Towards a health data commons in LMICs

Demonstrating fair sharing and reuse of health data in sub-Saharan Africa

Daniel Kapitan, Julie Fleischer, Chris Ihure, Rob Wiegman, Iris Bokkes, Mark van der G

Contents

List of Figures

List of Tables

Data commons as a catalyst for achieving UHC

The world lags significantly in its pursuit to reach the Sustainable Development Goal (SDG) of universal health coverage (UHC).[1] It is a widely held belief that digital technologies have an important role to play towards achieving this goal in low- and middle income countries (LMICs).[2] Yet, no clear cut development paths are known as to how digital health can be used effectively and efficiently on a large scale in such low resource settings.[3]

We believe that digital technologies indeed have an important role to play in achieving UHC. More specifically, we posit that patient-centric data access and reuse is an essential element in systematically improving care delivery in LMICs. While data access and reuse has been put center stage in many studies, it is a complex and dynamic field of work. Academic research often fails to bridge the gap towards practice. Realizing a commons-based ecosystem of for data sharing is difficult, as it needs to overcome potentially conflicting interests of actors involved in the system. Ongoing rapid developments of digital technologies themselves make it difficult to access investment decisions in an already resource constrained context.

Despite the challenges, we want to demonstrate that patient-centric data access and reuse is feasible, today. We take the approach of *"show, don't tell"*. Through implementing demonstrator projects that contribute towards the creation of a **Health Data Commons (HDC)**, we show that health data sharing can be achieved in LMICs at acceptable cost and low technical risk. This document describes the learnings from the HDC project. Through these demonstrators we aim to initiate a paradigm shift as to how data sharing can be realized such that it can act as a catalyst towards achieving UHC.

Building on the openHIE framework

The HDC project takes the openHIE framework[4] as a starting point, being the most generic and commonly used health information interoperability framework. This framework has by and large been adopted by sub-Saharan African countries[5], including Nigeria[6], Kenya[7] and Tanzania.[8]

The HDC framework and its components are shown in figure 1. The bottom layer (yellow) comprises the point of service (PoS) systems, which includes the systems used by clinicians, health professionals,

[1] World Health Organization and World Bank (2021)
[2] Kickbusch et al. (2021) and World Health Organization (2021)
[3] McCool et al. (2022) and Neumark and Prince (2021)
[4] ohie.org
[5] Mamuye et al. (2022)
[6] Nigeria data exchange architecture for the national data repository
[7] KHISIF
[8] "Tanzania Health Enterprise Architecture" (2020)

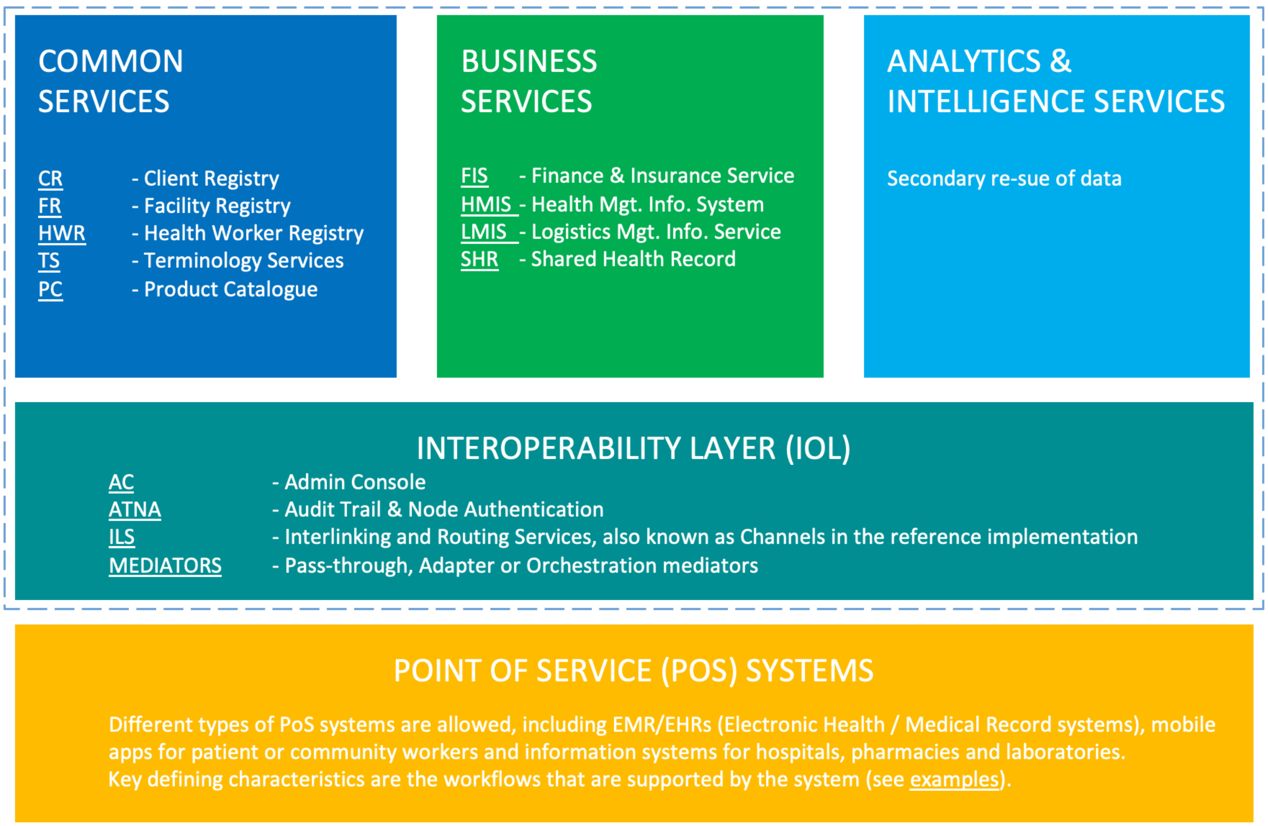


Figure 1.: OpenHIE framework

community health workers and the like. Examples of such systems are OpenMRS electronic medical records (EMR) system and the RapidSMS mHealth application, which are used to access and update a patient's records, register activities and record healthcare transactions.

The second, middle layer (teal) constitutes the interoperability layer (IOL), which acts as a gateway for exchanging information between systems. Any type of information exchange, be it between two PoS systems, or between a PoS system and business services (explained below), is mediated through this interoperability layer. The interoperability layer provides functionality such as routing, translation services, auditing and authentication.

The top layer of framework comprises three distinct domains. Common Services (blue) include a variety of registry services that are designed to uniquely identify and track unique patients, facilities, healthcare products, and terminology that are used throughout the health data commons. Business Service (green) are designed to support the delivery of care within the health system. The District Health Information System version 2, for example, is a well-known and widely used Health Management Information System for collecting, analyzing, visualizing and sharing data through combining data from multiple PoS systems for a given geography or jurisdiction.[9]

The HDC framework explicitly adds a third domain in the top layer, which is not included in the OpenHIE specification, namely Analytics & Intelligence Services (light blue). The rationale for this addition is to facilitate secondary use of health data for academic research, real-world evidence studies etc. within the nascent concept of health data spaces.[10] Note that the Kenyan Health Information Systems Interoperability Framework (KHISIF) has also explicitly included the analytics domain.

Shifting the paradigm for creating a health data commons

Over the last few years, the importance of interoperability of systems and reuse of data has become evident. One of the key challenges in establishing interoperability is the dilemma associated with economies of scope: start small, and run the risk of not achieving common standards. Start large, and get bogged down in talking rather than building a new standard infrastructure. Many actors in the healthcare domain believe that interoperability can only be achieved through top-down, centrally led design and implementation. At the same time, it is notoriously difficult to run such large scale programs effectively and efficiently. We believe this paradox is actually surmountable. Modern digital technologies and solution patterns for achieving interoperability are extensible and modular by design, and can be successfully implemented in a decentralized, federated fashion.

The hourglass model

To illustrate this point, consider how the internet came to be since its inception. Beck (2019) argued that the success and ubiquitousness of the internet stems from its foundation on minimal, extensible standards. This so-called 'hourglass model' of the internet is shown in figure 2.

[9] DHIS2

[10] see for example the International Data Spaces Association

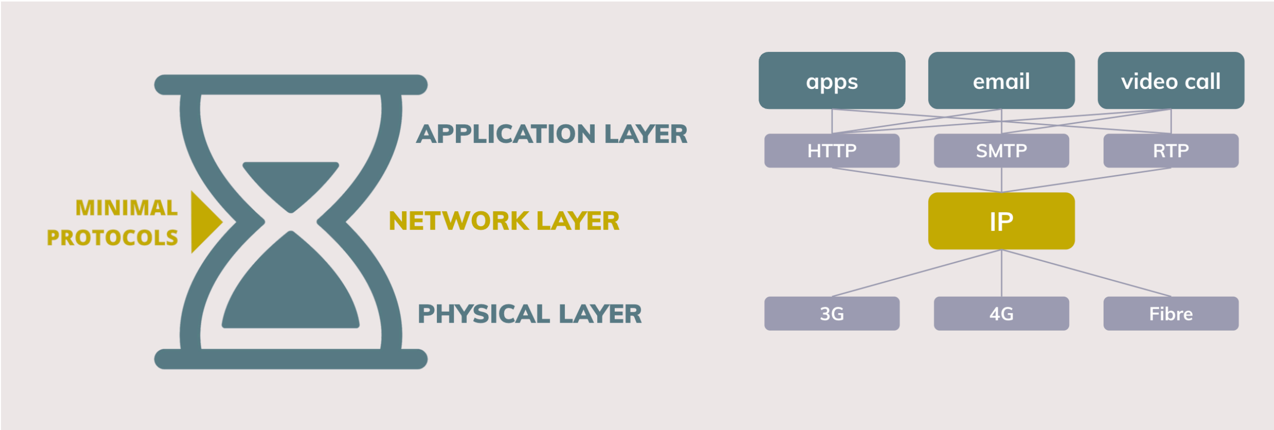


Figure 2.: The hourglass model of the internet

In essence, the hourglass model states that defining minimal standards (protocols) at the waist of the hourglass is crucial towards supporting a wide range of applications and supporting services at the top and bottom of the hourglass. For the internet, the IP protocol constitutes the center of the hourglass. Although the IP protocol is very limited in itself (it can only send and receive data packets), it is exactly because of its minimal standard that the IP protocol is able to support a great diversity of applications on top of it (web apps, email, videoconferencing) and allows implementation using a great diversity of supporting services below it (in this case, the many different types of physical networks). However, despite this flexibility and extensibility, the hourglass model strictly states that the components that lie above the central layer cannot directly access the services that lie below it.

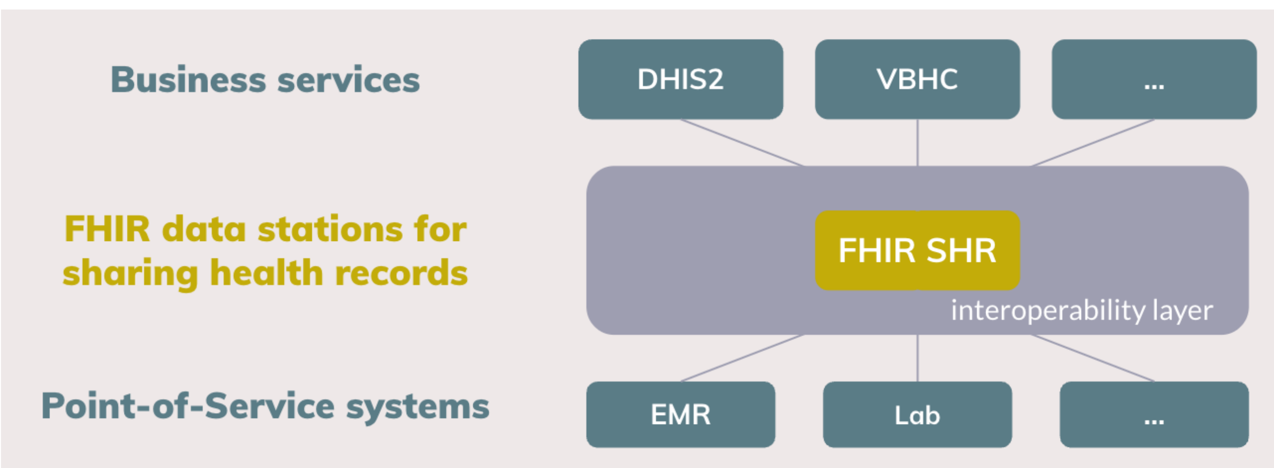


Figure 3.: The hourglass model for health data

The hourglass model can be used as a foundational concept for a new paradigm on interoperability for health data, as is shown in figure 3. Essentially, we propose an integration of the design principles of OpenHIE, Fast Healthcare Interoperability Resources (FHIR) and to make data Findable, Accessible, Interoperable and Reusable (FAIR).

FHIR as the *de facto* standard for health data exchange

FHIR has become the *de facto* standard for clinical information exchange in the healthcare sector, both for routine health data exchange[11] and research settings[12]. The recently announced collaboration between HL7 International (the governing body of FHIR) and WHO exemplifies this trend.[13] To date, however, the lion's share of FHIR-related projects in LMICs focus on achieving openness and interoperability at the level of Point-of-Service systems. For example, the SMART on FHIR standard provides a consistent approach to security and data requirements for health applications and defines a workflow that an application can use to securely request access to data, and then receive and use that data.[14]

Beyond the value of FHIR for Point-of-Service systems, we propose to adopt FHIR as the elementary standard for realizing the Shared Health Record (SHR) component that is specified in the OpenHIE framework. We envisage FHIR SHR data stations at the center of the hourglass as the minimal standard for health data sharing. Higher lying business services can access health data as is persisted in the FHIR SHR, while lower lying Point-of-Service systems can interact with the FHIR SHR to read, write and combine the shared health record. As the FHIR standard is extensible in itself, that is, data sets can be composed out of the 100 elementary building blocks[15], the implementation and roll-out of data stations can happen in a gradual manner. In the work presented here, we take the International Patient Summary (IPS) that consists of about 15 FHIR building blocks as the basis for the shared health record.[16] As the use and reuse of health data grows, more data can be added to the FHIR SHR.

Alignment with FAIR principles

Our paradigm also aligns with the principles put forward by the GO FAIR community.[17] For example, these principles have been applied in implementing the VODAN-Africa data infrastructure for monitoring COVID-19.[18] Building on the experience from the VODAN Africa project, Gebreslassie et al. (2023) have analyzed whether the FHIR standard can indeed be leveraged for the FAIRification process. They conclude that FHIR as a native solution is a viable option through the protocols and specifications it supports, or with the community implementation guides.[19]

[11] Ayaz et al. (2021).
[12] Duda et al. (2022), Vorisek et al. (2022).
[13] https://www.who.int/news/item/03-07-2023-who-and-hl7-collaborate-to-support-adoption-of-open-interoperability-standards
[14] Mandel et al. (2016).
[15] the so-called FHIR Resources
[16] FHIR IPS Implementation Guide
[17] https://go-fair.org
[18] van Reisen et al. (2021)
[19] For more details, see Schultes (2023).

Implement HDC with modern technology

The OpenHIE community was formed a decade ago in 2013, and whilst the conceptual elements of the framework are still relevant and useful, a lot has changed in terms of available software tools and implementations. One of the key contributions of the HDC project is to implement demonstrators that comply to the openHIE guidelines using modern digital technologies. More specifically, we aim to leverage the following open source components:

- **OpenHIM as core interoperability platform**[20]: thanks to Jembi Health Systems and the open source community, the OpenHIM has matured over the last decade. The current version 8 is being used in various countries.[21]
- **HAPI FHIR and Bulk FHIR API**: HAPI FHIR is a complete implementation of the HL7 FHIR standard for healthcare interoperability, and is one the most commonly used open source implementation in LMICs.[22] We demonstrate that this component can be used effectively to implement a Shared Patient Record that supports both primary use (exchange of single records between users) and secondary use (reuse of data of a whole population for analytical purposes). Furthermore, we particularly focus on using the bulk FHIR API as a means for supporting analytical workflows that require access to and processing of data in bulk.[23]
- **Federated analytics and federated learning**[24]: existing medical data is not fully exploited for secondary reuse primarily because it sits in data silos and privacy concerns restrict access to this data. However, without access to sufficient data, it is difficult to imagine how digital health can effectively contribute towards achieving UHC. Federated analytics and federated learning provides stepping stone towards addressing this issue. We envisage a health data commons where the FHIR SHR is accessible as a data station that participates in a federated analytics framework. Many mature open source components are available to this purpose.

A new paradigm for creating a health data ecosystem

We propose an integration of the design principles of OpenHIE, FHIR and FAIR for a new paradigm wherein FHIR data stations are the minimal standards for health data sharing. Beyond these design principles, however, it is not necessary to implement and deploy this ecosystem of FHIR data stations in a top-down, large-scale program. The internet was not built that way either. Instead, any health facility, county or consortium can start building and contributing towards implementing FHIR data stations that adhere to the FHIR standards and the principle of the hourglass model. Building the internet of health data, one data station at a time.

[20] openhim.org
[21] See this list of projects known to the openHIE community.
[22] https://hapifhir.io
[23] website FHIR Bulk Data Access, Mandl et al. (2020), Jones et al. (2021).
[24] Rieke et al. (2020)

Demonstrators built for MomCare project

PharmAccess launched MomCare in Kenya (2017) and Tanzania (2019) with the objective to create transparency on the journeys of pregnant mothers. The program is built on three pillars: journey tracking, quality support and a mobile wallet. MomCare distinguishes two user groups: mothers are supported during their pregnancy through reminders and surveys, using SMS as the digital mode of engagement. Health workers are equipped with an Android-based application, in which visits, care activities and clinical observations are recorded. Reimbursements of the maternal clinic are based on the data captured with SMS and the app, thereby creating a conditional payment scheme, where providers are partially reimbursed up-front for a fixed bundle of activities, supplemented by bonus payments based on a predefined set of care activities.

In its original form, the MomCare program did not support interoperable data exchange. In Kenya, M-TIBA is the primary digital platform, on top of which a relatively lightweight custom app has been built as the engagement layer for the health workers. M-TIBA provides data access through its data warehouse platform for the MomCare program. However, this is not a standardized, general purpose API. In the case of Tanzania, a stand-alone custom app is used which does not provide an interface of any kind for interacting with the platform. Given these constraints, to date the MomCare program uses a custom-built data warehouse environment as its main data platform, on which data extractions, transformations and analysis are performed to generate the operational reports. Feedback reports for the health workers, in the form of operational dashboards, are made accessible through the app. Similar reports are provided to the back-office for the periodic reimbursement to the clinics.

