# A national framework for electronic health records

------ Draft ------

iSPIRT Foundation

3rd February, 2017

## Summary

The volume of health data in digital form is growing and, in response, the penetration of IT infrastructure in healthcare delivery systems is rising. In parallel, due to factors such as migration, individuals are served by numerous health providers, of different sizes, geographies, and constitutions. Thus, it is time for the government to propose policy recommendations to ensure security and interoperability of patient data systems. In this note, we present a framework for patient data exchange between healthcare providers and for the creation of *longitudinal individual health records*; we refer to this as electronic health records (or EHR) in the rest of the document. We first outline the design principles guiding our proposed framework, then outline our proposal presenting a few example use cases. Our framework builds on the digital locker and consent frameworks endorsed by the ministry of electronics and information technology, uses principles of Internet-scale architectures and is designed with data security and patient-centricity in mind. Our hope is to support the government in its efforts to produce a world-class infrastructure for healthcare data in India.

## Introduction

Healthcare, in India, is very diverse and fragmented. A typical patient, over the course of her life, interacts with multiple hospitals, diagnostic labs, independent medical practitioners, pharmacies, insurance companies, insurance agents, third-party administrators (TPAs), and various healthcare intermediaries (e.g., ASHA workers), who are all critical parts of the healthcare delivery chain. Owing to various socio-economic factors (like high migration and social mobility rates), individuals are served by a large spread of such providers, of different sizes, geographies, and constitutions.

Providers have started to deploy IT infrastructure for process efficiencies and for managing quality of care. New healthcare technology companies have emerged, offering technology for communication between patient and provider as well as for various data management and administrative tasks, ranging from things such as patient self-care to managing personnel and inventories. As a result, there is a surge in digital health data in the country. New national-level schemes such as NHPS[1] and e-hospital[2] and the digitization of existing interventions for maternal/child care, TB, etc will generate digital data on hundreds of millions of citizens, and further contribute to the surge of health data.

The IT systems of different providers are being developed independently of each other, without adherence to common standards. This fragmentation has the undesirable consequence of the systems not communicating well with each other, fostering redundant data collection across systems, inadequate patient identification, and, in many cases, privacy violations. This fragmentation problem is also recognized in the ministry of health and family welfare (MoHFW)'s recent notification of electronic health record (EHR) standards[3].

[1] http://www.pradhanmantriyojana.co.in/national-health-protection-scheme-insurance-cover-rs-1-lakh-poor-bpl/
[2] http://www.nic.in/projects/e-hospital
[3] http://mohfw.nic.in/WriteReadData/l892s/2857976581461059607.pdf

As IT infrastructure problems emerge in healthcare, progressive changes are occurring in data management policy in the country. The digital locker framework and consent framework have been notified by the Ministry of Electronics and Information Technology (MEITY). The Reserve Bank of India has introduced new data aggregation policy emphasizing auditable and revocable user consent as a central requirement for companies to acquire financial data about individuals[4].

A similar policy could provide solutions to the impending data management problem in healthcare IT systems as well as guide the formulation of data exchange policy in healthcare. In introducing such a policy, we have a golden opportunity to provide an open, interoperable and secure interchange platform for health data, extensible to include other types of digital medical data in the future. With thoughtful design and market interventions, we can enable quick adoption of such a platform across public and private entities.

## EHRs will significantly benefit the overall health landscape of India

Proper management of health data and enabling data exchange is necessary for increasing efficiency of care delivery and increasing value-based care. This can happen in the following ways:

1. **Easier access to patient history**: By enabling access to longitudinal individual health records, we will empower doctors, to make better diagnosis, enable early risk detection, eliminate repeated lab tests and make healthcare spending more judicious.
2. **National health informatics:** Government departments and public health researchers use a variety of field data to track health indicators for various diseases and conditions such as infant mortality rates (IMR) and spread of vector-borne diseases. With a national EHR framework, these indicators could be calculated with the use of up-to-date EHR data, hence making them more reliable and of high quality. Health analytics based on EHR data can guide decisions on where healthcare funding and resources must be directed to address key health issues.
3. **Efficient epidemic surveillance:** India's infrastructure for surveillance is in formative stages. Most epidemiological work today is human-intensive[5]: trained epidemiologists, working as medical practitioners and surveyors, conduct activities around epidemic detection, reporting and response. EHRs can provide near real-time epidemic detection. Using EHR data for epidemic is an active area of research in other countries as well[6].
4. **Facilitating medical research:** If data about patients is easily available and organized, it will fuel medical research which, in turn, can improve health indicators in the long run (e.g., medical research on maternal health data can lead to new interventions to reduce infant mortality). Medical research, conducted using large volumes of data, can lead not only to better disease

