESS

39. Awadallah, R. & Tammaro, A. M. The road to FAIR energy data: motivations, shortcomings, and use cases. Int. Inf. Libr. Rev. 57, 177–192 (2025). https://doi.org/10.1080/10572317.2025.2479280

ARTICLE IN PRESS