Within the this context of the MomCare program, we have built a number of demonstrators (components) to show how interoperability can be achieved using the HDC framework. figure 4 demonstrators illustrates how each of these components are positioned within the framework, where naming of components follows OpenHIE terminology.

- **FHIR adapter mediator**: data from the existing MomCare app is extracted and transformed into FHIR standard. The code is implemented in such a way that it can be readily deployed as OpenHIM mediator;[25]
- **Shared Patient Record in Bulk FHIR format**: data from all patients and participating health facilities in the MomCare program are persisted as Shared Patient Records using the Bulk FHIR data format. While the FHIR standard was originally designed as an interoperability standard for information exchange, we demonstrate its use for persistent data storage as means for realizing the Shared Health Record component. We demonstrate that the data can be stored and made accessible using standard cloud storage technologies to enable secondary reuse of bulk data through federated analytics;
- **Business Services**:
    - **Automated DHIS2 reporting**: existing management reports can be reproduced and generated from the FHIR-transformed data, thereby reducing double entry and opening up possibilities to improve management reporting;

[25] OpenHIM Mediator Library

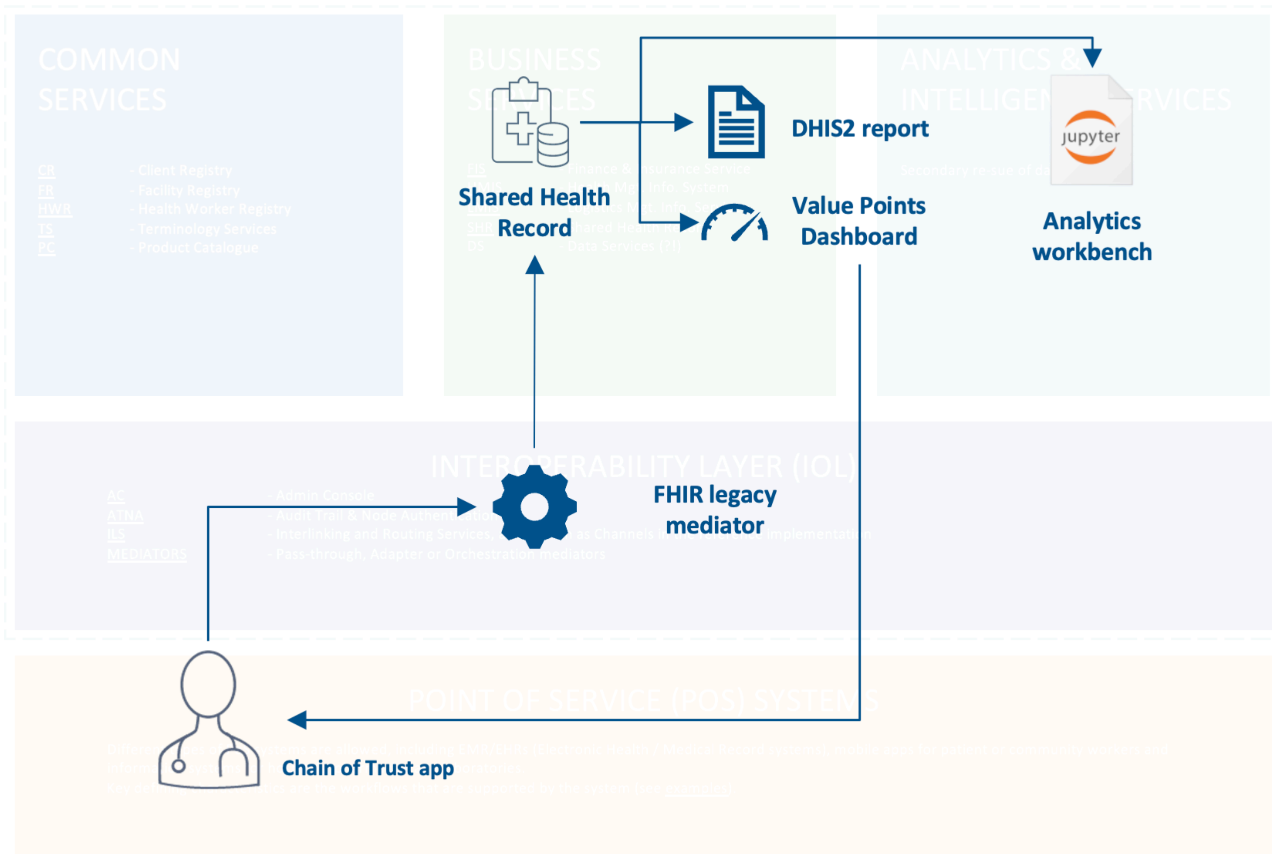


Figure 4.: Demonstrators realized in MomCare Tanzania

  - **Value-Points Dashboards**: value-based care models and conditional payment schemes can be supported using the FHIR-transformed data. By using FHIR, data across different point of care service providers can be share and integrated, thus allowing the tracking of a full care journey from start to finish. Having visibility on the full patient journey is one of the key requirements for support value-based healthcare programs;

- **Federated analytics workbench**: we demonstrate the integration of the Shared Health Record with an federated analytics workbench. We add the global *de facto* standard of interactive, notebook-based computing of the Jupyter project[26] to the openHIE framework.

Above mentioned demonstrators have been build using open source technologies and implemented using standard cloud technologies. Within the setting of the MomCare project, Azure was chosen as the cloud infrastructure on which the components have been deployed, but the components can be deployed on any cloud hosting provider.

Key findings

Required skills and development effort

The HDC project was initiated in Q3 2022, while the majority of the development work was conducted in December 2022 until July 2023. The core team consisted of 5 people with the profiles listed below, with two people taking up the software developer role, and one person for each of the other profiles.

0.0.1. Software developer

- Perform all development work, both the back-end (data platform) and front-end (web app)
- Cross-platform and cross-language development skills are required (Java, Python, Javascript) as the open source components that were just vary greatly
- Also responsible for cloud engineering and managing the production environment (DevOps), in this case Azure

0.0.2. Data scientist

- Develop data processing pipelines, including transformation of source data into FHIR
- Perform data analysis and write standardized queries for the reports
- Develop interactive data visualization components, which were subsequently included in the dashboard app

[26] https://jupyter.org/

0.0.3. Digital architect

- Design overall architectural framework of HDC
- Research and evaluate open source components to be used for specific use-cases
- Writing implementation guides and knowledge transfer to key stakeholders

0.0.4. Solution lead

- Define functional requirements of reports and dashboard via mock-ups
- Interface to key users in the field

Table 1.: Mandays spent on HDC project

Role	Total mandays
Software developer	90
Data scientist	30
Solution lead	10
Digital architect	30
TOTAL	**160**

Table 2.: Estimated recurring development effort required for new use-cases

Use-case	Recurring development effort (mandays)
Implement FHIR mediator	10
Implement new dashboard on FHIR data	10 - 20
Deploy and configure FHIR SHR for a region	20

table 1 shows the amount of time spent on the project. The development effort includes both the initial exploration and development of re-usable components, and the recurring effort required if these components are to be used for a new project. table 2 provides a rough indication what the recurring development effort would be for such new use-cases.

Lessons learned

Many of the lessons we learned whilst working on the MomCare Tanzania case confirm the truisms associated with doing data-related, digital development work. First, even with clear architectural guidelines in hand, there is a wealth of digital public goods and open source software to choose from. Choosing the best fit for a given use-case is often not evident, and in fact should be a continuous effort when scaling up the paradigm proposed here. We feel there is good work to be done in bridging the various worldwide communities, for example, the FHIR community, the OpenHIE community and the FAIR community.

Second, it was that standardisation really does add value and reduce the development overhead. Within a couple of months, we managed to deploy a FHIR data station using the HAPI FHIR server. That data station immediately gave us the possibility to query data uniformly and control access to personally identifiable information. We experienced the benefits of having standardized data for analysing data from different sources. We also found that applying data standards improves visibility on data quality.

Finally, we found that adopting best - or at least common - practices in our development work improved productivity and efficiency. For example, over the course of the project, three different team member contributed and improved on the FHIR translation scripts. Having detailed technical guidelines helped us to handover work efficiently, iterate quickly and reuse code.

Acknowledgements

We gratefully acknowledge the following open source digital public goods without which these demonstrators could not have been realized (alphabetical order):

- Apache Arrow and Apache Parquet are used as language independent data formats for the Shared Health Record component, allowing efficient storage and processing of small to very large datasets.
- duckdb is used as primary database system for performing analytics on Shared Health Record component.
- Project Jupyter as the *de facto* global standard for data analytics using interactive notebooks.
- svelte and svelte.kit frameworks for buidling offline-resilitient progressive web apps.
- Quarto publishing system. This document was authored using Quarto books.
- Vega-Altair and Vega-Lite data visualisation framework, used in creating dashboards and reports.

List of abbreviations

Abbrevation	Definition
API	Application Programming Interface
CR	Client Registry (openHIE component)
CoT	Chain of Trust, the name of the app used in MomCare Tanzania
DHIS2	District Health Information System, version 2, a type of HMIS
FIS	Finance and Insurance Service (openHIE component)
FHIR	Fast Healthcare Interoperability Resources
FR	Facility Registry (openHIE component)
GDPR	General Data Protection Regulation
HL7	Health Level 7 International, the governing body of the FHIR standard
HDC	Health Data Commons
HIE	Health Information Exchange
HMIS	Health Management Information System (openHIE component)
HWR	Health Worker Registry (openHIE component)
IOL	Interoperability layer (openHIE component)
LMIC	Low- and middle-income country
LMIS	Logistics Management Information System (openHIE component)
OHS	Open Health Stack
OpenHIE	Open source community and specification of a HIE
OpenHIM	Open Health Information Mediator, an open source reference implementation of the interoperability layer of openHIE
PC	Product Catalogue (openHIE component)
PET	Privacy-enhancing technology
PII	Personnally Identifiable Information
PoC	Point-of-Care system (openHIE component)

Abbrevation	Definition
PoS	Point-of-Service system, alternative name of Point-of-Care system
PR	Patient Registry, alternative name for Client Registry (openHIE component)
REST	Representational state transfer, a software architectural style used in web APIs
SQL	Structured Query Language
SHR	Shared Health Record (openHIE component)
TS	Terminology Service (openHIE component)
UHC	Universal Health Coverage

Part I.

Narrative

The following chapters describe the narrative and perspective on health data commons, and the context of MomCare Tanzania as a testbed.

1. Envisioning a health data commons

1.1. The need for a data commons

Digital transformations call for a new understanding of the concepts of public health and universal health coverage (UHC).[1] Digital technologies are changing notions of health and wellbeing and offering new tools through which public health goals can be achieved. However, this does not mean that achieving UHC in a digital world will only depend on a rapid pace of adoption of new technologies. On the contrary, given the wide scope of UHC, it becomes clear that achieving UHC in a digital world will inevitably require more than the adoption of new technologies in health and health care as a means of simply increasing efficiency or cutting costs.

The first key question will be whether digital technologies help increase the availability, accessibility, acceptability, and quality of health services as we know them. The second (and related) question concerns the changing nature and direction of health care, and the possibility of making it more preventive, personalised, and mobile through the use of such technologies. Finally, the third question relates to the extent to which digital transformations will enfranchise patients and communities (and particularly vulnerable groups including children and young people) and evolve their relationship with health professionals and providers, thus helping shape the health system according to the needs of the patients and communities. These three dimensions of digitally transformed UHC is visualised in figure 1.1.

To illustrate these dimensions, consider the following examples for low- and middle incomce countries (LMICs).

1.1.1. Example 1

...

1.1.2. Example 2

Digital interventions targeting people with noncommunicable diseases (NCDs) show potential in LMIC settings.[2] mHealth programs designed to support engagement in behaviors associated with diabetes, stroke, and cardiovascular disease management have resulted in improved clinical outcomes, health behaviors, and compliance with treatment, although not all published studies have shown positive

[1] This introduction is largely based on the Governing Health Futures 2030 report by the Lancet & Financial Times Commission.

[2] taken from McCool et al. (2022)

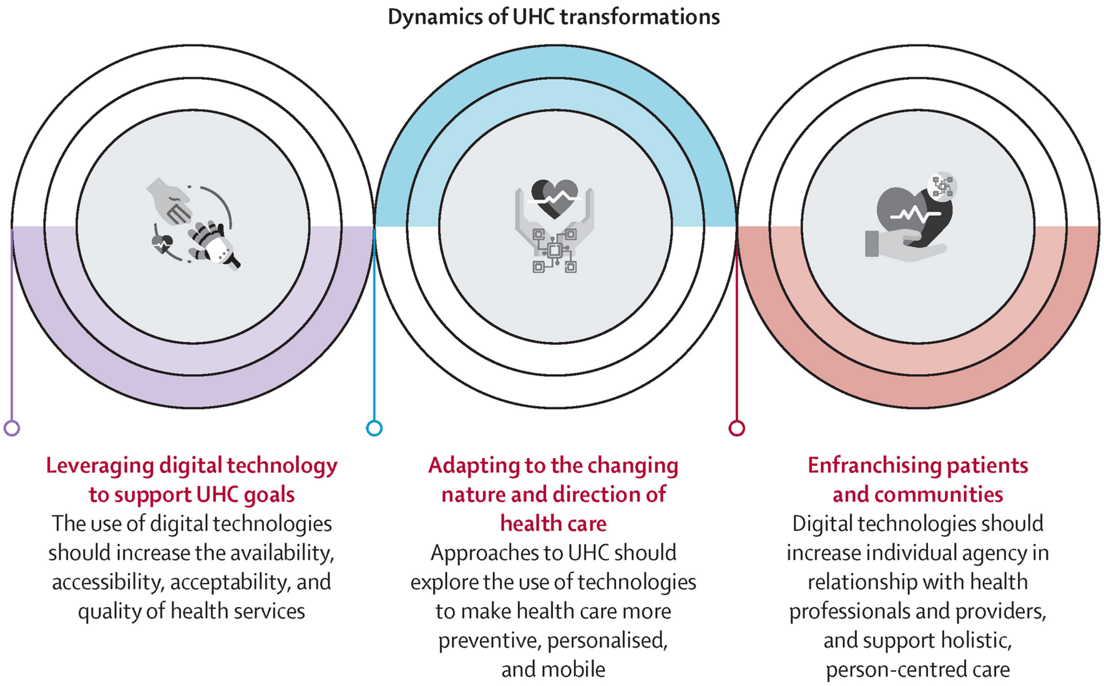


Figure 1.1.: Three key dimensions of a digitally transformed UHC

effects. mHealth has also been used successfully in the remote monitoring of people with long-term conditions and in the provision of personalized medical advice based on the data received. mHealth interventions designed to promote physical activity and healthy diets for NCD prevention have shown promise in LMICs, providing a viable mechanism to improve diet and physical activity behaviors. A review of randomized controlled trials (RCTs) on the effectiveness of mHealth interventions on physical activity and diet outcomes in developing countries reported consistent findings with systematic reviews of mHealth interventions carried out in HICs.

Ongoing examples of digital interventions for NCDs include:

- add example + URL
- add example + URL
- add example + URL

1.1.3. Example 3

...

The notion of utilising digital technologies as a means to catalyse UHC is dependent on a new approach to the collection and use of health data. The Lancet and Financial Times Commission stresses the need for applying

> "… the concept of data solidarity, with the aim of simultaneously protecting individual rights, promoting the public good potential of such data, and building a culture of data justice and equity."

Despite the long-standing adoption of broad strategies and declaration of principles to guide digital transformations, in practice, many countries still do not have effective approaches to digital health that data solidarity at the centre. Choosing who can access data and use it is in fact a central societal question we have to answer. We need to develop a sustainable information ecosystem that shifts the power balance, and control over data, back to the societies. This can be achieved through democratic management of data as a commons. Hence, we take the concept of (health) data commons as one of the foundational concepts with which we seek to align our work.[3]

1.2. Design blueprint for data commons

For data to be governed as the commons, we need an appropriate design of the data ecosystem. Defining the basic elements of a design blueprint for data commons is necessary for a plurality of institutions, initiatives and infrastructures to work together, or at least in parallel, on attaining shared goals. Design of data commons needs to consider three pillars, as shown in figure 1.2.

1.2.1. Stewarding access

Data commons deploy various forms of access to ensure on one hand that generative characteristics of data as a resource are not limited, and on the other that it is shared in a way that is sustainable, preserves rights and minimizes risks. Design decisions that concern Stewarding Access establish the rules and means for deciding who gets to access data and under what conditions. There is a tension here that needs to be resolved: between Open Access commons and stronger, permissioned forms that limit access through more refined governance. Stewardship also entails maintenance of the data and related infrastructures – as ensuring access requires large amounts of effort to collect, store and maintain quality of data.

Sharing framework

There can be no data commons without accessible data. Therefore legal frameworks and tools are necessary to first make the data available as data commons, and then to allow access and downstream uses. Business-to-business (B2B), business-to-consumer (B2C) and business-to-government (B2G) data sharing or data altruism are some of the frameworks that are available. Decisions also concern enabling continuous access. Access may require accepting a specific license that the community adopted for data use.

[3]We follow the definition of a data commons as described by work of Tarkowski and Zygmuntowski (2022).

