Case studies in standardization- comparative case studies of South Africa, India and Sierra Leone.

Interoperability is the ability to exchange data between two or more systems.

Integration here can be understood as the process of joining distinct systems in such a way that they appear as being a whole in a particular perspective.

The South African example

In 1994, as a legacy of apartheid, the new South Africa inherited one of the least equitable health care systems in the world; 60% of the resources were used by the private sector, serving only about 20% of the population. In order to rectify the wrongs of the past, the new government launched the Reconstruction and Development Program (RDP) with a title that clearly expresses its intent: the reconstruction and development of communities that suffered under apartheid. The restructuring of the health sector is based on a decentralized system of health districts. Development of a new national health information system to support the restructuring of the health sector is part of this plan.

During apartheid (1948–1993), the health services were extremely fragmented according to race, type of service, and the system of "homelands." Until May 1994, there were 14 departments of health at the central level: the "general" Department of National Health and Population Development, 3 specific "white," "Asian," and "colored" administrations, and 10 for "blacks," "homelands," and "self-governing states."

As one consequence of this fragmentation, there were no comprehensive national standards for data collection, and each province used different data sets, definitions, and standards. Furthermore, apartheid was characterized by extreme inequity in health services provision and health status between populations and racial groups. Equity in health services provision and health status has therefore been a major political target in post apartheid South Africa. The important issue in this case is that striving for equity between geographical areas and racial groups required a system of national standards to measure and monitor the extent to which this policy is being achieved and to pinpoint areas where more resources and efforts are needed.

The Health Information Systems Program (HISP) was initiated as part of the RDP in 1995 with the aim to develop integrated district based information systems supporting the new emerging decentralised structures in South Africa. Starting in 3 pilot districts in Cape Town, the focus was on strategies for developing shared data standards as a way to counteract the extreme fragmentation of health services and health programmes serving different population groups making up the legacy of apartheid. Figure 3.6 below illustrates the typical situation in a district in South Africa, where a multitude of different services all report to different head office outside the district, and no information was shared or coordinated within the district.
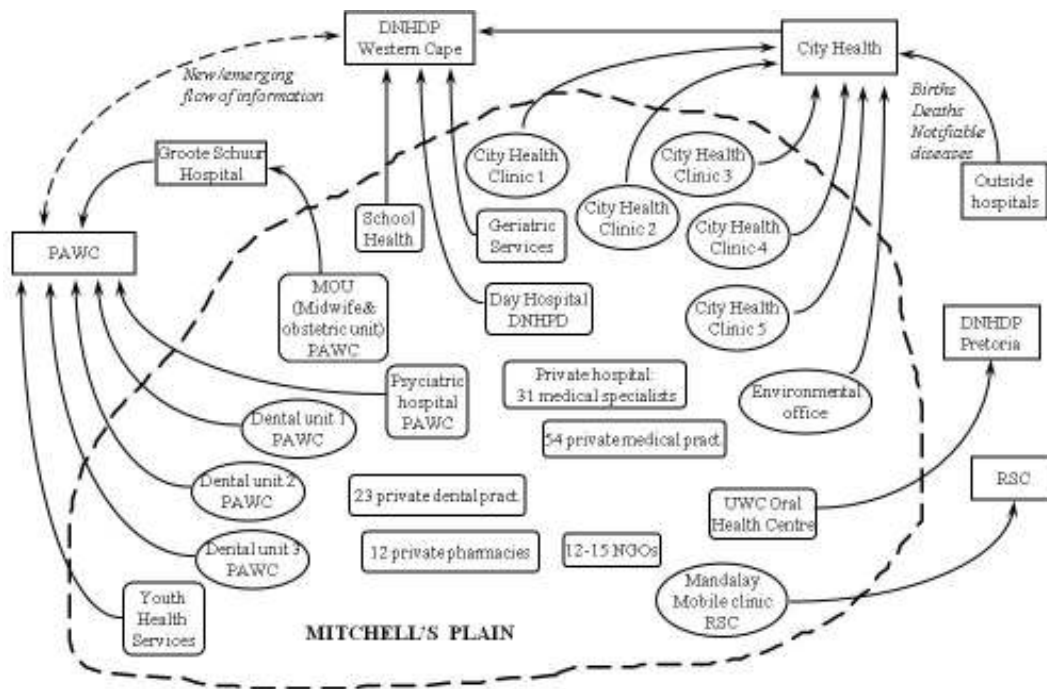
Figure 3.6: Legacy of apartheid; fragmented and centralised data flows, Mitchell's Plain district in Cape Town. The figure featured in the document proposing to establish HISP written by the Western Cape Strategic management Committee 1994.