[4] https://rbi.org.in/Scripts/NotificationUser.aspx?Id=10598&Mode=0

[5] Dikid *et al.* *Emerging & re-emerging infections in India: An overview*, Indian Journal of Medical Research, Jul 2013. Also see a news report on the emerging epidemiology intelligence practice in India: In the works: An epidemiology intelligence service, Jul 2013.

[6] Zheng *et al. Epidemic Surveillance Using an Electronic Medical Record: An Empiric Approach to Performance Improvement*, PloS 2014. Available online.

response but also to disease prevention[7]. Early disease detection and prevention mechanisms are currently weak in India and EHRs can help improve the situation.

## Design principles for an EHR framework

We believe that any technology framework for enabling health data exchange should follow the below design principles.

1. **Patient is the owner of his/her data:** As specified in ministry of health and family welfare (MoHFW)'s recent notification of electronic health record (EHR) standards[8]

> *"The physical or electronic records, which are generated by the healthcare provider, are held in trust by them on behalf of the patient. The contained data which are the sensitive personal data of the patient is owned by the patient herself. The medium of storage or transmission of such electronic medical record will be owned by the healthcare provider. The 'sensitive personal information (SPI) and personal information (PI)' of the patient is owned by the patient themselves. Refer to IT Act 2000 for the definition of SPI and PI."*

   As a fundamental principle, health data *about* an individual that are generated by a service provider (e.g., a hospital) are owned by that individual. Any instance in which such data are to be shared by the service provider with a different entity must involve consent and authorization from the data owner (except, perhaps, in exceptional cases like sex determination, and epidemic detection). This individual ownership principle is also the foundation on which the digital locker and consent frameworks are built.

2. **Privacy by design:** User data needs to be protected from abuse and compromise. The EHR framework must define data sharing mechanisms in a manner that ensures privacy of user data ground-up. Tools to protect privacy of data must be in-built in the framework and best-practice guidelines should be in place for the framework users (hospitals, insurers, and other stakeholders) to ensure privacy of data.

3. **Minimal changes to existing health record formats:** The diverse IT products used by healthcare providers in India do not follow a common standard for data storage and do not communicate with each other. We propose to enforce minimal to no changes to existing IT products. We can ensure quick adoption of the framework if it is designed to work with the existing IT infrastructure as-is. We expect IT systems to improve and adopt standards over time.

4. **Clear incentive for providers to participate:** Almost 80% of Indians seek private healthcare and pay out of pocket. In order to ensure the private healthcare providers participate, the framework should support built-in incentive structures to ensure rapid and universal adoption. Incentives and regulatory control need to be suitably balanced, in order to ensure maximum participation and innovation, and still ensuring user privacy and control.

[7] Precision Health Research at Stanford University: http://med.stanford.edu/precisionhealth.html
[8] http://mohfw.nic.in/WriteReadData/l892s/2857976581461059607.pdf

5. **Minimalistic design**: The framework should be minimalistic in terms of the number of requirements and constraints it places on framework users. It should be easy to adopt and it should be easy to build specialized applications on top of the framework. Rules for data formats to be used by stakeholders must be minimal. As an example, UIDAI has only four mandatory fields to include in a person's Aadhaar ID.

6. **Open APIs:** The framework should provide an *open* and *standard* set of application programming interfaces (APIs) for creating, accessing and updating records in EHRs, as proposed in the Policy for Open APIs by MEITY[9]. The API definitions should be simple and follow the principles of minimalism and privacy by design. Some parts of the framework could be publicly available (as public goods) for any framework user to utilize.

7. **Internet scale:** The framework should ensure high availability of health data from EHRs and high efficiency of data access by data requesting entities. Availability and efficiency guarantees should be similar to Internet-scale systems (such as Google, Facebook, Amazon Web services).