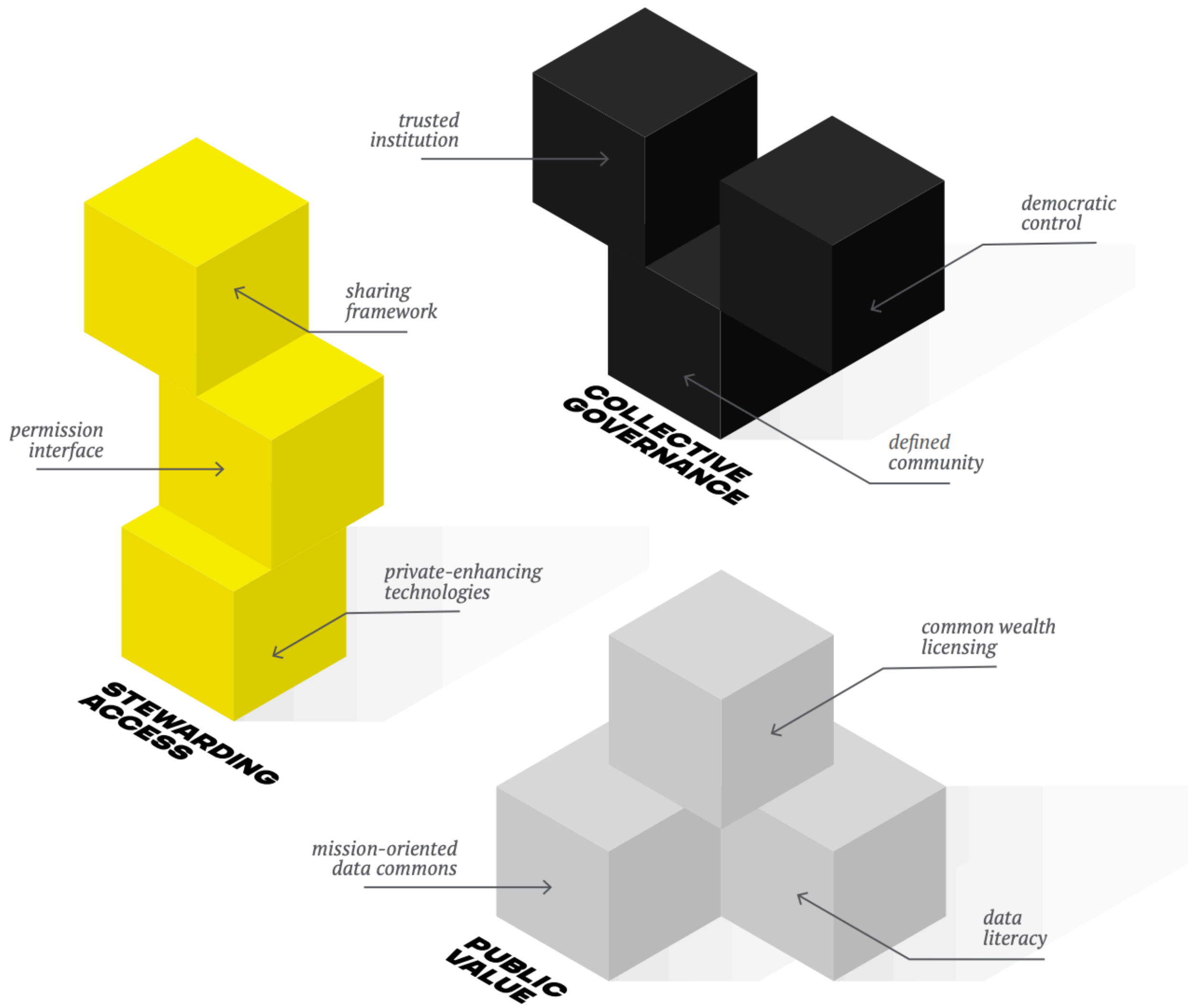


Figure 1.2.: Design blueprint for a data commons

Permission interface

Where Open Access data commons permit everyone to access and use data, other forms of commons need to be based on permissioned access. Thus, a permission interface needs to be designed. The interface may monitor, register and assess impact of requests to access data. Ensuring that the identity of an actor that is requesting ac- cess to data is transparent allows for greater accountability, also in terms of preventing harm and levying sanctions if data commons have been abused. Permissioned access is particularly relevant for creating a health data commons.

Privacy-enhancing technologies

Satisfying data protection by design (GDPR) for personal data can be achieved by conscious architecture choice. Since it is the societal objective that is important, not technological novelty in itself, greater protection of rights should be achieved with privacy enhancing technologies (PETs) such as Open Algorithms (which "move algorithm to data"), federated learning, pseudonimization, distributed vetting ledger and others. While the focus of the data governance debate is on privacy, care should be taken to preserve and enhance other rights as well.

1.2.2. Collective governance

Data commons are linked to the community which manages them, and in many cases generates the data as well. Any other arrangement would constitute an appropriation of the resource, and disempowerment of the people. To establish collective governance over data, there must be either an existing or a newly established entity that can become trusted institutional vehicle for data commons.

Defined community

In order to ensure democratic governance, the community that is the primary holder of rights in data needs to be defined. In this way, collective rights in data can also be better assigned and represented. Yet this is often challenging with regard to digital data, as traditional community or group formation frames do not apply. The challenge lies as much in conceptualizing the community, as in defining the right institutional level of civic life at which the collective interests should coalesce.

Trusted institution

A trusted institutional actor capable of stewarding the commons is a necessary element of data governance design. Data commons institutions are needed due to limitations of both grassroots organizing and market incentives. Institutionalizing the data commons, and thus supporting them with dedicated infrastructure, funding and capacity, renders them independent from market or state pressures.

Democratic control

For the community to have greater autonomy, it has to be directly involved in decision-making. Different forms of democratic participation or accountability can be deployed, including supervisory councils, citizen panels and assemblies, sortition and quadratic voting. Different forms of democratic control can be deployed at all levels of social life, from the local and municipal level to the governance of national datasets.

1.2.3. Public value

A successful data commons strategy needs to take into account not just the management and provision of data, but also the need to ensure that the gener- ated data-based products and services increase the common good. The notion of public value is useful to emphasize concrete, observable benefits produced for the society as a whole, and not just for the community that manages a data commons. By providing public value, data commons can restructure the data value cycle, change the balance of power and introduce a regenerative function to the data ecosystem. A public value perspective also pays attention to positive externalities of data commons, such as increased data literacy or experiences with civic participation.

Mission-oriented data commons

Data commons initiatives should be guided by the values upheld by the community and oriented towards societal goals. Thus, access to data is not a goal in itself, but should lead to socially beneficial uses. A mission-oriented approach ensures that data commons benefit the society in an egalitarian, inclusive manner, for example by prioritizing or incentivising data use for socially important aims.

Common wealth licensing

There is a need to build a new generation of licensing tools that allow access and use rights to be managed in as standardized way as possible. As a general principle, a license for data access and use should aim to build the shared wealth of community, by sharing the products and revenues back with the commons, and with the society – instead of just producing commercial value.

Data literacy

All commons have to remain sustainable by not only regenerating their stock, but also the capacity of the community to continue commoning. In the case of data commons, this means supporting projects of redistributive justice and reducing inequalities in the capacity to obtain value from data commons. Broadly understood data literacy includes not just individual education and training, but increases in the capacity of different actors, institutions and communities to make beneficial uses of data.

1.3. Building a data commons, one data station at a time

The aim of the Health Data Commons (HDC) project is to demonstrate a real-world implementation of such a data commons ecosystem specifically for healthcare using the framework shown in figure 1. This document focuses on the **technical** aspects of the design, implementation choices and learnings from various projects that have been conducted at PharmAccess Foundation since 2022. As such, the work presented here pertains to stewarding access: how can a HDC be implemented that supports a practical framework for health data sharing, incl. permission interface and support of privacy enhancing technologies. To do so, the following design principles have been adopted.

1.3.1. Hourglass model: build a data network, not a single solution

Over the last few years, the importance of interoperability of systems and reuse of data has become evident. One of the key challenges in establishing interoperability the dilemma of how to start: start small, and run the risk of not achieving common standards. Start large, and get bogged down in talking rather than buiding a new standard infrastructure. To tackle this challenge, we follow the concept of the "hourglass" model (figure 1.3). The hourglass model is an approach to layered system architecture where a middle layer is intentionally constricted in order to support flexibility in the implementation of layers above and below. Above the spanning layer are applications, and below the spanning layer are supports. Beck (2019) provides a formal analysis of the hourglass model which states that

> *a weaker layer specification has fewer possible applications but more possible supporting layers than a stronger layer specification.*

Figure 1.3.: The hourglass model as first described by Beck (2019)

We believe that the hourglass model provides a plausible road towards an ecosystem of health data as a common good, where **data sharing is facilitated through a decentralized, federated network of data stations**. These data stations are designed in such a way that it provides the minimal standardization to allow stewarding access all parties is the ecosystem, including healthcare facilities, government organizations, commercial companies etc. The data is not shared through centralised platforms, but is organised through local data stations. In analogy with the Internet, each data station is an independent node that acts as a webserver, interacting with other nodes to create a data sharing network.

1.3.2. FHIR as the de facto data standard in healthcare

Fast Healthcare Interoperability Resources (FHIR) has become the de-facto standard for clinical information exchange in the healthcare sector, both for routine health data exchange[4] and research settings[5], and as is exemplified by the new collaboration between HL7 International (the governing body of FHIR) and WHO.[6] FHIR is increasingly being used in LMICs as well: the WHO reference app for antenatal and neonatal care is built on the FHIR-based OpenSRP digital platform.[7] The mHealth4Afrika project takes a comprehensive community-based approach in co-designing and validating a modular, multilingual system based on FHIR.[8] Ejo Health piloted a solution Rwanda whereby community workers were provided a tablet, and demonstrated time-savings for administrative tasks and improvements in terms of safety and ease of work through digitizing the record-keeping process using the FHIR-based Aidbox platform.[9]

To date, the lion's share of FHIR-related projects in LMICs focus on creating open digital health platforms, that is, achieving openness and interoperability at the level of Point of Service systems The object of openness are the software components themselves, where openness is achieved through open sourcing FHIR implementations, such as HAPI FHIR[10] or through APIs as specified by the FHIR standard. For example, the SMART on FHIR standard provides a consistent approach to security and data requirements for health applications and defines a workflow that an application can use to securely request access to data, and then receive and use that data.[11]

Above and beyond the value of FHIR for open Point of Service systems, we believe that with the recent release of the Bulk Data Access API[12], FHIR can play a pivotal role in enabling open health data commons. The Bulk Data Access API, which by December 2022 has been incorporated in all major FHIR implementations, can handle requests on cohorts with multiple patients rather than just one patient at a time, in an easy to use formatted single file. Combined with the existing semantic interoperability that FHIR provides through its Resources (the data components which allow flexible

[4] Ayaz et al. (2021).
[5] Duda et al. (2022), Vorisek et al. (2022).
[6] https://www.who.int/news/item/03-07-2023-who-and-hl7-collaborate-to-support-adoption-of-open-interoperability-standards
[7] https://smartregister.org/.
[8] http://www.mhealth4afrika.eu/.
[9] https://www.ejohealth.com/, https://sovereignty.network/kickstarter.
[10] https://hapifhir.io
[11] Mandel et al. (2016).
[12] https://hl7.org/fhir/uv/bulkdata/index.html, Mandl et al. (2020), Jones et al. (2021).

composition and combination of a wide range of health data) we now have a means for supporting analytical workflows that require access to and processing of data in bulk. It is this new functionality of achieving semantic interoperability for bulk data through FHIR that we want to bring attention to as a means for building a health data commons.

1.3.3. FAIR data stations as the cornerstone for a data commons

Inspired by the work of van Reisen et al. (2021) in implementing the VODAN-Africa data infrastructure for monitoring COVID-19, we synthesize the concept of the hourglass model and the FHIR standard into “fair data stations” as the foundational element for creating a federative, networked health data commons. FAIR data[13] are data which meet principles of findability, accessibility, interoperability, and reusability. This concept has recently gained traction, particularly in the context of research data. The GO FAIR Initiative lists 18 implementation networks that are currently underway.[14]

Gebreslassie et al. (2023) have analyzed that the FHIR standard can indeed be leveraged for the FAIRification process. They conclude that FHIR as a native solution through the protocols and specifications it supports, or with the community implementation guides, is a viable option for the FAIRfication process of health data. Furthermore, the widespread availibility of FHIR implementations also enables a transition strategy to enable data sharing of non-FHIR-based systems. The FHIR facade model provides a way to transition towards FHIR: rather than creating a FHIR repository to house the required data, the facade model data is fed directly from other repositories and converted to FHIR resources on demand. There are two ways to build a FHIR facade (figure 1.4):

- **storage-less facade** translates FHIR REST calls to queries to the underlying database or services of the original system. The internal information model is mapped to FHIR - find what FHIR resources and attributes represent data structures in your system. Such a facade passes all the calls to the original database.
- **facade with intermediate FHIR server** uses a generic FHIR Server for storing data that is going to be served over API. The same mapping of the internal information model to FHIR is performed but then synchronized data in the FHIR database with the data in the FHIR server that does the rest of the work.

1.3.4. Putting it all together

The design principles outlined above are integrated into the openHIE framework as follows:

- The Shared Health Record (SHR) component is the fair data station, the elemental building block of the data commons which is implemented using the FHIR Bulk API standard;
- The Interoperability Layer (IOL) provides the key functions for connecting a network of data stations, including

[13] Read more on Wikipedia.
[14] See GO FAIR website and specifically the page on implementation networks. Last accessed 15th August 2022.

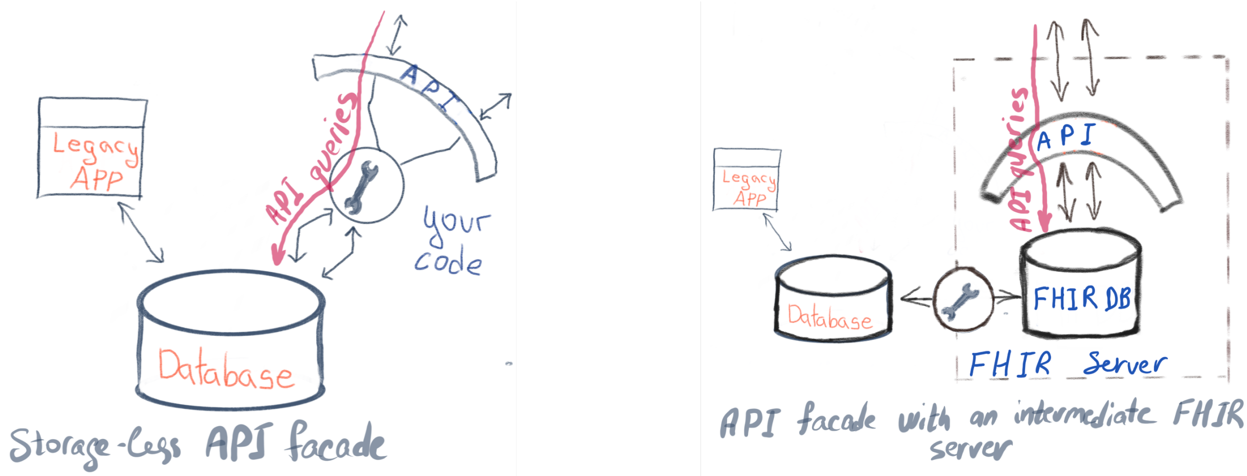


Figure 1.4.: Two ways for implementing a FHIR facade. Taken from Smirnov (2019).

  - Mediators as a storage-less facade for integrating non-FHIR legacy systems;
  - Interlinking and routing services for search, metadata and other discovery services;

- Many of the Common Services will be implemented using the FHIR standard, where components such as Terminology Services and Facility Registries as specified in detail.

In the following chapter we introduce Momcare Tanzania, which is used as the testbed for this architecture.

2. MomCare Tanzania

2.1. The MomCare programme

STORYLINE: FROM SHOWING HOW LOCAL DIGITIZATION/TRANSPARENCY LEADS TO BETTER JOURNEYS TO DEMONSTRATING THAT THIS CAN BE DONE PLATFORM AGNOSTICALLY. AND HOW THIS CAN LEAD TO EVEN BETTER JOURNEYS.

In Sub-Saharan Africa, 200,000 women die annually from pregnancy-related complications, representing 68% of maternal deaths worldwide. Value-based healthcare (VBHC) could help improve maternal and neonatal health outcomes, by promoting a data-driven and patient-centered approach. This presentation will highlight the impact, lessons, and feasibility of implementing VBHC for maternal, neonatal and child health (MNCH) in LMICs., Improving MNCH outcomes requires early and regular antenatal care, facility-based delivery, and postnatal care, tailored to the needs of individual mothers. Therefore, PharmAccess launched 'MomCare' in Kenya in 2017 followed by Tanzania in 2019 [1].

Using the VBHC framework, MomCare incorporates three dimensions of care, namely i) financing for a package of maternal care; ii) quality standards (SafeCare) for healthcare providers and iii) actionable data to incentivize patient and provider behaviors (figure 2.1). Mothers received a digital wallet (M-TIBA in Kenya) that entitled them to a care bundle encompassing the entire pregnancy journey including postnatal and neonatal care. Quality assessment was done through SafeCare during onboarding and repeated every year. Women consented to data- collection through claims / data entry systems, sms-surveys and call surveys. This allowed us to track utilization behavior, well-being, care experience and outcomes throughout the journey as well as risk-mitigation by clinics. We implemented a pay- for-performance model and created dashboards with insights and benchmarking information. Clinics received actionable feedback including risk stratification and an overview of mothers to call to encourage facility-based delivery. Field teams supported clinics through periodical data-based feedback sessions to stimulate continuous quality improvement.

MomCare has been implemented in over 70 clinics across Tanzania and Kenya, supporting over 55,000 mothers. Throughout the program we measured improved adherence to maternal care, better SafeCare scores and improved risk mitigation. Providers actively engaged with data-based insights, actionable feedback, and the pay-for-performance system. We were able to identify unmet needs like mental health support for (teenage) mothers or breastfeeding support and adapted the care bundle accordingly. Importantly, the program brought transparency on risk mitigation and costs. MomCare is proof that VBHC can be successfully implemented in LMICs to improve maternal care.

[1] For more background, see Huisman et al. (2022), Sanctis et al. (2022), Shija et al. (2021) and Mrema (2021)

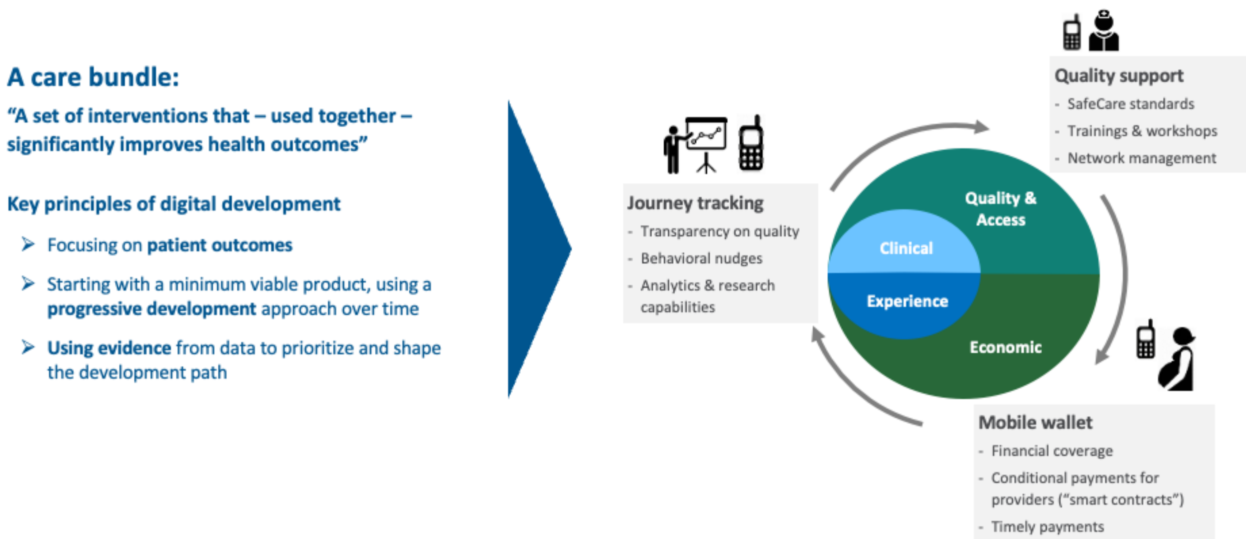


Figure 2.1.: Dimensions of the MomCare programme

2.2. Tracking the journey of pregnant mothers

MomCare distinguishes two user groups, each supported with a specific type of mobile device (figure 2.2). Interaction and communication with mothers is facilitated through a feature phone (figure 2.2a), which enables them to access a mobile wallet, obtain coverage for services including antenatal checkups, delivery (normal & C-section), post-natal checkups and immunizations. Pregnant mother are reminded of their upcoming visits through SMS. SMS is also used to send questionnaires to the mother about her care experience and wellbeing.