The strategy for moving from a fragmented centralised to a decentralised integrated information and management system is depicted in the following Figure 3.7 from Atlantis sub-district outside Cape Town; creating a repository for all information being reported from the district within the district and share it so that it can be used by all stakeholders. In this book, we label this as the data warehouse strategy.

HISP was responsible for efforts towards standardization and software development. Standardization efforts aimed at developing minimal data sets - a uniform and minimal set of data elements (a data set) with clear definitions to be reported by all health facilities using simple paper reporting forms. These were essential in that the data elements should cover the key information needs across and be used to calculate the most important indicators. Implicit in this approach was that the reporting on the minimal data set would be established in addition to the data already being collected by the various health programs and organizational structures. Software development and prototyping efforts led to the creation of the first version of DHIS application which supported the implementation and use of the minimum datasets.

In order to reach consensus on data sets, widespread negotiations and consultation with different health programs and services were carried out starting from the Western Cape Province. After about 9 months of intensive negotiations driven by local managers in collaboration with the HISP team, the first essential data set was implemented in all local government health facilities in the Cape Metropole in May 1997, and later in the whole of Western Cape. The first version of the DHIS was implemented and used to capture and analyze monthly data at district, regional, and provincial levels in Western Cape from 1998.

Simultaneously, in the Eastern Cape Province a unified monthly minimal data set was implemented as from January 1998, in all primary health care centers. The DHIS software application was used to manage the new minimum data sets in the two provinces. Interestingly, the two minimum data sets resulted from two different processes in two different provinces, were very different with only about 50% overlap. The DHIS had already at that time a flexible meta data structure that could be used to manage multiple data sets; while the data elements that were the same in both data sets could be seen as a "shared" core data set, the standards that where specific for each province could be seen as a locally defined extension of the shared core, and the concept of organizing the data sets in a hierarchy was born.

In a situation where there were no shared national standards, and in the other provinces, no shared provincial standards, the fact that two provinces had developed their own shared standards, following similar approaches and managed by the same software application appeared as a significant success. From 1999, the DHIS and HISP approaches were endorsed as official national approaches and projects to establish provincial data sets and agreed processes were started in all provinces, and the first national essential data set was agreed upon in June 2000. While all provinces maintained their own extended data sets, the national data set made up the shared core which all provinces needed to collect and report on. This data set was revised several times since then. Over time, additional data elements were added, either as a process of expanding existing program data sets, or to accommodate new vertical programs (e.g., HIV/AIDS programs). By 2005, it was expanded to become a national indicator dataset (NIDS), reflecting the increased focus on the use of indicators, as compared to the earlier focus on data elements.

While some health programs, such as within the HIV/AIDS segment, maintained their own full-fledged systems and only provided extracts to the NIDS, managed by the DHIS, others, such as the Extended Program on Immunization (EPI), was fully included in the essential data set and DHIS routines. The example of EPI: In 2000 EPI was collecting 44 data elements through their own system, whereas 5 data elements were collected through the minimal data set. They then realized that the completeness and overall data quality were much higher for the few data elements being collected through the national "minimal" system and DHIS, than for the much larger data set they collected themselves. As a consequence, the EPI reduced their data elements from 44 to 13 elements and included them in the national reporting system. In this way, the national system grew in strength and quality and was able to provide "everybody" with better data than they had when each one of them maintained their own system.

The Indian example

From around 2000, the South African approach to integration by using unified data set and reporting tools was attempted applied in other countries, typically in smaller scale settings. The example of India illustrates how these bottom-up approaches need to be modified and adapted to the more stable and entrenchedbureaucratic structures of other countries.

The HISP initiative from South Africa was adopted first in India in December 1999 in the state of Andhra Pradesh in the South. Initially, the HISP team, inspired by the South African experience, and maybe a little naïve about the "heavy hand" of the centre, tried to develop minimum data sets following a situation analysis of 12 pilot clinics in one district of the State. However, these attempts were largely unsuccessful as the national mandated data sets could not be modified by levels below. Even though health is a state subject in India, historically the data collection formats have been largely governed through nationally defined vertical health programs, in which international donors have a significant influence. Unable to make a dent in this process of modification, HISP India then adopted an approach of automating existing formats on an "as is" basis, hoping that once HISP and DHIS is more deeply entrenched they may have some leeway in influencing the design of the datasets.