## Our proposal

We propose a general framework for data exchange between health IT systems and for creation of electronic health records, following the above design principles. Our proposal is based on the digital locker framework and the consent frameworks introduced by MEITY[10]. We first summarize these two frameworks and then present our proposal.

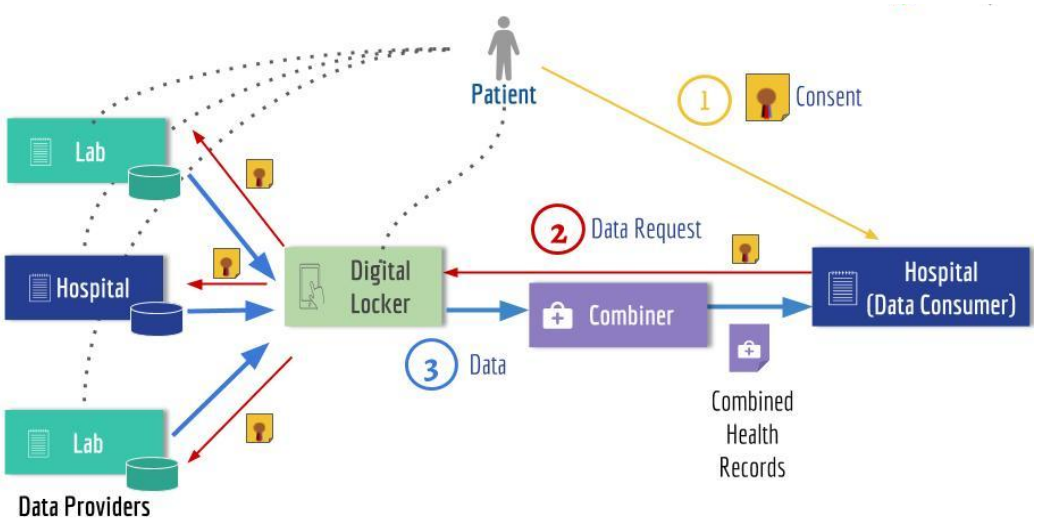
MEITY put forward a **consent framework** to facilitate safe, API-driven, user-controlled data sharing between different IT systems. Under this framework, any service provider seeking digital data about an individual, called the *data consumer (DC)*, from another service provider that generates the data, called the *data provider (DP)*, must obtain consent from the individual *in electronic form*. This electronic form of consent is referred to as a *consent artefact* and comprises a machine-readable file that captures different parameters involved in the data share transaction such as the names of the entities involved, the type and duration of the data being shared, the purpose of sharing and the lifetime of consent. Each artefact contains a unique identifier of the individual (e.g., his Aadhaar number) and has a digital signature associated with that identifier embedded in it. The data consumer must provide the consent artefact to the provider when requesting data and the latter must share data only when it has validated the artefact and the embedded signature. The use of digital signatures tied to a strong digital identity ensures that consent artefacts cannot be forged and significantly reduces the chances of unauthorized entities getting access to user data. This, amongst other features, makes the consent framework vastly superior to other data authorization mechanisms in online systems.

[9] http://meity.gov.in/sites/upload_files/dit/files/Open_APIs_19May2015.pdf
[10] http://meity.gov.in/content/public-consultation-digital-locker-interoperability-documents

Along with the consent framework, MEITY has also introduced the **digital locker framework**, which uses the consent framework to enable secure, streamlined access to user data from different data providers. A digital locker is essentially a standardized intermediary which interfaces between DCs on one side and DPs on the other. It provides a mechanism to store user data in a federated manner: each locker communicates with a set of data providers and manages access to data generated by the providers. The locker stores either a copy of the data itself or a link to where the data resides at source and it communicates with other lockers in the network to build an aggregated view of all user data. Any DC seeking access to an individual's data from multiple DPs must sign up with a digital locker provider and request the locker provider to fetch data on its behalf, which the locker does by requesting other locker providers and the relevant DPs. Similarly, an individual himself can get a consolidated view of his data by getting an account at a locker provider. Digital lockers use the consent framework to ensure unauthorized entities are not allowed access to an individual's data. Digital lockers also provide different grades of access control, e.g., for highly sensitive forms of data, the DC may be required to collect fresh consent from the individual for every view of that data i.e. copying of data may be disallowed. Finally, strong audit and notification mechanisms are built into both the frameworks, which provide further security guarantees.