Health workers are equipped with an Android tablets through which they can access and use an app in which visits, care activities and clinical observations are recorded (figure 2.2b). The app is designed to give actionable feedback and insights to facilitate daily operations. Questionnaires to assess living conditions and comorbidities are also collected through the app. Note that the exact technical details of the app vary between Kenya and Tanzania, which will be discussed in more detail later.

Finally, providers have access to dashboards that provide insight on their patient population, quality of the journeys and allow benchmarking with other facilities.

!! Better to turn table below into graphic and demonstrate data exchange needs !!

	User	Functionality	Hardware	Software
Kenya	Clinician	Billing *	Paper billing form	-
	Billing clerk	Billing		MTIBA
	Clinician	Recording medical history	Purple booklet	-
	Payment provider	Process payments	Computer	MTIBA & Mpesa (?)

	User	Functionality	Hardware	Software
	Mother	Education, keeping medical history, entitlement	Purple booklet	-
	Kenyan clincian (nurse)	Journey tracking	Computer or Android tablet	Patient Journey App
	Clinic management (Kenya)	Feedback		PowerBI dashboard
	Mother	Receive appointment reminders and give feedback	(Feature) phone	SMS
	Call center agent	Plan and calls and record questionnaire responses	Computer (?)	SurveyID
	Mother	Provide end of journey feedback	(Feature) phone	Call
Tanzania	Clinician	Recording medical history and billing		
	Payer (PAF)			Excel
	etc			
	etc			

Reimbursements of the maternal clinic are based on the data captured with SMS and the app, thereby creating a conditional payment scheme, where providers are partially reimbursed up-front for a fixed bundle of activities, supplemented by bonus payments based on a predefined set of care activities. TO DO: explain conditional pyament scheme here (figure 2.3).

2.3. Payment Model Types

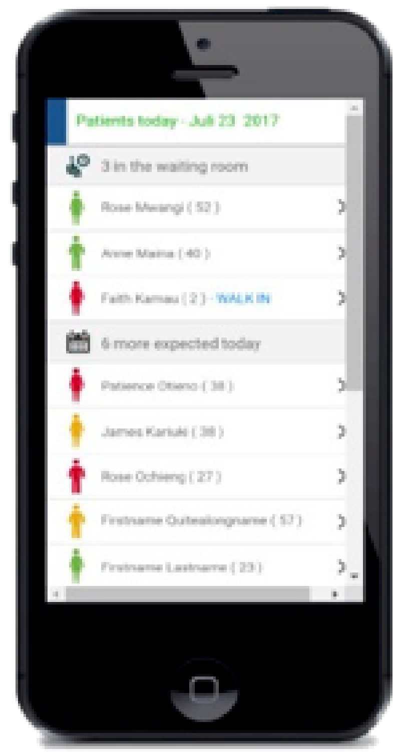
2.3.1. Payer / clinic perspective

- per-visit-bundles
- journey score
- value points
- capitation
- incentives for mothers TO DO: elaborate above

2.3.2. Mother perspective

- quotes from mothers (Liam)

(a) Feature phone used by pregnant women



(b) Android tablet used by health workers

Figure 2.2.: Mobile devices used in MomCare programme

2.4. Care Model Interventions

- ambulance
- Moms2Be
- mental health

TO DO: add graphs with results.

2.5. Data collection

- MTIBA
- PJA
- CoT
- SMS
- End of Journey calls

As fragmented as the above data collection appears, this does represent

2.6. MomCare App & Dashboards

- data collection (CoT)
- operational support (PJA & CoT)
- Dashboards

2.7. Current limitations and challenges

Manyara region has reques Challenges: TO DO: explain the main drivers/issues that we want to address with the demonstrator Hanang wants full district view and we are looking at sustainability: MomCare apps are maternity only + custom made = conflict Various systems in public / FBO/

2.7.1. Lack of visibility outside MomCare network

Pregnancy starts and continues at home

- may be in community health system
- may already need advice or mental healthcare that can be given remotely #### Delivery / referral elsewhere
- we don't see the mother

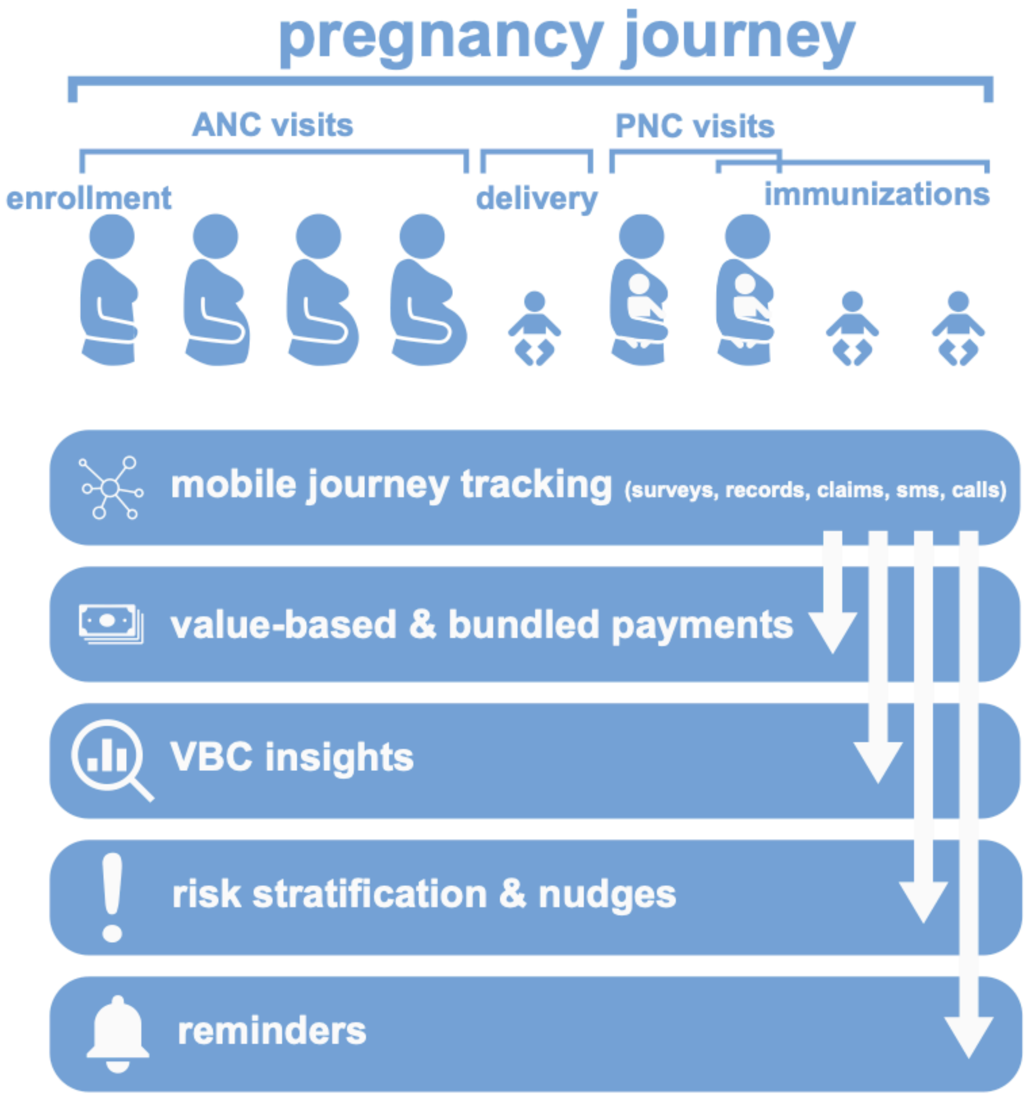


Figure 2.3.: pregnancy journey

- providers don't see her data leading to worse care Pregnancy journeys are more than just clinic visits. Women tend to be pregnant for a long time before they have their first ANC and may already have a need for (mental) healthcare advice or information on pregnancy. #### Real-life implementation requires combining data services from multiple individual services
- entitlement
- 'brick' clinical services
- 'click' clinical services
- community clinical services
- payments, visibility and execution
- PROMs collection

2.7.2. Non-standard implementation stand in the way of scaling up data exchange

In its current form, the MomCare programme uses custom-made digital platforms. In Kenya, M-TIBA is the primary digital platform, on top of which a relatively lightweight custom app has been built as the engagement layer for the health workers. M-TIBA provides data access through its data warehouse platform for the MomCare programme, however, this is not a standardized, general purpose API. In the case of Tanzania, a stand-alone custom app is used which does not provide an interface of any kind for interacting with the platform. Given these constraints, to date the MomCare programme uses a custom-built data warehouse environment as its main data platform, on which data extractions, transformations and analysis are performed to generate the operational reports. Feedback reports for the health workers, in the form of operational dashboards, are made accessible through the app. Similar reports are provided to the back-office for the periodic reimbursement to the clinics.

2.7.3. Double-entry

In summary, if we want to scale-up, we want to standardize this approach. This leads us to the demonstrator in Tanzania, explained in the next chapter.

2.8. MomCare Tanzania as our testbed

MomCare Tanzania has bee operational since 2019. Initially, the data gathering and feedback was part of a performance based program called 'Tumaini la Mama'. When this program ended, the journey tracking became part of the MomCare programme. Recently Manyara regional authority (see figure 2.4) has approached PharmAccess to continue providing data and transparency on pregnancy journeys and scale it up to the full district of Hanang (approximately 17.000 pregnancies per year). The district is reported to have low performance in maternal indicators as compared with other districts in the region. There are 35 clinics in the district:

- 9 clinics already in MomCare
- 18 with reasonable connectivity – can onboard without changes
- 8 with intermittent connectivity – need an upgrade of the app to onboard

Figure 2.4.: Manyara region, Tanzania. The regional capital is the town of Babati. According to the 2012 national census, the region had a population of 1,4 million, which was lower than the pre-census projection of 1,5 million.

Part II.

Demonstrators

In the following chapters we will describe the various components for which we have implemented demonstrators within MomCare Tanzania.

3. FHIR mediator

3.1. The Interoperability layer (IOL)

An interoperability layer (IOL) is a system that enables simpler interoperability between disparate information systems. In the context of the health data commons platform we seek to demonstrate here, the IOL aims to connect the Point-of-Service systems, the commons services and business services. As such, the interoperability Layer receives transactions from client systems, coordinates interaction between the different components of the ecosystem and provides common core functionality to simplify the data exchange between systems.

The openHIE framework provides a detailed specification of the interoperatbility layer.[1] Within the scope of our demonstrator projects, the following functional requirements are the most relevant:

3.1.1. Core functions

- Provides a **central point of access** for the services of the HIE. For example this interface will provide access to the CR, PR, FR and SHR. This central point of access simplifies security management and provides a single entry point into the HIE.
- Provides **routing functions** that allow messages to be routed to the correct service provider systems within the HIE
- Provides a **central logging mechanism** for the messages sent through the exchange. This function should log copies of the messages that travel through the interoperability layer for audit and reporting purposes.

3.1.2. Mediation functions

- **Adaptor mediators** accept a request and transform/adapt that request into another format before sending the request on to its final destination
- **Orchestration mediators** accept a request and uses that request to execute some business process. This could involve making webservice calls to one or more other services to gather additional information about the request or to process it further. Finally a response is collated and returned to the OpenHIM.

The technical implementatation specification of OpenHIM states that mediators can be built using any platform or programming language, with the **only** restriction that it interfaces with openHIM via RESTful APIs.[2] In the following we demostrate a FHIR adaptor mediator, whereby data from

[1] OpenHIE Interoperability Layer - Use Cases and Requirements.
[2] OpenHIM documantation Mediators.

the PoS system (in this case, the CoT app of MomCare Tanzania) can be extracted and stored in the SHR using FHIR version 4. This approach is in fact identical to the storage-less FHIR facade solution pattern described in Section 1.3.3. As the OpenHIM platform provides many of the core functions, such as authentication, routing, logging etc., our demonstrator focuses on mapping and translating a legacy system onto the semantics of FHIR Resources.

3.2. FHIR adaptor for MomCare Tanzania

The main objectives of this demonstrator were to assess the complexity and effort required to implement i) the data transformation scripts, to be used for standardized reporting and secondary analytics (see demonstrators in Chapter 5 and Chapter 9); and ii) the workflow orchestration required for the FHIR adapter. In view of simplicity and speed of development, we have chosen to implement the logic of the mapping in SQL, such that the extraction and mapping of the FHIR Resources could be performed directly on the CoT database (MySQL version 5). This approach is in line with ongoing effort within the FHIR community to support JSON data in modern database engines.[3]. Since nearly all major database systems support JSON, the approach demonstrated here can be applied to most relational database systems.

The results of the data transformation into 10 FHIR v4 resources is shown in table 3.1, listing the number of records per resource type. The conceptual data model of the existing MomCare app could readily be transformed into the semantics of FHIR Resources using SQL and validated with a FHIR Python library.[4] The largest challenge during the transformation process pertained to the absence of unique business identifiers for patients and healthcare organizations. For patients, either the mobile phone number or the healthcare insurance number was taken, depending on availability. A combination of name, address and latitude/longitude coordinates were used to uniquely identify organizations and locations, as Tanzania does not have a system in place for this purpose.

Table 3.1.: Mapping of data model of Chain of Trust app into FHIR Resources (v4).

CoT concept	FHIR v4 resource	Business key	# records
Patient	Patient	`patients.id`	28,161
Patient.EDD	Observation(Pregnancy: EDD)	reference to `patients.id` and `effectiveDateTime = patients.updatedOn`	28,587
Clinics	Organization	`clinics.id`	70
Clinics.lat/lon	Location	Lat-lon coordinates	70

[3] SQL-on-FHIR working group.
[4] fhir.resources.

CoT concept	FHIR v4 resource	Business key	# records
Enrollments	EpisodeOfCare	`enrollments.id`	20,571
Visit	Encounter	`visits.id`	174,998
Diagnoses	Condition	`CONCAT(visits.id, diagnoses.id)`	157,162
Asset	Procedure, MedicationAdministration	visit_diagnoses table links clincal findings to Encounters	1,098,129
Appointment	Appointment	tbd	tbd
BaselineSurvey, Survey	Questionnaire	`baseline_surveys.id`, `surveys.id`	4
BaselineAnswers, SurveyResponse, Questions	QuestionnaireResponse	`baseline_answers.id`, `survey_responses.id`, `questions.id`	tbd

As an example, the SQL code for extracting the Patient resource is shown below.

```
SELECT
JSON_MERGE_PATCH(
  JSON_OBJECT('resourceType', 'Patient'),
  JSON_OBJECT('identifier',
    CASE WHEN p.mAfyaCardNumber IS NOT NULL AND trim(p.mAfyaCardNumber) <> '' THEN
      JSON_ARRAY(
        JSON_OBJECT(
          'use', 'temp',
          'value', REPLACE(p.ID, ' ', '-') -- FHIR.Patient.identifier.value need to match regex "
        ),
        JSON_OBJECT(
          'use', 'usual',
          'type', JSON_OBJECT(
          'coding', JSON_ARRAY(
            JSON_OBJECT(
              'system', 'http://terminology.hl7.org/CodeSystem/v2-0203',
              'code', 'NIIP',
              'display', 'National Insurance Payor Identifier (Payor)'
            )
          )
        ),
        'value', REPLACE(p.mAfyaCardNumber, ' ', '-') -- FHIR.Patient.identifier.value need to ma
        -- can't add system because Tanzania National Health Insurance Fund doesn't have URI
      )
```

```
  )
  ELSE
    JSON_ARRAY(
      JSON_OBJECT(
        'use', 'temp',
        'value', REPLACE(p.ID, ' ', '-') -- FHIR.Patient.identifier.value need to match rege
        -- can't add system because Tanzania National Health Insurance Fund doesn't have URI
      )
    )
  END
),
JSON_OBJECT('active',
  if(p.active = 0, cast(FALSE as json), cast(TRUE as json))
),
JSON_OBJECT('name',
  JSON_ARRAY(
    JSON_OBJECT(
      'text', CONCAT_WS(' ', TRIM(p.firstName), TRIM(p.lastName)),
      'family', TRIM(p.lastName),
      'given',
        CASE WHEN p.middleName IS NOT NULL AND trim(p.middleName) <> '' THEN
          JSON_ARRAY(
            REPLACE(p.middleName, ' ', '-'), REPLACE(p.lastName, ' ', '-')
          )
        ELSE
          JSON_ARRAY(
            REPLACE(p.lastName, ' ', '-')
          )
        END
    )
  )
),
JSON_OBJECT('telecom',
  CASE
    WHEN p.phoneNumber is not null THEN
      JSON_ARRAY(JSON_OBJECT('system', 'phone', 'value', p.phoneNumber, 'rank', 1))
    WHEN p.alternatePhoneNumber is not null THEN
      JSON_ARRAY(JSON_OBJECT('system', 'phone', 'value', p.phoneNumber, 'rank', 1))
  END
),
JSON_OBJECT('gender', 'female'),
JSON_OBJECT('birthDate', p.dateOfBirth),
JSON_OBJECT('contact',
  JSON_ARRAY(
```

```
      JSON_OBJECT(
        'organization', JSON_OBJECT(
          'reference', CAST(c.id AS CHAR(50)),
          'display', CONCAT_WS(', ', c.name, c.locality, c.region),
          'type', 'Organization'
        )
      )
    )
  )
) AS patient

FROM
  PATIENTS AS p
  LEFT JOIN CLINICS AS c ON p.clinicId = c.id
WHERE
  p.lastName <> ''
```