However, the above opportunity did not come through easily, and after 5 years of struggle, HISP was terminated as the state government had changed following the elections which also had the fall out of a change in the bureaucratic structures. HISP moved first to Kerala state, and then other states such as Gujarat, Jharkhand, and Madhya Pradesh. In each of the states, it was found that while the state health secretaries agreed to the logic of the "Minimum Data Set" approach, when it actually came to making concrete changes in what data gets collected and why, they would back off saying that "we cant change as the national level wants this data." So, in all the cases invariably, HISP India designed the respective state applications on a "as is" basis implying the collection of nearly 3000 data elements on a monthly basis. However, by making data from across program areas and for the whole state available for analysis through the DHIS data warehouse, important lessons on data quality and the eventual usefulness of the various data were drawn by a multitude of stakeholders, lessons that later provided valuable input in the national reform process.

In response to a grim state of public health affairs despite the achievements in the economic growth sphere, in 2005 a national mission was launched called the National Rural Health Mission (NRHM). The NRHM was established with the explicit aim of bringing about architectural corrections within a health systems framework in order to make health care, especially to the marginalized poor, more equitable, affordable and accountable. Amongst the health systems identified for introducing these architectural corrections was the area of HIS, including making these systems more integrated and decentralized.

While there had been previously (prior to 2008) various attempts both through the national level and various states to carry out HIS related reforms, systems remained fragmented and data driven. In 2008, another initiative was taken on redesign which took technical inputs from the HISP team, including one person who had been deeply involved in the HISP process in South Africa. The redesign started with a situation analysis of the HMIS of 4-5 states to identify what were the existing systemic problems in the HIS. This analysis led to an inductively derived set of 6 principles which could be applied to HIS redesign including of

datasets, reporting formats, indicator lists, and feedback cycles, as well as the DHIS data warehouse application.

In carrying out the situation analysis, HMIS data from some states was analyzed for 2007-2008. The problems identified across these states were quite uniform; a very high number of data elements being collected, of which a high number of zeros or blanks were reported (ranging from 50 to 60%). While a lot of data elements were being routinely collected, less than 5% of them were actually being used for the generation of indicators reflecting very poor utilization of information for action. Further, a lot of disaggregated data was being collected (eg each data element being broken up by caste groups – Schedule Castes, Schedule Tribes, Others) on a routine basis, which would be better off being collected on an annual basis through surveys. Another systemic challenge which existed in the HIS was the historical practice of the field nurses collecting "area based data" which involved them going to the houses and asking for example, if a delivery has taken place and recording it even though it may have taken place in a facility (like a hospital or medical college) which may be outside the jurisdiction of the nurse. A problem of duplication would arise as the same delivery noted by the nurse in her facility (the sub centre) records was also recorded by the hospital or medical college where the delivery physically took place.

To address these identified problems at a systemic level, HIS guiding principles were inductively derived to help design based on the overall philosophy of developing a HIS that could support the broader NRHM agendas of decentralization and use of information for action. These principles included reducing duplication of reporting, use of surveys for collecting data not used routinely, build the HIS on indicators rather than raw numbers, and establish a hierarchy of indicators for the different levels, as in South Africa.

After a long period of intensive negotiation across the different divisions, some decisions on redesign were taken which led to a nearly 90% reduction of data elements to what existed earlier. Two of the national programs (Immunization and Blindness Control) were integrated into the existing routine HIS. With other national programs (eg TB and Malaria) basically adopted the strategy of "wait and watch" to see how the new HIS would fare before deciding whether or not they would agree to integrate. Till then, it was decided to adopt a strategy of "reverse integration" where the core indicators required from the disease programs would be imported from their systems, electronically or manually as the case may be.

In this way, the national HIS was redesigned with the aim of simplification, integration and also decentralization. Simplification came in the form of reduction of data elements, the creation of facility specific datasets, and clearly distinguishing between recording and reporting formats. Integration came with the incorporation of datasets from two of the existing vertical programs into the routine HIS, and agreeing on a phased strategy for the other programs. Decentralization was designed by reducing work load of field workers with respect to the amount of data they had to collect, and setting up an environment where they could focus more on how they could use indicators to manage their everyday processes. The entire set of new recording and reporting formats were then under signature from the Mission Director, NRHM, sent to all the states with the instructions that they should immediately switch over to the new formats, and all the existing formats (except those disease specific programs not yet integrated in the HIS) would need to be discontinued with immediate effect.