## The Primary Use Case: Integrated Patient Records

We describe our proposed framework using an example illustrating the primary use case of the framework. Suppose a patient visits multiple labs, hospitals and other care providers over a period of time. Each care provider has its own IT system storing health data about the patient and is thus a potential data provider. Suppose that this patient now visits a new hospital, which needs to get access (as a data consumer) to the health data stored in the IT systems of the other care providers the patient visited earlier. Using the digital locker concept, this new hospital can obtain the patient's prior data and create a combined master health record of the patient in real time. This combined master record is what we refer to as the EHR of the patient.



The above figure shows a single digital locker provider for simplicity; in general, there could be many. Each of these must register as a digital locker provider using MEITY's registration guidelines. The past care providers of the patient must each sign up with at least one of the digital locker providers who store links to the patient's data with that provider (or the data itself) as and when such data are generated. The digital locker framework provides APIs to be implemented as part of such a sign-up. Each patient record is indexed against a strong, de-duplicable patient identifier like the patient's Aadhaar number. Each patient record itself has a unique record ID (called *document ID* in the digital locker framework) associated with it, which is unique across all records kept by all data providers.

The data consumer hospital must also sign up with a digital locker provider of its choice. To create the patient's EHR, it does the following:

1. First, it **collects consent** from the patient to access his data from all registered care providers that have the patient's data. One consent artefact for every care provider is generated and each artefact has information like the name of the care provider, the ID of the patient record(s) at the provider being accessed and an identifier of the patient[11]. The requesting hospital stores all consent artefacts generated in the process.
2. Second, the hospital **sends a data request**, along with the consent artefacts, to its digital locker provider who forwards the request to the relevant data providers, directly or via other locker providers. Each data provider is sent only the artefact that is relevant to it (contains consent for accessing data at that provider). Upon validating the consent artefact, each data provider sends the requested data back, which eventually gets aggregated at the requesting digital locker. All these exchanges take place via standard APIs specified in the digital locker framework.
3. Third, optionally, a software component called the **combiner** processes the aggregated data to produce a single health record, the EHR of the patient. The combiner works in an automated or semi-automated manner to learn the schema underlying the individual records and to produce a merged record in a standard format. The combiner may be a specialized service provided by an entity different from the others, or may be provided by digital locker providers.
4. Fourth, the data consumer hospital receives the aggregated data, combined or uncombined, as the case may be.

A few important points regarding our approach of generating EHRs:

a. **Patient data security and privacy** are provided via the security controls built into the digital locker and consent frameworks. Consent as defined in these frameworks is strong, difficult to forge and non-repudiable and, as stated above, there are additional controls in the digital locker framework to ensure data safety.
b. **There is no central store of EHRs.** Different hospitals could sign up with different digital locker providers. As data providers, they issue the medical records they create to the digital locker they sign up with. As data consumers, they get access to the longitudinal individual health records. The framework takes care of aggregating the records across different digital lockers.

[11] Although multiple consent artefacts are generated by the system, the user interface can still be simple and enable the patient to consent to all data collection in a single action.

c. We do not specify a mechanism for the combiner but note that it is possible to combine data for certain data types in a **"schema-less"** manner (i.e. without asking the data providers to specify the schema as part of the data), by applying modern machine learning techniques for schema mapping[12]. This is particularly the case with diagnostic data. For other forms of data (clinical data, prescriptions, etc), semi-automated approaches e.g., machine learning in combination with human feedback on mapping errors, would be needed.
d. Ensuring **a universal patient identifier** is essential for successful data aggregation and combining. We believe that Aadhaar is the most suitable choice for this identifier. It is already available to more than 95% of Indian residents and it has mechanisms for de-duplication built in. Other IDs currently don't have the prevalence and the security advantages that Aadhaar enjoys. Populating existing databases with an identifier like Aadhaar will require a significant behavior change on the part of users (users will need to carry the ID for obtaining care) and this should be achievable through suitable incentive schemes.

Some entities may want to implement the digital locker and consent APIs for fetching data in their existing IT system. This would allow them to participate as data providers without having to sign up with a licensed digital locker provider. Other entities can fetch data from them *directly* i.e. without using a digital locker as an intermediary[13].