While this query may seem verbose, writing such extraction scripts can be done within typically one day provided the developer/data analyst has prior knowledge of the data model of the PoS system. In this example, three JSON functions are used to construct the Patient resource, element by element:

- `JSON_OBJECT` is used to extract `{key: value}` from the relational database. By nesting `JSON_OBJECT` function calls, the FHIR resource is constructed element by element;
- `JSON_ARRAY` is used to enable list of values for a given key, for example, multiple names.
- `JSON_MERGE_PATCH` is used to handle missing values: the FHIR specification is quite strict that `{key: value}` pairs should only be included if there is in fact a value recorded in the system. Using `JSON_MERGE_PATCH`, all keys that don't contain a value are discarded.[5]

Besides the SQL scripts, the mediator is packages in a Docker container such that it can be deployed as a microservice. Specific for MomCare Tanzania the FHIR adaptor was put in production on the Azure cloud platform. More Source code for this demonstrator is available on GitHub.[6]

3.3. Details per resource

3.3.1. Patient

- Insurance number is incomplete, but in theory only insured mothers are treated
- User can choose which `externalID` is used: QR-code (patientID), phonenumber or insurancenumber
- Expected Delivery Data in Observation(Pregnancy)

[5] see this discussion on technical details of handling missing values in FHIR.
[6] https://github.com/PharmAccess/datacare-fhir-transformation-container

- `expectedDeliveryDate` contains most recent date entered; `EDDold` contains previously recorded EDD
- As of summer 2022, `loc_id` of the sub-village is recorded
- Sometimes patients lose their QR code. If we can't find/reconstruct this code, a new QR code i.e. patientID is generated

3.3.2. Observation

FHIR International Patient Summary defines specific Observations related to pregnancy:

- Expected Delivery Date (EDD)
- Outcome
- Status

3.3.3. Clinics

- iProvideID is from PharmAccess database used across different programmes
- lat-lon coordinates are used for uniquely idenfiying clincs, as there are no official codes to identify healthcare providers

3.3.4. Location

- Hierarchy of regional subdivision: country (0) –> region (1) –> district (2) –> division (3) –> ward (4) –> village (5) –> sub-village
- ISO-3166-2:TZ has codes for regions (wiki)
- No official codes exist for divisions and below, so we stick to names
- shapefiles up to ward level (3) are available (link)
- Too many missing values, not used in MVP |

3.3.5. EpisodeOfCare

- `endDate` taken as expectedDeliveryDate + 20 weeks, which is standard follow-up period

3.3.6. Encounters

- Actual visit of pregnant woman at clinic site |
- As of 2022, `immunization` is added as new encounter type; prior to that, immunizations were registered as `pnc`

3.3.7. Condition

- visit_assets table contains all procedures performed and/or medication administered
- Mapping internal diagnoses: A1, A2 and A3 are internal codes. These are mapped as follows
  - A1 –> Z34, Supervision of normal pregnancy
  - A2 –> O80, Single spontaneous delivery
  - A3 –> Z39, Postpartum care and examination

3.3.8. Appointments

- Not linked to Encounters
- Not mandatory to fill in nor update status

3.3.9. Referrals

3.3.10. Questionnaire

Questionnaire data is coming from different resources in the COT database. The Baseline survey tables are the result of questionnairs taken via the app. The Survey tables of questionnaires taken via SMS and the questions table **TODO**.

3.3.11. QuestionnaireResponse

3.3.12. TO DO:

- Assets `isResource` pertains to consumable goods
- Referrals are incomplete, records referral to other provider for e.g. abortion
- Users care Caregivers
- Risklevel is derived from risk-scoring model in datawarehouse

3.4. Concluding remarks

We have demonstrated that extracting FHIR Resources from a PoS is viable by directly querying the relational database using JSON functions. This approach can be replicated for other PoS systems and can be integrated into OpenHIM as a mediator.

4. Shared Health Record

4.1. A FHIR-based Shared Health Record

The Shared Health Record (SHR) facilitates the sharing of clinical information between health information systems to enable better patient care, thus improving health outcomes.[1] The SHR is a means of allowing different services to share health data stored in a centralized data repository. It contains a subset of normalized data for a patient from various systems such as an electronic medical record or the Laboratory Information Management System. This record is queried and updated between the different institutions and systems that are authorized to do so. The SHR is distinct from a data warehouse; it is an operational, real-time transactional data source.

A shared health record is normalised if all metadata items such as patient, provider, and facility identifiers are resolved to appropriate universal identifiers (as opposed to their local identifiers as used by a client system). In addition, all terminology codes in use need to be mapped to an appropriate reference terminology to ensure that the information is consistently understood.

One of the key choices for the HDC platform is to use use the FHIR standard to realize the SHR service. In doing so, a number of design choices need to be addressed.

4.1.1. Using FHIR as a persistent data store

FHIR was originally designed for exchanging data only. With the advent of Bulk FHIR API, it is increasingly being used as a standard for persisting a combined record like the SHR specification. In fact, by implementing a FHIR facade with an intermediate server as described in Section 1.3.3, one can imagine that such an intermediate server can in fact function as the SHR if it were to collect and integrate data from various (legacy) Point-of-Service systems. Taking this one step further, so-called approaches for "analytics-on-FHIR" are under active development. The Open Health Stack (OHS), for example, includes detailed designs and reference implementations for so-called FHIR Analytics shown in figure 4.1. These solution aim to reduce the impedence mismatch between the heavily nested structure of FHIR Resources and the need for having the data readily available in tabular format for analysis using SQL. Many major cloud providers provide persistent FHIR data storage solutions, including Google FHIR Store, Azure Health Data Services and Amazon HealthLake.

It is within the context of these technological developments that we have chosen to adopt FHIR as the core standard for SHR. Nevertheless, it is Pedrera-Jiménez et al. (2022) : FHIR not the preferred choice for persistent record. Still, we do take that approach because - our use-cases are relatively simple

[1]This introduction is taken from OpenHIE.

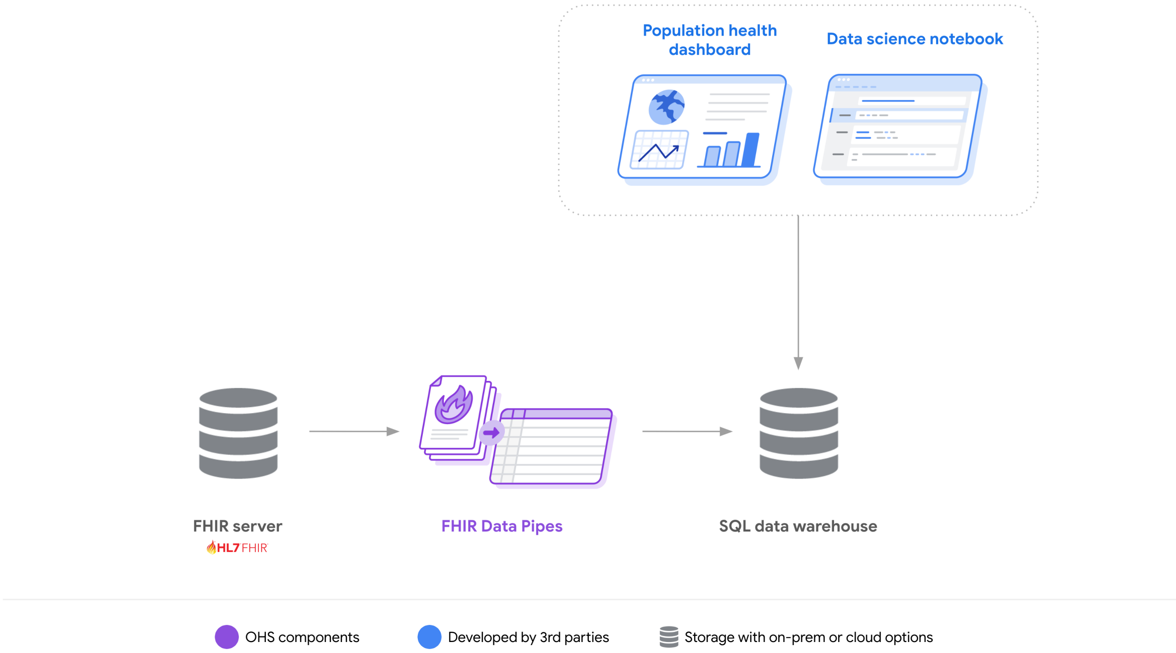


Figure 4.1.: Open Health Stack FHIR Analytics

–> FHIR semantics are sufficient, no need to more detail clinical models - bet on wide-scale adoption by industry and vendors, much more so than OpenEHR and ISO - We take potential technical debt of upgrading versions in future for granted

4.1.2. Versions and profiles

- Choose for IPS as profile plus additions
  - Encounter
  - ServiceRequest: referrals
  - Questionnaire
  - QuestionnaireResponse

4.1.3. Resources versus Bundles

Certainly! In FHIR (Fast Healthcare Interoperability Resources), there is a distinction between FHIR resources and FHIR documents. Let's delve into the details:

1. FHIR Resource: A FHIR resource is an atomic unit of information in the FHIR specification. It represents a discrete piece of healthcare-related data. Each FHIR resource is designed to capture specific types of information, such as patient demographics, allergies, medications, lab

results, or clinical observations. Examples of FHIR resources include Patient, Observation, MedicationStatement, AllergyIntolerance, and Condition.

FHIR resources follow a standardized structure defined by the FHIR specification. They consist of a set of elements that hold the data, including attributes, extensions, and references to other resources. Resources can be uniquely identified and can be accessed, created, updated, or deleted independently. They are designed to be granular and reusable, allowing for efficient exchange and interoperability of healthcare data.

2. FHIR Document: A FHIR document, also known as a FHIR bundle, is a collection of FHIR resources grouped together for a specific purpose or context. It serves as a container for organizing related resources into a single entity. FHIR documents are used to transmit or exchange multiple resources as a single unit.

A FHIR document can include different types of resources, such as patient records, clinical notes, diagnostic reports, or care plans. For example, a discharge summary document may contain a Patient resource, Condition resources representing diagnoses, MedicationStatement resources for prescribed medications, and other relevant resources.

FHIR documents provide a convenient way to package and transmit multiple resources together, maintaining their relationships and context. They allow for the exchange of comprehensive and structured healthcare information, facilitating interoperability between systems and supporting clinical workflows.

In summary, a FHIR resource represents a discrete piece of healthcare data, whereas a FHIR document is a collection of related resources grouped together for a specific purpose or context. FHIR resources provide granular and reusable units of information, while FHIR documents enable the transmission and organization of multiple resources as a single entity.

4.2. Chosen technologies for data layer

- SQL
- duckdb
- arrow & parquet, reference Liu et al. (2020)
-

5. FHIR-based reporting

This section describes the functionality that is implemented to generate standardized tables and reports from fhir resources.

5.1. Base reporter: Standardized tables

We provide a base reporter that extracts the most important information for the momcare project from the fhir resources into standardized tables. These standardized tables serve as a starting point for further analysis.

Patient timeline table

In the MomCare project, the patient's journey over time is tracked. This means that most important information would be captured with a table that tracks each procedure, treatment, or diagnosis that a patient receives over time. For this reason we implemented a functionality to create the patient timeline table.

The *patient_timeline* table consists of all procedures and diagnosis that a patient is subjected to over time. It is created based on the following fhir resources:

- patient
- procedure
- condition
- encounter
- observation
- organization

The *patient_timeline* is of the following format:

patient_id	encounter_id	visit_id	provider_name	procedure_date	code	code_system	type	description	visit_Type	visit_type_code	visit_type_code	visit_type_no	gestation_week
12345678	34	52	Dawar Dispensary	2022-10-10	O98.7	ICD10	condition	HIV in pregnancy	Antenatal care	1245250	201	2	20

patient_id	encounter_id	visit_id	provider_name	procedure_date	code	code_system	type	description	visit_Type	visit_type_code	visit_no	visit_type_no	gestation_week
12345678	34	52	Dawar Dispensary	2022-10-10	NaN	snomed	procedure	HIV test	Antenatal care	424525001	1	1	18
12345678	123	2	Ngarenanyuki RC Health centre	2022-11-18	Z34	ICD10	condition	HIV test	Postpartum care	424525001	1	1	44
12345678	123	2	Ngarenanyuki RC Health centre	2022-11-18	NaN	snomed	procedure	Malaria Rapid Test	Antenatal care	133906008	3	3	30
12345678	123	2	Ngarenanyuki RC Health centre	2022-11-18	NaN	snomed	procedure	Blood pressure	Antenatal care	424525001	1	1	8

next to the patient_timeline also some basic information about the patient is of importance, for this reason the *patient_information* table is created:

patient_id	birthDate	enrollment_date	age_at_enrolment	age_group	delivery_date	expected_delivery_date
abc123def456	2002-12-10	2023-04-22	20	AGE_GROUP_20_24	NaT	2023-09-06
987654321	1990-09-10	2020-04-01	29	AGE_GROUP_25_29	2020-10-01	2020-10-08

For further analysis those two tables are often combined into one *patient_timline_combined* table

questionnare_response table

Besides the patient's yourney the patient is also requested to answer different questionnaires. The DataCare module contains the functionality to flatten the fhir Resource questionnaireResponse to create a table containing all questionnaires' questions and answers for each patient. This table is called the *questionnaire_response* table and is of the following format:

patientId	ClinicName	date	survey_id	question_id	question	answer	...
12345678	Tumaini Hospital	2021-10-07	1	44	Do you have a previous C-Section scar?	false	...
12345678	Tumaini Hospital	2021-10-07	1	42	Is this your first child?	true	...
12345678	Tumaini Hospital	2021-10-07	2	36	At what age did you have your first child?	<20	...
12345678	Tumaini Hospital	2021-10-07	2	35	How many times have you been pregnant before?	0	...

5.2. Usage Base Reporter

Based on the base reports described above, different reports of interest can be created:

- **DHIS2 reports** provide insight in how many mothers visit a clinic, what conditions occur amongst mothers that visit that clinic, and what procedures mothers obtain in a clinic.
- **Value point reports** provide the value points (described in the *value points* section) obtained by clinics by threathing their patients.

In figure 5.1, a high level overview is given of how the base reports provide input for the DHIS2 and value points reports

5.2.1. TODO

In the analytics workbench an example note book is provided, showing how to use the base reporter.

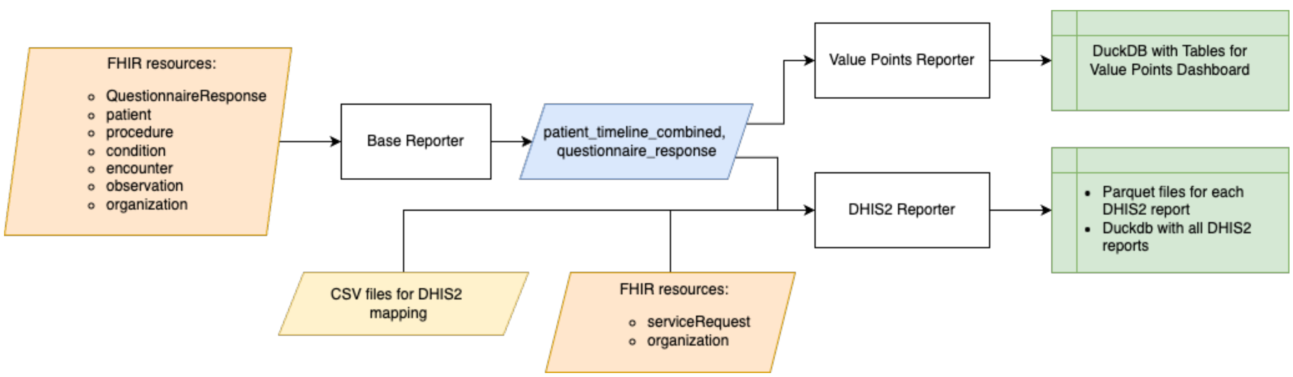


Figure 5.1.: Reporting schema, *orange: 'raw' fhir data, yellow: external data, blue: intermediate data, green: processed data*

6. DHIS2

DHIS2 is a tool used for collection, validation, and presentation of patient-based statistical data. To use this tool, we need tables in a standardized format that can be used by DHIS2. From now on we call those tables DHIS2 reports. In addition, governments use the DHIS2 reports to obtain insights in the performance of different clinics. For MomCare three DHIS2 reports are of interest:

- Antenatal Care (ANC) Report, containing all statistics about care for pregnant women before delivery of their child.
- Delivery Report, containing statistsics about the care and health of the mother during the delivery, and informartion about the child
- Postpartum Care (PNC) Report, containing statistics about the care of the mother after delivery of her child. This report usually also contains statistics about he child. However, this information was not available and for that reason the PNC report created is not complete.

The DHIS2 reports are of the following format:

serial_number	description	year	month	clinic	age_group_10_14	age_group_15_19	age_group_20_24	age_group_25_29	age_group_30_34	age_group_35_plus
1	Mothers expected /projected to deliver at service station	2022	7	Charlotte Hospital	0	0	1	0	1	0

Where the description shows the DHIS2 topic, and for each clinic and each age group the occurence in that year and month is counted.

6.1. DHIS2 reporter

All code to create the DHIS2 reports can be found the Datacare-fhir-data-insights-cot repository. The COT data was first converted into standardized FHIR resources. The Fhir resources were then processed into standardized tables such as a patienttimeline, which lists all procedures and diagnosis that the mother received over time. The conversion of Fhir data in the standardized reports occurs with the base reporter module as described in the standardized reporting section. Based on these standardized reports, the DHIS2 reporter module is used to create the DHIS2 reports.

All Fhir resources required for the DHIS2 reports are:

- patient, providing information about the age of the mother
- condition, providing information about all diagnosis the mother received
- procedure, providing information about all procedures the mother received, and information about the time of each procedure
- encounter, providing information about the type of visit (ANC, delivery, or PNC)
- observation, providing information about the expected delivery date
- organization, providing information about the clinics that the mother visisted
- serviceRequest, providing information that indicates whether the mother was referred and for what reason
- questionnaireResponse, providing all questionnaire responses of the mother

In figure 6.1 the flow to create a DHIS2 report has been visualized.

After runnning the base reporter module †o create the base tables, the DHIS2 reporter module is used to create the ANC, delivery, and PNC DHIS2 report. Within the DHIS2 reporter the following actions take place:

- **Creation of mapping tables:** A mapping between DHIS2 topic and the corresponding condition/procedure or other characteristic in the data was created in the form of CSV files. These files make it easy to map rows in the *questionnaire_response* and *patient_timeline* tables to topics in the DHIS2 reports. These CSV files are converted to SQL tables to use in the DHIS2 reporter. This section is dashed, because it is only required one time to load the mapping tables in the database. However, when the mapping files would be updated, this section should be ran again.
- **Enrichment of the patient timeline:** The patient_timeline table is enriched with extra information that is of importance to create DHIS2 reports.
- **Coupling DHIS2 topics:** The DHIS2 topics are coupled to the patient timeline and questionnaire response data using the mapping tables. In addition for mor and separate functions for more unique topics.
- **Creation DHIS2 reports:** The DHIS2 report is created based on the coupled patient_timeline and questionnaire data.