The Indian example, as contrasted with the South African case was a top down national level driven initiative. While in South Africa, the climate and political will was to create "new systems," in India it was trying to make changes within a deeply entrenched system in which various stakeholders preferred the status quo. Standards established in India included the data elements, data sets, reporting formats and reporting relationships – a mix of technical and practice related. The case of standardization in South Africa, and to some extent also the case from India, emphasized flexibility in standard setting *vertically* in the hierarchy, analogue to what we have labeled vertical integration, such as e.g. integrating the line of management from the policy setting and managerial levels to the peripheral operational level. In the standard setting you are free to add your own standards, for your own domain, which may include levels below, as long as you adhere to the standards of the level above. Vertical integration is very much about managing levels of granularity, levels of aggregation. In the India case, what could not be effectively agreed was on the hierarchy, with the national statistics division insisting on receiving all the data that was being collected at the facility level. The national level insisted on collecting "monitoring" data instead of "evaluation" data, which by design could not be useful for them to convert to action..

The Sierra Leone case

In 2002, the civil war that had raged Sierra Leone since 1990 was declared over. Having cost 50 000 lives, and made many more refugees, the war had had a devastating impact on the small West African country, which is currently ranked among the *last* on the United Nation's development index.

Sierra Leone is a Health Metrics Network's (HMN) pilot country for HIS. The HMN formed a strong partnership with the Ministry of Health in 2005, thus forming a solid political base for the integration project. After an initial HIS assessment revealed substantial challenges in relation to a myriad of data reporting structures and weak ministerial data management, HISP got involved in developing a detailed plan of action during the last half of 2007, the implementation of which commenced in early 2008. Two efforts to integrate data reporting had been carried out the previous two years, each of them consisted of the development of a new paper reporting form that included parts of other forms and data sets and a computer database in which to capture the data. The most recent integration effort was from early 2007 and focused on Mother and Child health and immunisation services and came with a customised database application based on EpiInfo. It was partly overlapping the preceding integration effort, which also included disease data, and which was still in use, also with a database application. In addition came several other forms overlapping both of the initiatives. Numerous errors and inconsistencies were identified, but the developers had left the country and the effort had no more funds. Furthermore, the UNAIDS CRIS application for HIV/AIDS data was implemented in all districts, also overlapping with the other systems on Mother and Child Health data.