## Other Use Cases

Here are some more examples of use cases that can be realized using our framework.

### Insurance

Data is critical to insurance-based healthcare delivery, and with the national health protection scheme (NHPS) coming up, the use of insurance to cover healthcare costs is going to dramatically increase in the country. Using our data exchange framework, insurance companies can get access to richer data (richer than what is currently collected in a claims processes) and use this to evaluate claims more rigorously and to conduct better analytics. For example, given access to integrated patient records of beneficiaries, insurers can perform detailed portfolio analysis and price their insurance offerings more objectively. Electronic consent from patients enables them to directly interface with hospitals' Electronic Medical Record (EMR) systems and other sources of data and to increase automation of data collection.

### Second opinions

Consider a patient who goes to Hospital A where he is diagnosed with cancer. The patient wants to get a second opinion from Hospital B, where the doctor could use the local EMR software to try to fetch

[12] There has been much work on schema mapping algorithms for Web documents e.g., http://ilpubs.stanford.edu:8090/851/1/2008-8.pdf. It seems feasible to extend this work to the case of medical data although this will require more experimentation.
[13] Medical software vendors in other countries have built their own custom APIs to enable provider-to-provider data sharing (e.g., Epic, a US-based EMR provider, has done this: http://www.deaconess.com/Deaconess-Electronic-Health-Record/MyChart/Care-Everywhere-FAQs.aspx). We propose the use of *standard APIs* (based on the digital locker framework) for such sharing.

patient data residing in Hospital A's lab information management system (LIMS). The latter may have already implemented APIs for enabling other hospitals to fetch data from its LIMS. The patient provides consent to Hospital B from within the EMR software being used there and the EMR uses the resulting consent artefact to request data from Hospital A's LIMS. Upon validating the consent, the patient's data is returned. The doctor at Hospital B can now recommend additional tests to the patient based on what is unavailable from the obtained data. This is a simple use case which demonstrates how data can be exchanged between different care providers without the use of the digital locker.

### Medical research

EHRs built using our framework can also be applied for medical research. Medical research query workflows typically rely on a base query of the form "**Find patients with X**" where X could be any condition involving multiple patient features which, in turn, could be distributed temporally. Such a query enables the researcher to extract patient cohorts with a given condition.

To enable cohort extraction queries to be effectively executed in our framework, a few processes need to be in place. First, data providers need to implement the ability to respond to queries of the form "**Return all records of patient P**" where *P* is any patient identifier. Such queries, when run across data providers, allow digital lockers to aggregate *all* data about a patient in a single step. Additionally, the consent collection process for running such queries could be made more flexible in order to enable researchers to quickly gather patient data. The consent framework provides the ability to collect "**one-time consent**" from end users, using which individuals can grant permission to a data consumer to repeatedly access data about them from the same provider ad infinitum using the same consent artefact. Given the repetitive nature of data capture in medical research, one-time consent would be a suitable choice here. Consent could be collected from patients either by the medical researcher or by the care provider ahead of time.

Second, some entity needs to implement a mechanism to store aggregated patient data and to respond to queries of the form "**Find patients with X**" in a suitable query language. This functionality could be implemented by the medical researcher's digital locker provider itself or an intermediary who takes the aggregated data from the latter, combines it and answers such queries. We refer to this entity as the **query processor**.

Third, another intermediary between the medical researcher and its digital locker provider, would need to implement a *de-identification module,* in order to protect patient privacy. De-identification may involve a combination of different techniques like removal of columns (e.g., removal names, locations), introduction of small amounts of noise in query responses and "generalization techniques" (e.g., instead of reporting the age of a patient as the value 25, report it as a value in the range [20, 30]). Policy guidelines for performing de-identification on health data for medical research need to be set. We refer to the entity implementing de-identification as the **anonymizer**. The anonymizer may be the same entity as the query processor.

Given these three components, the medical research use case can be implemented as follows.

- First, the medical researcher, or each data provider participating in the research, collects one-time consent from patients whose data is to be used for the research.
- Second, the researcher maps the research task to a set of patient cohort queries of the form "**Find patients with X**" and each of these queries are issued to its query processor, along with the one-time consent artefacts required for the data fetch.
- Third, the query processor requests the researcher's digital locker provider to fetch data using the one-time consent artefacts. The locker provider gathers all the required data and submits them to the query processor.
- Fourth, the query processor runs the queries received earlier and submits responses to the anonymizer. The responses are de-identified by the anonymizer.
- Finally, the aggregated, de-identified cohort data are sent by the anonymizer to the researcher.