6.1.1. DHIS2 topics

For each topic in the DHIS2 reports a calculation or filtering of the patients to which the condition applies is required. In this section, we describe for each topic how it was obtained from the data.

The serial number and the description refer to the topics in the DHIS2 reports. The code is the code given to a procedure or condition, the system is the code system, in this case SNOMED for

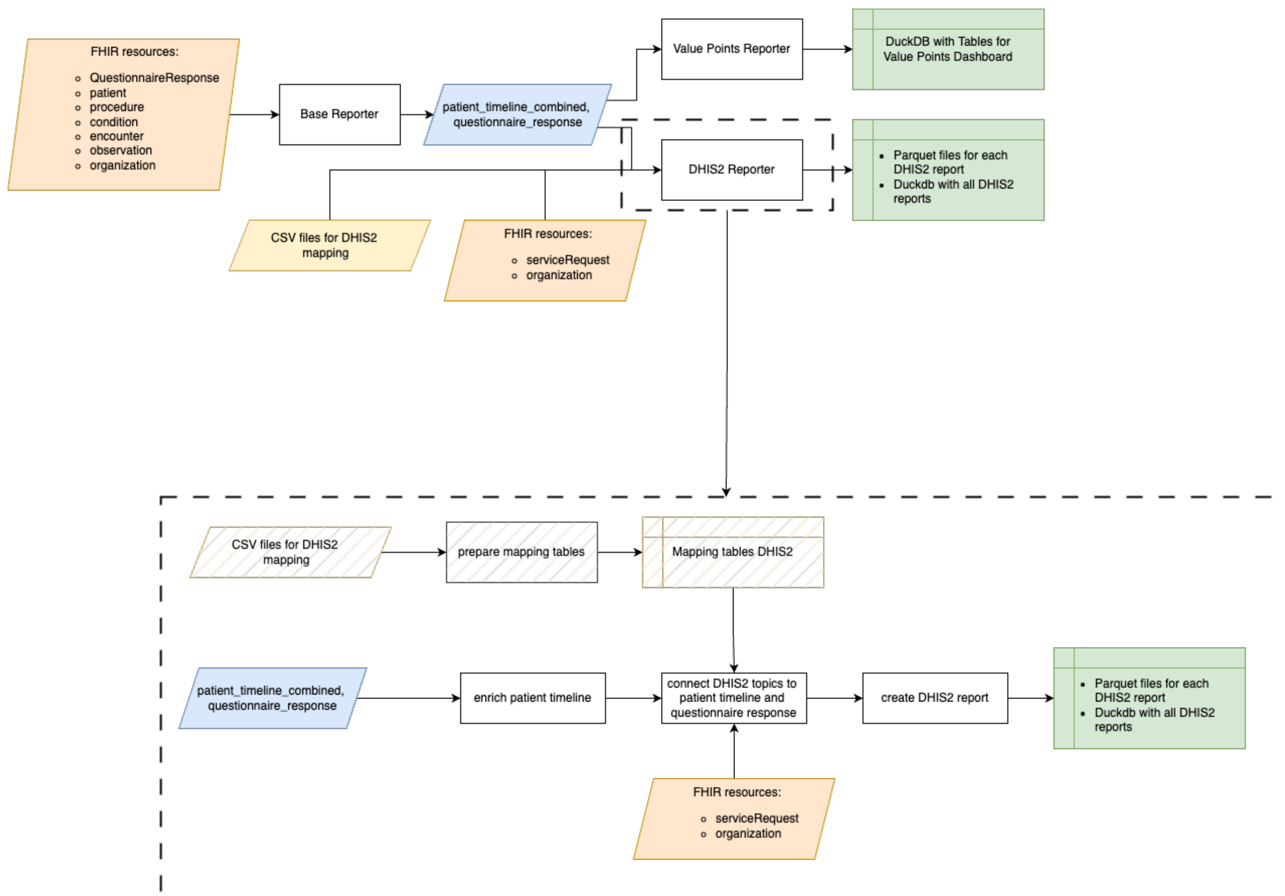


Figure 6.1.: DHIS2 reporting flow and required input, dashed blocks: optional actions, orange: 'raw' fhir data, yellow: external data, blue: intermediate data, green: processed data

procedures and ICD10 for conditions, origin refers to the fhir resource where the information can be found, the original description the description given in the COT system, and the extra filter contains exta information of importance for calcultion of the topic.

For questionnaire responses this includes the answer to the question. for many topics the VISIT TYPE (ANC, PNC or delivery) is of importance, therefore the VISIT_TYPE is often used as an extra filter. This information comes from the ENCOUNTER_RESOURCE. Besides this, for procedures and conditions, we have added calculated fields:

- VISIT_TYPE_NO: the number of visit for a specific visit type (anc, pnc delivery)
- VISIT_NO: the number of visit (each visit counts)
- HIV_TEST: counts if it is the first second, third or xxxth HIV test
- gestation_week: contains the week of gestation for a patient at the time of the visit.
- age: age of the mother at enrollment. Note that, for example, when a mothers birthday is on the 22nd of August, we consider the age of the mother on the 21 of August still as 21.
- MALARIA_TREATMENT: the number of malaria treatment

ANC reports

serial_number	description	code	system	origin	original_description	extra filter
1	Number of Projected pregnant women	na	na	metric	na	given by clinic
2	First ANC visit	na	na	metric	na	VISIT_TYPE = 1 and VISIT_TYPE = ANC
2a	Week of pregnancy below 12 weeks (< 12 weeks)	na	na	metric	na	gestation_wee < 12 and VISIT_TYPE = ANC and VISIT_TYPE = 1

serial_number	description	code	system	origin	original	extra filter_description
2b	Week of pregnancy above 12 weeks (>= 12 weeks)	na	na	metric	na	gestation_week >= 12 and VISIT_TYPE = ANC and VISIT_TYPE_N = 1
2c	All revisit clients	na	na	metric	na	VISIT_NO > 1
2d	All pregnant women with 4+ ANC visits	na	na	metric	na	visit_type_no >= 4 and Visit_type = ANC
2e	Number of pregnant women with HB test at first ANC visit	23244001	snomed	procedure	Hemoglobin	VISIT_TYPE = ANC and VISIT_TYPE_N = 1
3	Pregnant women given TT2+ vaccine	146080110000	snomed	procedure	Tetanus toxoid	VISIT_TYPE = ANC
4a	4th and more pregnancy	na	na	questionnaire	How many times have you been pregnant before?	Answer Response: >4,4,5,5+

serial_number	description	code	system	origin	original_description	extra filter
4b	Below 20 YEARs	na	na	metric	na	age < 20
4c	First pregnancy at 35 and above YEARs	O09.6	ICD10	condition	primigravida	AGE >= 35
4c	First pregnancy at 35 and above YEARs	na	na	questionnaire	At what age did you have your first child?	answerResponse = >35
4c	First pregnancy at 35 and above YEARs	na	na	questionnaire	Were you older than 35 when you first got pregnant?	answerResponse = true
4d	Low HB < 8.5 g/dl - Anemia first visit	O99.0	ICD10	condition	Anaemia in pregnancy	VISIT_TYPE = ANC, VISIT_TYPE = 1
4e	High blood pressure (BP => 140/90 hg)	O13	ICD10	condition	Pregnancy - induced hypertension	VISIT_TYPE = ANC

serial_number	description	code	system	origin	original_description	extra filter
4f	Tuberculosis	O98.0	ICD10	condition	Tuberculosis	VISIT_TYPE = ANC
4g	Sugar in urine	69376001	snomed	procedure	Urine test for Glucose	VISIT_TYPE = ANC
4h	Protein in urine	441948005	snomed	procedure	Urine test for Protein	VISIT_TYPE = ANC
4i	Pregnant women tested for syphilis	169698000	snomed	procedure	Syphilis test / VDRL or PRP	VISIT_TYPE = ANC
4j	Diagnosed with syphilis infection	O98.1	ICD10	condition	Syphilis	VISIT_TYPE = ANC
4k	Treated for syphilis infection	75247008	snomed	procedure	Syphilis treatment	VISIT_TYPE = ANC
4l	Partners tested for syphilis infection	na	na	questionnaireResponse	Partners tested for syphilis infection	answer = true
4m	Partners diagnosed with syphilis infection	na	na	questionnaireResponse	Partners diagnosed with syphilis infection	answer = true

serial_number	description	code	system	origin	original_description	extra filter
4n	Partners / spouse treated for syphilis infections	na	na	questionnaire	Partners/spouse treated for syphilis infections	Response = true
4o	Pregnant women diagnosed with non syphilis Sexually Transmitted Diseases (STDs)	N77.1	ICD10	condition	bacterial vaginosis	VISIT_TYPE = ANC
4o	Pregnant women diagnosed with non syphilis Sexually Transmitted Diseases (STDs)	A56	ICD10	condition	chlamydia	VISIT_TYPE = ANC
4o	Pregnant women diagnosed with non syphilis Sexually Transmitted Diseases (STDs)	O54.10	ICD10	condition	genital herpes	VISIT_TYPE = ANC
4o	Pregnant women diagnosed with non syphilis Sexually Transmitted Diseases (STDs)	A54.9	ICD10	condition	gonorrheal	VISIT_TYPE = ANC
4o	Pregnant women diagnosed with non syphilis Sexually Transmitted Diseases (STDs)	N73.9	ICD10	condition	pelvic inflammatory disease	VISIT_TYPE = ANC
4o	Pregnant women diagnosed with non syphilis Sexually Transmitted Diseases (STDs)	B37	ICD10	condition	Candidiasis	VISIT_TYPE = ANC
4p	Pregnant women who received correct treatment for non syphilis STDs	?	snomed	procedure	Other STD treatment	VISIT_TYPE = ANC
4q	Partners / spouses diagnosed with non syphilis STDs	na	na	questionnaire	Partners/spouse diagnosed with non syphilis STDs	Response = true

serial_number	description	code	system	origin	original_description	extra filter
4r	Partners / spouses who receive correct treatment for non syphilis STDs	na	na	questionnaireResponse	Partners/spouses received correct treatment for non syphilis STDs	answer = true
5a	Pregnant mothers with known HIV infection before starting ANC clinic	na	na	questionnaireResponse	(Current Medical Condition)Do you have any of these health conditions?	answer = HIV
5b	Pregnant women who received councelling before HIV testing	313077009	snomed	procedure	HIV councelling (before test)	VISIT_TYPE = ANC
5c	Pregnant women receiving first HIV test at ANC clinic	171121004	snomed	procedure	HIV test	VISIT_TYPE = ANC and HIV_TEST = 1

serial_number	description	code	system	origin	original	extra filter_descriptio
5d	Pregnant women found to be HIV positive - first test	O98.7	ICD10	condition	HIV in pregnancy	VISIT_TYPE = ANC and HIV_TEST = 1
5e	Pregnant women below 25 YEARs found to be HIV positive - first test	O98.7	ICD10	condition	HIV in pregnancy	VISIT_TYPE = ANC and HIV_TEST =1 and age < 25
5f	Pregnant women councelled after HIV testing	313077009	snomed	procedure	HIV councelling (after test)	VISIT_TYPE = ANC
5g	Pregnant women tested for HIV with spouse/partner	171121004	snomed	procedure	HIV test	VISIT_TYPE = ANC
5h	Pregnant women receiving second HIV test	171121004	snomed	procedure	HIV test	VISIT_TYPE = ANC and HIV_TEST=
5i	Pregnant women found to be HIV positive by second HIV test	O98.7	ICD10	condition	HIV in pregnancy	VISIT_TYPE = ANC and HIV_TEST=

serial_number	description	code	system	origin	original_description	extra filter
5j	Partner/spouse receiving first HIV test at ANC clinic	na	na	questionnaire	Partner/spouse receiving first HIV test at ANC clinic	Response = true
5k	Partner/spouse found to be HIV positive by first test	na	na	questionnaire	Partner/spouse found to be HIV positive by first test	Response = true
5l	Partner/spouse receiving second HIV test at ANC	na	na	questionnaire	Partner/spouse receiving second HIV test at ANC	Response = true

serial_number	description	code	system	origin	original_description	extra_filter
5m	Partner/spouse found to be HIV positive by second test	na	na	questionnaire	Partner/spouse found to be HIV positive by second test	Response = true
5n	Pregnant women and partners/spouses with discordant HIV results after testing at ANC clinic	na	na	questionnaire	Partners/spouses with discordant HIV results after testing at ANC clinic	Response = true
5o	Pregnant women receiving infant feeding counselling	243094003		procedure	Breastfeeding education and assessment	VISIT_TYPE = ANC

serial_number	description	code	system	origin	original	extrafilter_description
5p	Pregnant women at below 25 YEARs receiving first HIV test	171121004	snomed	procedure	HIV test	VISIT_TYPE = ANC and HIV_TEST = 1 and age < 25
6a	Pregnant women receiving LLIN	?	snomed	procedure	Mosquito net	VISIT_TYPE = ANC
6b	Pregnant women testing Malaria using MRDT/BS	412690006	snomed	procedure	Malaria Rapid Test	VISIT_TYPE = ANC
6b	Pregnant women testing Malaria using MRDT/BS	?	snomed	procedure	%MRDT%	VISIT_TYPE = ANC
6c	Pregnant women diagnosed with malaria	B54	ICD10	condition	Malaria	VISIT_TYPE = ANC
6d	Pregnant mothers given IPT2	777371007	snomed	procedure	SP tabs for Presumptive malaria treatment	VISIT_TYPE = ANC and MALARIA_TRE = 2
6e	Pregnant mothers given IPT3	777371007	snomed	procedure	SP tabs for Presumptive malaria treatment	VISIT_TYPE = ANC and MALARIA_TRE = 3

serial_number	description	code	system	origin	original_description	extra filter
6f	Pregnant mothers given IPT4	777371007	snomed	procedure	SP tabs for Presumptive malaria treatment	VISIT_TYPE = ANC and MALARIA_T = 4
7	Pregnant mothers given Iron/Folic Acid (I, F, IFA) (enough to last until next visit)	63718003	snomed	procedure	Folic acid/Ferous	VISIT_TYPE = ANC
7	Pregnant mothers given Iron/Folic Acid (I, F, IFA) (enough to last until next visit)	?	snomed	procedure	Ferrous	VISIT_TYPE = ANC
7	Pregnant mothers given Iron/Folic Acid (I, F, IFA) (enough to last until next visit)	63718003	snomed	procedure	Folic Acid	VISIT_TYPE = ANC
7	Pregnant mothers given Iron/Folic Acid (I, F, IFA) (enough to last until next visit)	387402000	snomed	procedure	Ferrous Sulphate	VISIT_TYPE = ANC
8	Pregnant women given antihemitics (Albendazole /Mebendazole)	387311004	snomed	procedure	Mebendazole	VISIT_TYPE = ANC
9	Pregnant Women counselled for family planning	397619005	snomed	procedure	Family planning	VISIT_TYPE = ANC
10	Pregnant Women referred	na	na	service	na	Report total nr of patients in referral table

serial_number	description	code	system	origin	original	extra filtered_description
11	Pregnant Women referred to CTC	na	na	serviceReport	na	total nr of patients in referral table where reason is HIV

delivery topics

serial_number	description	code	system	origin	original	extra filtered_description
1	Mothers expected /projected to deliver at service station	na	na	Observation	na	use expected delivery date and count nr of unique patients

serial_number	description	code	system	origin	original_description	extra filter
2a	Mothers delivered at the facility	11466000	SNOMED	procedure	Caesarean Section	VISIT_TYPE = delivery
2a	Mothers delivered at the facility	177184002	SNOMED	procedure	%Delivery%	VISIT_TYPE = delivery
2a	Mothers delivered at the facility	237311001	SNOMED	procedure	Breech delivery	VISIT_TYPE = delivery
2b	Deliveries/Birth before arrival at the facility (BBA)	na	na	questionnaire	%where was the baby born%	answerResponse like %other%
2c	Delivered by assistance of traditional birth attendants(TBA)	na	na	questionnaire	%was there a birth attendant present?%	answerResponse=true
2d	Delivered at home without assistance of TBA	na	na	questionnaire	%was there a birth attendant present?%	answerResponse=false
3a	Delivered after 12 hours since onset of labour	O63	ICD10	condition	Long labour	VISIT_TYPE = delivery

serial_number	description	code	system	origin	original_description	extra filter
4b	Vacuum (MV)	?	SNOMED	procedure	Vacuum (VM)	VISIT_TYPE = delivery
4c	Breech delivery (BR)	237311001	SNOMED	procedure	Breech delivery	VISIT_TYPE = delivery
4d	Caesarian section (CS)	11466000	SNOMED	procedure	Cesarean Section	VISIT_TYPE = delivery
5a	APH	O46	ICD10	condition	Antepartum Hemorrhage	VISIT_TYPE = ANC
5c	High Blood Pressure	O13	ICD10	condition	Pregnancy - induced hypertension	VISIT_TYPE = ANC
5d	Pre-eclampsia	O14	ICD10	condition	Pre - eclampsia	VISIT_TYPE = ANC
5e	Eclampsia	O15.0	ICD10	condition	Eclampsia	VISIT_TYPE = ANC
5f	Malaria	B54	ICD10	condition	Malaria	VISIT_TYPE = ANC
5g	Anaemia	O99.0	ICD10	condition	Anaemia in pregnancy	VISIT_TYPE = ANC

serial_number	description	code	system	origin	original_description	extra filter
6a	PPH	O72	ICD10	condition	Postpartum haemorrhage	VISIT_TYPE = PNC or delivery
6d	Obstructed labour	O64	ICD10	condition	Fetal Malpresentation	VISIT_TYPE = delivery
6d	Obstructed labour	O65	ICD10	condition	Pelvic abnormality	VISIT_TYPE = delivery
6d	Obstructed labour	O66	ICD10	condition	Other obstructed labour	VISIT_TYPE = delivery
6e	Retained placenta	O73	ICD10	condition	Retained placenta and membranes, without haemorrhage	VISIT_TYPE = delivery
6h	Sepsis	O85	ICD10	condition	Puerperal Sepsis	VISIT_TYPE = PNC