The situation by early 2008 was thus that in a typical district information office there were three software applications in more or less use, each capturing data sets overlapping with all the others, no communication between the applications, and many more data collection forms not captured by any software. "Isn't it possible to get all this data into one database? Can DHIS do that?" were the initial questions we were asked by the Ministry of Health when we started the planning in 2007. In order to respond to this request, we started to set up the DHIS as a data warehouse to be used at the district level which could extract and load data from the three other systems. This rather complex approach, given the state of the technical infrastructure at district level in Sierra Leone, was selected in order to not offend the various system owners. However, as things developed during 2008, it turned out that most districts stopped using HIV/AIDS-system CRIS; the HIV/AIDS staff employed by UNAIDS got many times the salary of the information officers, who then did not want to enter the data for them; "why don't they do it themselves, they don't have that much work to do!", as one information officer stated. Also the EpiInfo project run out of funding and stopped. As a consequence, the decision was taken to replace all applications with the DHIS as it was rolled out to the districts.
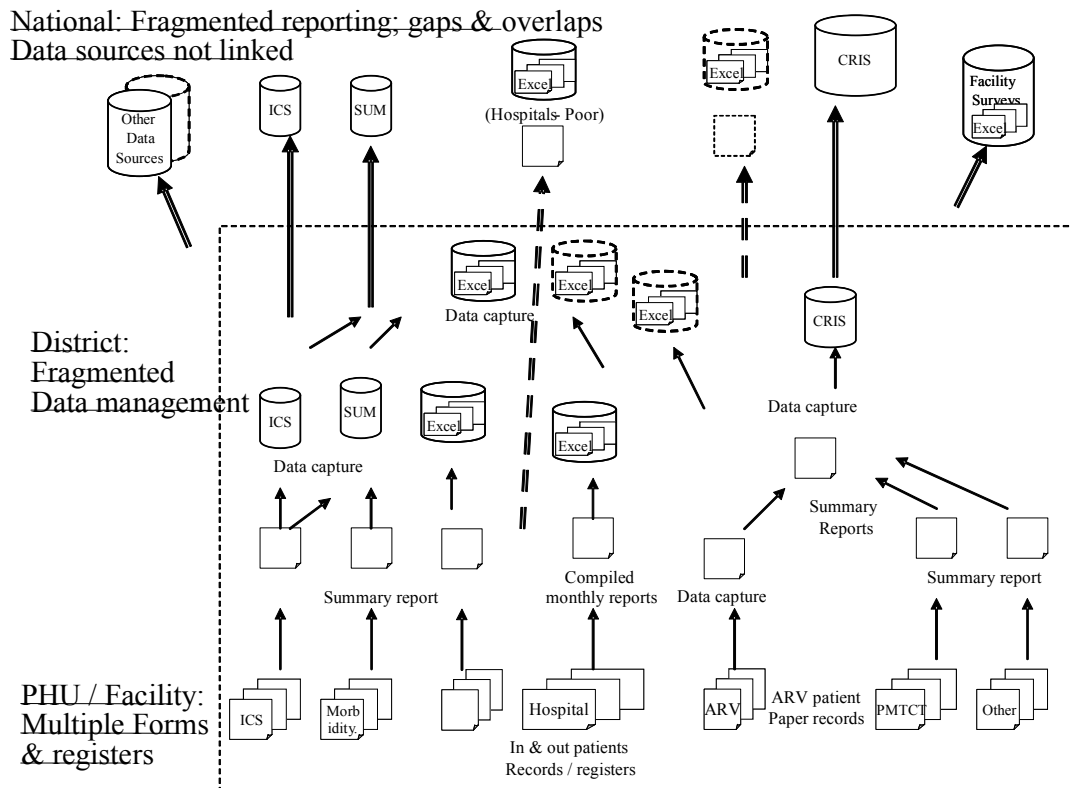
Figure 3.7: Prior to interventions, fragmented HIS in Sierra Leone.

Given the fragmented, overlapping and inconsistent character of the existing data sets, a major effort went into designing a data structure in the DHIS that enabled the capturing of all the data from the paper forms in use through a computer user interface mimicking each of the forms, while at the same time solving the inconsistencies behind the scene, in the database. Overlapping data that had been entered for one paper form using the corresponding "computer form" would then show up in the other computer forms it belonged to, without having to be entered into the database again. A consistent maximum data set had been derived from a multitude of overlapping forms.

Following a 2 times 3 weeks training of all district and national information officers in May and June, 2008, a major effort was directed to get all the data captured in the district and reported electronically to the national database. The most important result has been the documentation of the quality of data including the completeness. The most advanced district information officers have been important in spreading the lesson that "we can get all the data from all the forms into one database, why do we still have all these overlapping and complicated paper forms?" For the first time all the data collected were available in one database, which again convinced all stakeholders that it was actually possible to harmonise and integrate the data reporting from all programs. A revised set of data collection forms was drafted and circulated among stakeholders for feedback alongside a drafted list of indicators with definitions of the data needed to calculate them. In January 2009, a workshop with all the important stakeholders, program managers and district representatives agreed on revised and integrated data collection tools and data and indicator sets. Given that there were some

refinements to these collection tools, and the time it took to print and distribute them, these new forms were in effect in all facilities from January 2010.

As the new data collection forms were developed over the summer of 2009, the project started to approach the various international agencies involved in health service delivery in Sierra Leone. The districts, which for the first time not just had most of their data available in one data warehouse but also had electronic tools to ease the analysis and dissemination of this data, had been given extensive training in the use of information. Some best performing districts had started to engage civil society to improve their health indicators, a process that was fuelled by the national level publishing league tables ranking the districts on key indicators. UNICEF was the first major international organization that pledged to support the data warehouse structure and drop their own parallel reporting structures, and others followed suit. This integration at the organizational level was only made possible by creating an attractor; a system that had proven results in data completeness and quality. By the second quarter of 2009, the percentage of facilities that had reported the main data collection form reached above 90 for 7 out of 13 districts, up from 4 districts the quarter before. It was exactly this proven ability to increase data completeness that convinced the health development partners to support the DHIS. While it was not just the software by itself that increased completeness, the whole supporting structure of capacity building, information products, and quite aggressive dissemination of feedback back to districts to create an environment of friendly competition and peer pressure was what led the health development partners to accept the standards implemented in the software.