### National health informatics

Like private institutions, the government may also undertake medical research activities which could be of a more specific nature e.g., find out how many patients in a certain location have condition X (which could be a disease like TB or malaria) or find out how patients in the country with condition X are responding to treatment Y. These queries can also be mapped to patient cohort queries as in the medical research use case. For the current use case, a government department, like the HMIS division of the MoHFW, will play the role of a data consumer and collect consent from patients ahead of time i.e. before running analytics on their data. Patient consent collection could be bundled with other government programs e.g., at the time the public health department enrols citizens into its primary healthcare program, field officers could collect consent from enrollees for future data sharing with other government departments interested in health informatics. In certain situations (e.g., epidemic monitoring), consent collection may be waived by policy.

### Drug trials

EHRs could also be useful to pharma companies in conducting drug trials. Suppose a pharma company wants to test the effectiveness of a new drug against a common disease. They could take a sample of patients suffering from the disease and collect their consent to be able to view their health records containing drug-specific information from *any* hospital or caregiver. This allows them to query their health data periodically and track the effectiveness of the new drug over time. Because of the nature of EHRs in our framework (dynamically generated, and from multiple sources), such information can be collected by the company from a wider range of providers and faster.

## Incentives for participation

In order to incentivize participation in the digital locker framework, we propose that each data providing entity (hospital, lab, pharma company or others) should be allowed to earn **a base fee *per request* of any data record that it shares**. This would become an incentive for the data providers to start adopting the framework. Base fees should be of the order of the cost of storing and moving data; in general, this would be much less than the cost of care. Base fees should be different for textual and image data (e.g., scans) because the infrastructure required is different. Besides the base request fee, we propose **a base "issue" fee** that each data provider is eligible for, which is a one-time fee given to the

provider per data record at the time it agrees to share the record. This fee could be high initially to incentivize participation, and can diminish subsequently.

Digital locker providers will need to implement different approaches to cover their cost of operations and, in particular, to cover the fees they would need to pay to procure and store data. One approach is to charge data consumers for accessing health records from the locker system. Another is to charge patients for data-driven value-added services that lockers may provide. Data consumers whose operations depend critically on externally-sourced data, such as insurers, researchers and health informatics agencies, would be more willing to pay for data and thus, are likely to be served first by locker providers. Patients and care providers (hospitals etc), on the other hand, may be harder for them to acquire. Hospitals, for example, are better incentivized to generate patient data in-house than to source it from others (e.g., re-conduct diagnostic tests that have been conducted elsewhere). To engage them as data consumers, more indirect approaches would be needed. Here are two ways in which this could happen:

- **Pressure from insurers:** Where care is sponsored by an insurer, the latter has some control in the course of treatment. Insurers can direct hospitals away from their tendency to re-conduct tests or procedures by declining to pay for such tests. Policy instruments could be used to enforce prioritization of data re-use over re-generation in health insurance coverage.

- **Pressure from patients:** As competition amongst care providers increases, pricing will play an important role in attracting patients. Hospitals who participate in our framework as data consumers will enjoy a price advantage over others because of the low cost of acquiring health data it offers, and thus are likely to be preferred by patients.

## Conclusion

India is at a critical juncture in terms of health information systems. The volume of health data in digital form is growing fast and the penetration of IT infrastructure in healthcare delivery systems is small. This is a good time for the government to introduce policy to ensure that digital systems for managing patient data follow adequate security standards, are patient-centric and can interoperate seamlessly, as well as do this without suffering a high cost in changing existing systems. In this note, we have presented a general framework for enabling patient data exchange between healthcare providers and for creation of electronic health records, along with a few example use cases. Our framework builds on recently proposed government policy on data management (in particular, the digital locker and consent frameworks proposed by the ministry of electronics and information technology), requires very little change on the ground, uses well-recognized principles of Internet-scale architectures and is designed with data security and patient-centricity in mind. We hope this note will support the government in its efforts to transform healthcare IT in the country and in producing a world-class infrastructure for managing healthcare data.