serial_number	description	code	system	origin	original_description	extra filter
7a	Received antibiotic	255631004	SNOMED	procedure	Antibiotics	VISIT_TYPE = delivery
7b	Given uterotonic	?	SNOMED	procedure	Uterotonics	VISIT_TYPE = delivery
7c	Given magnesium sulphate	387202002	SNOMED	procedure	Magnesium Sulphate	VISIT_TYPE = delievery
7d	Placenta removed manually	?	SNOMED	procedure	Manual removal of placenta	VISIT_TYPE = delivery
7f	Received blood transfusion	116859006	SNOMED	procedure	Blood transfusion	VISIT_TYPE = delivery
8c	Women received oxytocin after delivery	777037006	SNOMED	procedure	Oxytocin	VISIT_TYPE = delivery
8d	Women received egometrine after delivery	126071000	SNOMED	procedure	Ergometrine	VISIT_TYPE = delivery
8e	Women received Misoprostol after delivery	776808001	SNOMED	procedure	Misoprostol	VISIT_TYPE = delivery

serial_number	description	code	system	origin	original_description	extra filter
9a	Total tested HIV at ANC	171121004	SNOMED	procedure	HIV test	VISIT_TYPE = ANC
9b	Tested positive at ANC	O98.7	ICD10	condition	HIV in pregnancy	VISIT_TYPE = ANC
9c	Total tested HIV during and after delivery	171121004	SNOMED	procedure	HIV test	VISIT_TYPE = delivery or PNC
9d	Tested positive during and after delivery	O98.7	ICD10	condition	HIV in pregnancy	VISIT_TYPE = delivery or PNC
9e	Opted for exclusive breast feeding (EBF)	na	na	questionnaire	%MamaCare Je una mpango wa kunyonyesha mtoto kwa%	Response answers like %Mfululizo kwa miezi 6 bila chakula kingine%

serial_number	description	code	system	origin	original_description	extra filter
9f	Those Opted for alternative feeding(RF)	na	na	questionnaireResponse	%Mama Je una mpango wa kunyonyesha mtoto kwa%	answers like %Kumpa maziwa ya mama na chakula kingine%
9f	Those Opted for alternative feeding(RF)	na	na	questionnaireResponse	%Mama Je una mpango wa kunyonyesha mtoto kwa%	answers like %Hutanyonyesha kabisa%
9g	Given ARV prophylaxis (Tail) at time of discharge	713548004	SNOMED	procedure	ARV prophylaxis	VISIT_TYPE = delivery
9h	Mothers given referral to CTC clinic	na	na	serviceRequest	na	patients in serviceRequests where reason = HIV
10a	Total number of children born alive					

serial_number	description	code	system	origin	original_description	extra filter
10b	Born alive with weight less than 2.5 kg	P05.0	ICD10	condition	Low Birth weight (< 2499 grams)	VISIT_TYPE = delivery
10c	Children born alive with weight =>2.5 kg	P05.0	ICD10	condition	Low Birth weight (< 2499 grams)	VISIT_TYPE = delivery and make sure condition is NOT met
10e	Baby Born fresh dead (FSB)	P95	ICD10	condition	Stillbirth (FSB)	VISIT_TYPE = delivery
10f	Baby Born with HIV-positive mothers	O98.7	ICD10	condition	HIV in pregnancy	VISIT_TYPE = delivery
10g	Baby received ARV drugs	713548004	SNOMED	procedure	ARV prophylaxis	VISIT_TYPE = delivery

serial_number	description	code	system	origin	original_description	extra filter
12a	Number of babies received breathing assistance by suction	232705007	SNOMED	procedure	Baby breathing assistance - sunction	VISIT_TYPE = delivery
12b	Number of babies received breathing assistance by stimulation	1230275006	SNOMED	procedure	Baby breathing assistance - stimulation	VISIT_TYPE = delivery
12c	Number of baby received breathing assistance by by Bag and Mask	?	SNOMED	procedure	Baby breathing assistance - bag and mask	VISIT_TYPE = delivery
13a	Number of babies who were breastfeed within one hour after birth	243094003	SNOMED	procedure	Breastfeeding education and assessment	VISIT_TYPE = delivery

serial_number	description	code	system	origin	original_description	extra filter
13b	Number of mothers who were given referral	na	na	serviceRequest	na	total nr of patients in referral table
14a	Maternal death occurred	O95	ICD10	condition	Maternal death	VISIT_TYPE = delivery
15a	Given counselling of family planning	397619005	SNOMED	procedure	Family planning	VISIT_TYPE = PNC
15b	Client who does total hysterectomy	236886002	SNOMED	procedure	Hysterectomy	na
15c	Client for whom IUCD was inserted	176837007	SNOMED	procedure	IUCD	na
15d	Client with inserted Norplant	169553002	SNOMED	procedure	%Norplant%	na
15e	Client received POP	255664004	SNOMED	procedure	Depo	na
16a	New baby born died	P96.8	ICD10	condition	Neonatal death	na

pnc

serial_code	Decription	Code	System	origin	original_description	extra filtering
1a	Toatal number of clients attendees within 48hs					
1b	Total number of clients attendees 3-7 days	na	na	combined metric	na	3<= DAYS_SINCE_DEL <= 7
1c	Total number of atendees within 7 days (1a+1b)	na	na	combined metric	na	DAYS_SINCE_DEL

serial_code	Decription	Code	System	origin	original_description	extra filtering
2	Total number of clients who finished all attendee (48hrs, 3-7days, 8-28days, 29-42)	na	na	combined metric	na	DAYS_SINCE_DELIV <= 42
3	Clients with severe anaemia (HB<8.5g/dl)	O99.0	ICD10	condition	Anaemia in pregnancy	VISIT_TYPE = PNC
4	Clients who had mental disorder after delivery	F53.9	ICD10	condition	Puerperal Mental Disorder, Unspecified	VISIT_TYPE = PNC
5	Client given vitamin A	36922[illegible]	SNOMED	procedure	Vitamin A	VISIT_TYPE = PNC
6	Clients with infected perineal tear	O90.1	ICD10	condition	Perineal tear infections	VISIT_TYPE = PNC
7	Client with Fistula	N82	ICD10	condition	?	VISIT_TYPE = PNC
8	Delivered out of health facility	na	na	combined metric	na	check patients whose expected delivery date is passed by 14 days and who did not have a delivery visit
8a	Delivered before arrival to health facility (BBA)	na	na	questionnaireResponse	%where was the baby born%	answer = other
8b	Delivered at traditional birth attendant (TBA)	na	na	questionnaireResponse	%was there a birth attendant present?%	true
8c	Home deliveries	na	na	questionnaireResponse	%where was the baby born%	home

serial_code	Decription	Code	System	origin	original_description	extra filtering
9a	Counselled on family planning	397619005	SNOMED	procedure	Family planning	VISIT_TYPE = PNC
9b	Given Condom	339731000000107	SNOMED	procedure	Condom	VISIT_TYPE = PNC
9c	Given Pills (POP)	255664004	SNOMED	procedure	POP	VISIT_TYPE = PNC
9d1	Given norplants (Implanon)	169553002	SNOMED	procedure	Norplant	VISIT_TYPE = PNC
9d2	Given norplants (Jadele)	169553002	SNOMED	procedure	Norplant (Jadele)	VISIT_TYPE = PNC
9e	Given inter uterine device	176837007	SNOMED	procedure	IUCD	VISIT_TYPE = PNC
9f	Sterilization (BTL)	287664005	SNOMED	procedure	Sterilization (BTL)	VISIT_TYPE = PNC
10	PMTCT	?	SNOMED	procedure	PMTCT	na
10b	HIV tested during postnatal (within 42days since delivery)	171121004	SNOMED	procedure	HIV test	VISIT_TYPE = PNC
10c	Diagnosed with HIV during post natal (within 42 days since delivery)	O98.7	ICD10	condition	HIV in pregnancy	VISIT_TYPE = PNC
10d	Clients with HIV who opted for exclusive breast feeding (EBF)	na	na	questionnaireResponse	%MimbaCase: Je una mpango wa kunyonyesha mtoto kwa%	answer ilike %Mfululizo kwa miezi 6 bila chakula kingine%
10e	Clients with HIV who opted for replacement feeding (RF)	na	na	questionnaireResponse	%MimbaCase: Je una mpango wa kunyonyesha mtoto kwa%	answer ilike %Kumpa maziwa ya mama na chakula kingine%

serial_code	Decription	Code	System	origin	original_description	extra filtering
10e	Clients with HIV who opted for replacement feeding (RF)	na	na	questionnaireResponse	%MamaCase: Je una mpango wa kunyonye-sha mtoto kwa%	%Hutanyonyesha kabisa%

6.2. Updating DHIS2 Topics

It is possible that the method to calculate certain DHIS2 topics must be updated. To do so there are multiple options:

- **Patient timeline mapping:** Alter or add to the mapping tables that are used to map the patient_timeline information. The mapping tables are stored in the master data folder on azure. When creating a mapping table with a new name, this must be updated in the configuration of the DHIS2 reporter, which can be found under *src/data/DHIS2/config.py*.
- **Questionnaire mapping:**Alter or add to the mapping tables that are used to map the questionnaire_response information. These mapping tables can be found under in the master data folder on Azure as well. When creating a mapping table with a new name, this must be updated in the configuration of the DHIS2 reporter, which can be found under *src/data/DHIS2/config.py*.
- **Separate calulcations:** If it is only one topic that must be updated, it can be more convenient to add to the mapping queries for the DHIS2 reports. For each report there is a separate mapping one can add to: *add_anc_mapping*, *add_delivery_mapping*, and *add_pnc_mapping*. These queries can be found under *scr/data/DHIS2/SQL_queries*.

6.3. DHIS2 TODO

Still not all DHIS2 topics have been (correctly) implemented in the reports:

- **Twin related topics in delivery report**: There are topics about twins that were not available in the current reports yet, but calculations are available in previous mapping code, that have yet to be converted in the DHIS2 reporter class
- **ANC report topic: Pregnant women tested for HIV with spouse/partner**. Currently it is checked if visit_type = ANC and if tested for HIV. but it is unknown if this is 'with spouse/parter'.
- **ANC report topic: Pregnant women testing Malaria using MRDT/BS.** Now it is only checked if a malaria rapid test was performed, but this should also include procedures with subject MRDT. MRDT is not yet mapped to fhir, because type = null and we only mapped visit items where the type is drug or test.

- **ANC report topic: Pregnant Women counselled for family planning.** This item is always classified as PNC, so either it does not occur in ANC or information is missed.
- **Delivery report topic: VACUUM (MV).** The mapping for this topic is already implemented, but this visit item is not yet mapped to fhir, because type is equipment, and not drug or test.
- **delivery report topic: Placenta removed manually.** The mapping for this topic is already prepared, but item is not yet mapped to fhir. because its type is null (not test or drug)
- **delivery and anc report referral related topics:** Now all referral patients are used for calculations, where previously only ‘mother’ type patients were evaluated. This information is not mapped to fhir, are non-mother patients filtered out?

Besides the specific DHIS2 topics there are also a few other todo’s:

- **Evaluation:** There are big differences with the previous DHIS2 reports that were based on COT data directly. Some differences were intended, some are not yet explained. This should be further investigated.
- **Procedure mapping:** The ICD10 codes are used to map the DHIS2 topics to the conditions, but for the procedures the SNOMED codes are not used yet. This is because we could not find SNOMED codes for all procedures, and for some of the procedures, we were uncertain whether the SNOMED codes were correct. Therefore, still the list of SNOMED codes should be checked and filled. When this list of SNOMED codes is completed, the mapping should be based on the SNOMED codes (currently its matched on the text).

6.4. USING DHIS2reporter

TODO: Implement example notebooks on analytics workbench! Example notebooks, showing how to run the Baser reporter and the DHIS2 reporter are available in the analytics workbench. These reporters can be run to manually create the standardized base reports or DHIS2 reports. In addition, one can also directly load the base reports (patient_timeline, questionnaire_responses) or DHIS2 reports, and directly start analyzing them.

7. Value Points

7.1. Introduction

In 2023, PharmAccess introduced ValuePoints (VPs) in Hanang province, a rural district with low digitalization rates, to visualize completeness of maternal care delivery and incentivize value. Integrated into the local digital ecosystem, VPs provide real-time insights at the patient journey level, enabling providers to deliver higher value, patient-centric care.

7.2. ValuePoint criteria

7.2.1. Service Indicators

Value Points are assigned to key service indicators across the pregnancy journey.

These indicators include:

Early ANC: An antenatal care (ANC) visit registered within the first 12 weeks of pregnancy (ANC $<$ 12 weeks)

ANC profile: Full antenatal care profile administered.

To count the mothers who had a full ANC profile, we check for visits that include the following procedures:

- vdrl
- hemoglobin
- blood grouping
- HIV test
- blood pressure
- urine test

In addition, the ANC profile should take place during the first visit.

Ultrasound before end of second trimester: We check for visits with an ultrasound procedure that occurred within the first 24 weeks of pregnancy. (Ultrasound < 24 weeks)

4+ ANCs: Four or more registered antenatal care visits. Clinics receive 0.5 points per visit for the first 4 visits. For further ANC visits no more points are received.

Facility Based Delivery: We filter mothers that have had a delivery visit at the facilities.

2PNC: Two or more registed postnatal care (PNC) visits. Clinics receive 0.5 points per visit for the first 2 visits. For further PNC visits no more points can be obtained.

7.2.2. Obtained Points

There is a limited time in which the different value points can be obtained. The cut-off date and the total number of value points that can be obtained for each topic are displayed in the graph below. The week of pregnancy (gestational age) for a mother is calculated based on her Expected Delivery Date (EDD).

Criterium	Points	Cut-off-date
ANC profile	1	12 weeks
Early ANC	1	24 weeks
Ultrasound before end of second trimester	1	24 weeks
4ANC	2	42 weeks
Facility Based Delivery	4	42 weeks
2PNCs	1	44 weeks

The date assigned to an obtained ValuePoint is the date that a procedure was performed.

7.2.3. Missed Value Points

If a mother does not obtain a value point before the cut-off date, we consider this value point as missed. Missed value points are coupled to a missed date. The missed date is calculated based on the EDD and the cut-off weeks.

A mother can go to different clinics during her journey. At this point, we only calculate missed value points for the clinic where the mother had her first visit.

7.2.4. Attainable Value Points

Attainable points are only calculated for the current date and defined for any topic that has not yet been obtained, but where the current date is less than the missed date.

Attainable points are only given to the clinic where the mother had her first visit.

7.2.5. Shared Value POints

As mentioned above a mother can visit different clinics. For this reason, the concept of shared value points was introduced.
Value points can be shared for:

- Ultrasound

- 4 ANC
- Facility Based Delivery
- 2 PNC

For Ultrasound and Facility Based DElivery the value points are always shared between the clinic where the procedure took place (procedure clinic) and the clinic where the first visit took place (enrollment clinic). The value points are shared as follows:

Criterium	Points Procedure clinic	Points enrollment clinic
Ultrasound	0.75	0.25
Facility Based Delivery	3	1

The date for these shared value points is defined as the date that the procedure was performed.

For the 4ANC and 2PNC criteria, a clinic only obtains value points if the mother had equal or more than 4 ANC visits or equal or more than 2 PNC visits in total. For each visit up to the 4th (ANC) or 2nd (PNC) visit a clinic receives 0.5 Value Points. For further visits, no value points are obtained.

7.2.6. Mothers per trimester

The mothers per trimester table shows the number of mothers that have visited the clinic and have an active journey (within EDD + four weeks) per month. Mothers are shown that have visited the clinic at least once during their journey. AS a mother can fall in two trimesters within one month, we have decided to show the status at the 1st of the month.

The mothers are divided in the different trimesters as follows:

trimester	start	end
first	week 1	week 12
second	week 13	week 26
third	week 27	week 42 or date of delivery
past EDD	week 42 or date of delivery	week 46

7.2.7. Value Points Reporter

Like the DHIS2 reporter the value points reporter uses the base tables created by the base reporter (see figure 7.1)

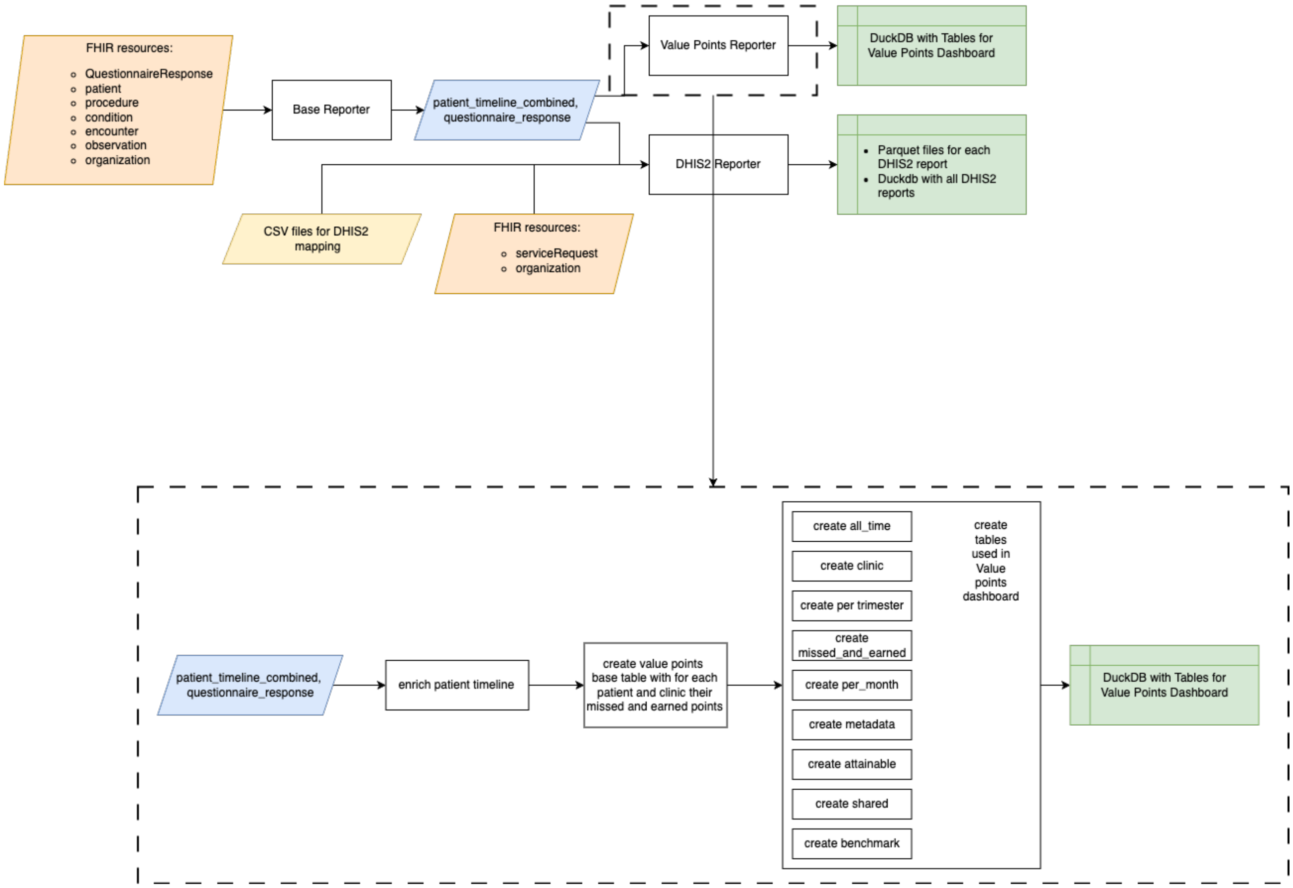


Figure 7.1.: Value Points reporting flow and required input, blue: intermediate data, green: processed data

After the base tables are created, the value points reporter has a function to enrich the base tables with elements that help with value point calculations.