The Sierra Leone example shows the power of the whole data warehouse approach to develop data standards in practice. By firstly, enabling the pulling in of all data from different sources into the DHIS2, and then using its analytical capabilities to make visible the problems of redundancies and gaps, a stronger motivation and buy in is obtained from the stakeholders to agree on data standards.

Example of Integration and interoperability: Developing and applying the SDMX-HD standard in Sierra Leone.
.

In late 2009, WHO had initiated a process to develop a data exchange standard for health metadata and statistics, building on an existing standard for financial transactions. This new protocol, SDMX-HD, was still not implemented by any application, far less in any real use-case setting by early 2010.

Nevertheless, HMN, MoH in Sierra Leone and HISP decided to push this forward with a pilot system that could both track patients on Anti-Retroviral Therapy (ART) and share this data with the data warehouse running on DHIS2, as stipulated in the original strategic plan for the country. OpenMRS, a medical record application, was set up in the main hospital in Freetown, the capital, and in the weeks before this took place in February 2010, major advancements were made to SDMX-HD, spurred by the fact that it was about to be implemented in two "live" applications. The main idea is that OpenMRS has the functionalities needed for doctors and nurses to track patients over time, while the aggregate data from this system is used for general health management at hospital, district, and national levels. This two-level use applies to other domains and applications also, such as logistics management, human resource management, lab systems, health finances etc. With SDMX-

HD set to go live in Sierra Leone at a given date, developers behind DHIS2 and OpenMRS not only pushed development forward by implementing this standard in the respective applications, they also contributed the necessary use-case to solve outstanding issues. The development of SDMX-HD made an HIS architecture of various applications for different business domains a reality; anyone able to share data on this protocol could now be "plugged in" in a country HIS.

While HMN and HISP were collaborating in Sierra Leone, Capacity*Plus,* a partner of HISP and HMN specializing in strengthening health workforce information systems, was partnering with the West Africa health Organisation (WAHO) to pilot an open source HRIS called iHRIS in Ghana. Learning about the "Sierra Leone architecture", WAHO took the initiative to include also the iHRIS suite of applications for HR management in this architecture, which again pushed the iHRIS implementation of SDMX-HD further. These initiatives, together with the fact that many countries in the West African region were in the process of introducing the DHIS and / or the iHRIS, led to the organisation of a workshop on training in DHIS2 and iHRIS in Accra, Ghana, September 2010. The WHO organised a "Connectathon" meeting for the SDMX-HD in parallel with the workshop. At the end of the workshop, the SDMX-HD standard was officially launched as implemented in the DHIS2, iHRIS, OpenMRS, and the WHO Indicator Measurement Registry (IMR) integrated framework. This initiative was further consolidated in November 2010 where HIS staff from all 15 WAHO member states were present. There, it was decided that HMN, WAHO, HISP, and Capacity*Plus* should collaborate to develop a centre of excellence at WAHO for supporting member countries in adopting the interoperable solutions which had grown out of Sierra Leone.

The three applications have a relationship of integration and interoperability with each other using this SDMX-HD standard. **Interoperability** is the ability to exchange data between two or more systems. In the figure below, we see exchange of data between DHIS and OpenMRS and between DHIS2 and iHRIS. This is where we have interoperability between, in this case, two systems. There is no interoperability between OpenMRS and iHRIS. **Integration** here can be understood as the process of joining distinct systems in such a way that they appear as being a whole in a particular perspective. In this case we see that patient data from OpenMRS and human resource data from iHRIS are "joined" and integrated in the DHIS2 data warehouse. OpenMRS and iHIRS are integrated, but there are no interoperability between them. DHIS2 is integrated with both OpenMRS and iHRIS.

Example: DHIS is calculating the indicator
Deliveries per midwife per Facility
for the month of May

Example SDMX-HD:
DHIS "asks" OpenMRS
to send following data:
 # deliveries by health
centre for month of
May.
OpenMRS aggregates
and send requested data

DHIS2: Data warehouse for aggregate data

Example SDMX-HD:
DHIS "asks" IHRIS to
send following data:
#midwifes by health
centre for month of
May.
iHRIS aggregates and
send requested data

Interoperability

OpenMRS : Medical records

Integration

iHRIS: Human Resource records