Then, the value points base report is created, which contains for each patient and clinic, the value points that have been obtaned. This table is called the value_points_base table.

All the tables that are used in the value points dashboard are then created based on the value_poiints_base tables.

Finally, The results are saved in a separate DuckDB file and uploaded to Azure storage bucket.

NOTE we agreed to only show data in the valuepoints dashboard that were created from 01-03-2023 onwards.

7.2.8. IN PROGRESS

TODO 2PNC should be checked if the PNC visits occured in first twee weeks after delivery (if delivered at facility) or if not delivered at facility if PNC visit is within 4 weeks of EDD. Currently just taken one cutof date..

Note on difference in pregnancy weeks in speaking and coding terms When stating week 12 as cut-off date, it is meant to be up to and including the 12th week of pregnancy, while In code this will be anything below 12: To clarify: a mother is pregnant for 86 days: $86/7 = 12.2$ weeks. We would say the mother is in her 13th week of pregnancy, and thus the mother has passed this cut-off date. In code, we check whether the week is $<=12$ which is false and thus we the cut-off date was passed.

TODO

Currently the Value Points Code contains a function to export the valuepoints tables as a separate duckdb. The way this is done is a bit cumbersome: first we copy the duckdb file that contains all data (also the raw data). Then we remove all schemas from there. Therafter we export the database, and import it again in a new duckdb file.

We have to export the database without schemas because it became very big (around 700MB) and removing schemas and tables did not bring the size down. The only method to bring the size down is currently to export the tables needed and import them again, then the size is around 1 MB.

In the future, there is a function *Vacuum*, which now does not seem to work, but is expected to remove all unnecessary space in a duckdb file?

7.3. Use of value points reporter

TODO Create a notebook as an example!

8. Offline capabale webapp

8.1. Why do we need this

8.2. Describe solution

8.3. Describe implementation

- Healthcare workers have access to Android tablets
- Internet access is intermittent at point of care. Assume that healthcare worker has access to internet when she/he takes tablet home
- Presenting results from data analytics needs to have solution for synchronization/offline usage
- We have chosen for a combination of svelte + duckdb for building such offline capable Android apps. Details are explained in Appendix A.
- This impacts the following elements in the reference architecture
  - Data visualization and user interaciton
  - Development, Engineering and DevOps: capability to deploy Svelte apps on and hosted app platform
  - Standards: compatibility with D3.js ecosystem for building interactive data visualizations. Vega-lite is standard is used in this component, and is also available in the workbench

9. Analytics on FHIR

9.1. Current standards for analytics

- Explain Jupyter notebook for interactive computing
- We want to support different languages
- Federated analytics
  - Most simple way: we mount the different data containers, which currently are ADLS Gen2 containers
  - In future: federated analytics libraries
- Option: computation in the blind

Interactive notebooks are the de-facto standard for analysts and researchers. Many open source solutions which can be self-hosted, most importantly Jupyter Project Many commercial offerings: All cloud vendors have own version of Jupyter already integrated in their environment To save engineering work, for now we opt for Azure ML Studio

The paradigm shift that we introduce is federated analytics Federated analytics = queries over decentralised databases Federated learning = training algorithms over decentralised data We use federated analytics as a term to cover both. For LMICs, starting with federated analytics (queries) is positioned on short term We add multi-party computation as optional component that can be used through interactive notebook environment

Impacts following components: Data Visualization and User Interaction Data Analytics Data processing architectures Standards Considerations for LRS

10. FHIR tables Kenya

As shown in the demonstrators outlining the FHIR adaptor for MomCare Tanzania and FHIR-based reporting, standardized tables based upon FHIR resources provide a strong fundament for futher analysis. These demonstrators also indicate that the largest challenge for the transformation of source data to FHIR Resources is the structure of the source data; especially when it lacks unique identifiers. The data structure of the source data in the MomCare Kenya setting is quite different from the MomCare Tanzania setting. Whereas in Tanzania the full data model is managed by PharmAccess, the Kenyan data models mainly derive their data from CarePay (an external partner that provides the mobile healthcare billing services).

The currently used tech stack that transforms the CarePay source data for analysis purposes is called CareAnalytics. Over time, many changes to the models within CareAnalytics have been made, but very few of them are clearly documented. This makes it complex to derive how tables are created (what source data are they based on? What transformations have happened? What models have been used for these transformations).

Therefore, we decided to experiment with the conversion of the source data to standardized FHIR tables without going through the process of creating nested FHIR resources. Hypothetically, this should allow the functionalities of the other demonstrators to be used while bypassing the complexities of the CareAnalytics tech stack.

10.1. Source data

We decided to try to use data as close to the source as possible. This was on two main reasons: * CarePay is in the process of adapting their data models. There is no point to building a transformation upon a changing base layer. However, the final data models are not yet close to being finished, and we would like to use the demonstrators in the short-term. * The use of source data allows this demonstrator to be applied to other projects where no (or a limited) tech stack is currently used. * It allows us to have full control of the transparency in documentation; an essential aspect for this projects reusability and interoperability.

With source data, we refer to the (daily) data dumps provided by CarePay as duckdb files. These files were converted to .parquet files, and can be accessed under the data > bronze folders in the Jupyter Notebooks on the Analytics Workbench.

10.2. Analytics Workbench

The Analytics Workbench is an open access platform where data can be accessed, managed, and transformed.

10.3. Standardized tables

The tables used are the same as in the Tanzanian demonstrator, a *patient_timeline* and *patient_information* table. (insert link to https://health-data-commons.pharmaccess.org/docs/reporting.html#patie timeline-table)

The *patient_timeline* table consists of all procedures and diagnosis that a patient is subjected to over time. It is created based on the following fhir resources:

- patient
- procedure
- condition
- encounter
- observation
- organization

The **patient_information** table contains basic information specific to the patient. For further analysis those two tables are often combined into one **patient_timline_combined** table.

The use of the same standardized tables as base reports, should allow the creation of the reports of interests using no to very little adaptation of the code. Reports that have previously been created include:

- **DHIS2 reports** provide insight in how many mothers visit a clinic, what conditions occur amongst mothers that visit that clinic, and what procedures mothers obtain in a clinic.
- **Value point reports** provide the value points (described in the *value points* section) obtained by clinics by threathing their patients.

10.4. Transformation of source data to FHIR tables

(Insert figures CP data structure overlay)

Stated vs Derived ? (TO DO)

10.5. Stated

10.5.1. Patient Timeline Table

The patient timeline is created by joining the FHIR resources for patient, encounter, claim, procedure and condition.

10.5.2. One-to-one mapping

Source table: **beneficiary** mapped to **patient**

Variable	FHIR resource
ID	patient_id
GENDER	gender
DATE_OF_BIRTH	birth_date
EXPECTED_DELIVERY_DATE	expected_delivery_date

Source table: **treatments** mapped to **encounter**

Variable	FHIR resource
ID	encounter_id
BENEFICIARY_REF_ID	patient_id
PROVIDER_REF_ID	visit_provider_id
DATE_TREATMENT	procedure_datetime

Source table: **claims** mapped to **claim**

Variable	FHIR resource
ID	claim_id
TREATMENT_ID	encounter_id
CREATED_DATE	created
MODIFIED_DATE	modified
STATUS	claim_status

Source table: **treatment_diagnosis** mapped to **condition**

Variable	FHIR resource
TREATMENT_ID	encounter_id
ICD10_CODE	icd10_diagnosis_code

Variable	FHIR resource
DIAGNOSIS_DESCRIPTION	diagnosis_description

10.6. Filtering patients

The patients included in the tables should only included mothers enrolled in the MomCare program. This means filtering the CarePay billing data based on;

gender –> only females age –> exclude children program_id –> include only specific numbers

> !! Where do we want to add the program_id as a resource?

10.7. Derived (Level: High Certainty)

10.7.1. Merging with item group label

An additional label 'item group' was added based on the **item_code** variable. This label was developed in the CareAnalytics data model, and was added here to facilitate the transformation of items to FHIR resources such as medication, procedures and other services.

Source table: **items**

Variable	FHIR resource
INVOICE_ID	item_invoice_id
CLAIM_ID	item_claim_id
QUANTITY	quantity
ITEM_NAME	modified
ITEM_CODE	claim_status
item_group	item_group

10.8. Timing in the journey

Dates need to be converted to date format.

Enrolment The moment of enrolment (**enrolment_date**) is defined as the moment when the patient has both an approved claim status and a registered visit to a health provider.

> Derived? Stage 2? More business rules than FHIR logic

Age at enrolment Age of enrolment is determined by calculating the difference between the **enrolment_date** and **birth_date**. Age groups are defined as the above age of enrolment being: Between 10-14; *age_group_10_14*, Between 15-19; *age_group_15_19*, Between 20-24; *age_group_20_24*, Between 25-29; *age_group_25_29*, Between 30-34; *age_group_30_34*, 35 or older; *age_group_35_plus*

Gestation week The gestation week is calculated as the difference in weeks between the **expected_delivery_date** and the **procedure_date**

> !! We should distinguish between the gestation week linked to the encounter (as described above) and the gestation week linked to the patient (this is the timing of where they are in the journey, so instead of *procedure_date* using the current date).

Delivery The delivery date is defined as the moment when a visit that has been categorized as a 'delivery' **encounter_type** has taken place.

10.9. Encounter types

Encounter types are defined by the type of items billed during that encounter. The type of item is derived from the **item_group**. ANC visit when an item in the 'anc visit' or 'ultrasound' **item_group** is billed. PNC visit when an item in the 'pnc visit' **item_group** is billed.

A delivery visit is defined as either

- The encounter when an item related to a delivery **item_group** is billed, including the following item groups: Delivery / Normal delivery / Single spontaneous delivery / Caesarean delivery / Cesarean delivery / Perineal laceration during delivery / Other assisted single delivery/ normal delivery - uncomplicated / Delivery complicated by fetal stress / Other complicated delivery / Multiple delivery / Single delivery by forceps and vacuum extractor / Delivery complicated by umbilical cord complications / Complications of anaesthesia during delivery / Outcome of delivery, unspecified
- The encounter when a diagnosis related to a delivery was given, including the following conditions: All O80 ICD10 codes, meaning conditions starting with O8.

10.10. Procedure types

> Possibly include the transformation of items based on CP codes to those FHIR resources Where: DO - medication > Medication PO - procedure > Procedure LO - lab > DiagnosticReport?

11. MomCare FAIR Data

11.1. FHIR-compliant MNCH dataset

Landing page for the MomCare (meta)data.

Bibliography

Ayaz, Muhammad, Muhammad F Pasha, Mohammed Y Alzahrani, Rahmat Budiarto, and Deris Stiawan. 2021. "The Fast Health Interoperability Resources (FHIR) Standard: Systematic Literature Review of Implementations, Applications, Challenges and Opportunities." *JMIR Medical Informatics* 9 (7): e21929.

Beck, Micah. 2019. "On the Hourglass Model." *Communications of the ACM* 62 (7): 48–57.

Duda, Stephany N, Nan Kennedy, Douglas Conway, Alex C Cheng, Viet Nguyen, Teresa Zayas-Cabán, and Paul A Harris. 2022. "HL7 FHIR-based Tools and Initiatives to Support Clinical Research: A Scoping Review." *Journal of the American Medical Informatics Association* 29 (9): 1642–53.

Gebreslassie, Tesfit Gebremeskel, Mirjam van Reisen, Samson Yohannes Amare, Getu Tadele Taye, and Ruduan Plug. 2023. "FHIR4FAIR: Leveraging FHIR in Health Data FAIRfication Process: In the Case of VODAN-A." *FAIR Connect* 1 (1): 49–54.

Huisman, Liesbeth, Shannen Mc van Duijn, Nuno Silva, Rianne van Doeveren, Jacinta Michuki, Moses Kuria, David Otieno Okeyo, et al. 2022. "A Digital Mobile Health Platform Increasing Efficiency and Transparency Towards Universal Health Coverage in Low- and Middle-Income Countries." *Digital Health* 8: 20552076221092213.

Jones, James, Daniel Gottlieb, Joshua C Mandel, Vladimir Ignatov, Alyssa Ellis, Wayne Kubick, and Kenneth D Mandl. 2021. "A Landscape Survey of Planned SMART/HL7 Bulk FHIR Data Access API Implementations and Tools." *Journal of the American Medical Informatics Association* 28 (6): 1284–87.

Kickbusch, Ilona, Dario Piselli, Anurag Agrawal, Ran Balicer, Olivia Banner, Michael Adelhardt, Emanuele Capobianco, et al. 2021. "The Lancet and Financial Times Commission on Governing Health Futures 2030: Growing up in a Digital World." *The Lancet* 398 (10312): 1727–76.

Liu, Dianbo, Ricky Sahu, Vlad Ignatov, Dan Gottlieb, and Kenneth D. Mandl. 2020. "High Performance Computing on Flat FHIR Files Created with the New SMART/HL7 Bulk Data Access Standard." *AMIA Annual Symposium Proceedings* 2019 (March): 592–96.

Mamuye, Adane L., Tesfahun M. Yilma, Ahmad Abdulwahab, Sean Broomhead, Phumzule Zondo, Mercy Kyeng, Justin Maeda, Mohammed Abdulaziz, Tadesse Wuhib, and Binyam C. Tilahun. 2022. "Health Information Exchange Policy and Standards for Digital Health Systems in Africa: A Systematic Review." *PLOS Digital Health* 1 (10): e0000118.

Mandel, Joshua C, David A Kreda, Kenneth D Mandl, Isaac S Kohane, and Rachel B Ramoni. 2016. "SMART on FHIR: A Standards-Based, Interoperable Apps Platform for Electronic Health Records." *Journal of the American Medical Informatics Association* 23 (5): 899–908.

Mandl, Kenneth D., Daniel Gottlieb, Joshua C. Mandel, Vladimir Ignatov, Raheel Sayeed, Grahame Grieve, James Jones, Alyssa Ellis, and Adam Culbertson. 2020. "Push Button Population Health: The SMART/HL7 FHIR Bulk Data Access Application Programming Interface." *Npj Digital Medicine* 3 (1): 1–9.

McCool, Judith, Rosie Dobson, Robyn Whittaker, and Chris Paton. 2022. "Mobile Health (mHealth) in Low- and Middle-Income Countries." *Annual Review of Public Health* 43 (1): 525–39.

Mrema, Anunsiatha. 2021. "Application of Digital Platform to Enhance Quality Improvement in Momcare Facilities in Manyara." *AIJR Abstracts*, July.

Neumark, Tom, and Ruth J. Prince. 2021. "Digital Health in East Africa: Innovation, Experimentation and the Market." *Global Policy* 12 (S6): 65–74.

Pedrera-Jiménez, Miguel, Spanish Expert Group on EHR standards, Dipak Kalra, Thomas Beale, Adolfo Muñoz-Carrero, and Pablo Serrano-Balazote. 2022. "Can OpenEHR, ISO 13606 and HL7 FHIR Work Together? An Agnostic Perspective for the Selection and Application of EHR Standards from Spain." TechRxiv.

Rieke, Nicola, Jonny Hancox, Wenqi Li, Fausto Milletarì, Holger R. Roth, Shadi Albarqouni, Spyridon Bakas, et al. 2020. "The Future of Digital Health with Federated Learning." *Npj Digital Medicine* 3 (1): 1–7.

Sanctis, Teresa De, Mary-Ann Etiebet, Wendy Janssens, Mark H. van der Graaf, Colette van Montfort, Emma Waiyaiya, and Nicole Spieker. 2022. "Maintaining Continuity of Care for Expectant Mothers in Kenya During the COVID-19 Pandemic: A Study of MomCare." *Global Health: Science and Practice* 10 (4).

Schultes, Erik. 2023. "The FAIR Hourglass: A Framework for FAIR Implementation." Edited by Barbara Magagna. *FAIR Connect* 1 (1): 13–17.

Shija, Liberatha, Johnson Yokoyana, Anunsiatha Mrema, Jonia Bwakea, Theodora Kiwale, and Neema Massawe. 2021. "Access to Essential Maternal Health Commodities Key to Improving Quality and Adherence to Maternal Healthcare Regimes: Experience from a MomCare Project in Northern Tanzania." *AIJR Abstracts*, July.

Smirnov, Pavel. 2019. "About FHIR Facades (Part I) - Two Approaches." https://www.health-samurai.io/articles/about-fhir-facades-part-i-two-approaches.

"Tanzania Health Enterprise Architecture." 2020.

Tarkowski, Alex, and Jan J. Zygmuntowski. 2022. "Data Commons Primer."

van Reisen, Mirjam, Francisca Oladipo, Mia Stokmans, Mouhamed Mpezamihgo, Sakinat Folorunso, Erik Schultes, Mariam Basajja, et al. 2021. "Design of a FAIR Digital Data Health Infrastructure in Africa for COVID-19 Reporting and Research." *Advanced Genetics (Hoboken, N.j.)* 2 (2): e10050.

Vorisek, Carina Nina, Moritz Lehne, Sophie Anne Ines Klopfenstein, Paula Josephine Mayer, Alexander Bartschke, Thomas Haese, and Sylvia Thun. 2022. "Fast Healthcare Interoperability Resources (FHIR) for Interoperability in Health Research: Systematic Review." *JMIR Medical Informatics* 10 (7): e35724.

World Health Organization. 2021. "Global Strategy on Digital Health 2020-2025." Geneva.

World Health Organization, and World Bank. 2021. "Tracking Universal Health Coverage: 2021 Global Monitoring Report." Washington, DC: World Health Organization.