

Generating and Knowledge Framework: Design and Open Specification

Usman Wajid¹, Chris Orton¹,
Andriana Magdalinou², Lydia
Montandon³

¹Information Catalyst, United Kingdom

²European Federation for Medical
Informatics, Lausanne, Switzerland

³Atos, Madrid, Spain

Corresponding Author. Usman Wajid,
Information Catalyst for Enterprise, Crewe,
CW15QR, United Kingdom; E-mail: uwajid@
gmail.com. ORCID ID: <http://www.orcid.org/0000-0002-3304-6166>.

doi: 10.5455/aim.2019.27.362-368

ACTA INFORM MED. 2019 DEC 27(5): 362-368

Received: Nov 26, 2019

Accepted: Dec 27, 2019

ABSTRACT

Introduction: The Information Aggregation (IA) component manages streaming and batch data deriving from a multitude sources in a scalable, efficient and reliable way to create Holistic Health Records (HHRs). Within this context, the IA component combines a number of diverse data sources into a common format and stores information in an available form to be used for analytics, simulations and decision making. **Aim:** The purpose of this paper is to provide an overview of the CrowdHEALTH project and the technical architecture of the CrowdHEALTH platform in order to put the aforementioned IA mechanism in context. This is followed by the design details and initial specifications of the first prototype of the IA component as well as its relationship with other components. **Methods:** The micro-service approach can be used to perform information aggregation and to update HHRs in the CrowdHEALTH platform. Micro-services are a variant of the service-oriented architecture (SOA) where applications are structured as a collection of loosely coupled services with defined interfaces. **Results:** Within the CrowdHEALTH architecture, the Information Aggregation component is situated between the Interoperability Layer and the CrowdHEALTH Datastore. The Information Aggregation component processes and aggregates interoperable data, before data aggregation in the HHRs and storage in the big datastore of CrowdHEALTH platform. The aggregation functions use big data management techniques and enhance the state of the art in specific areas such as the use of micro-services to perform synchronous aggregation operations on heterogeneous datasets. **Conclusions:** Although an initial version of the IA component was presented, the specifications and implementation level details will be further updated during the project's course.

Keywords: Data Aggregation, Health Records.

1. INTRODUCTION

The health data from diverse information sources (1) constitutes a collection of people profiles. Fitting the pieces together may provide a bigger picture of people's health status and improve knowledge discovery, research and personalized care (2). The Information Aggregation (IA) component enables the aggregation of different information sources to support the

creation of Holistic Health Records (HHRs) (3). The IA component manages streaming and batch data coming from various sources in a scalable, efficient and reliable way in order to develop Holistic Health Records (HHRs) (4). The aim of the IA component is to combine a number of isolated data sources into a common format and to store information in a form that can be available for analytics,

© 2019 Usman Wajid, Chris Orton, Andriana Magdalinou, Lydia Montandon

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

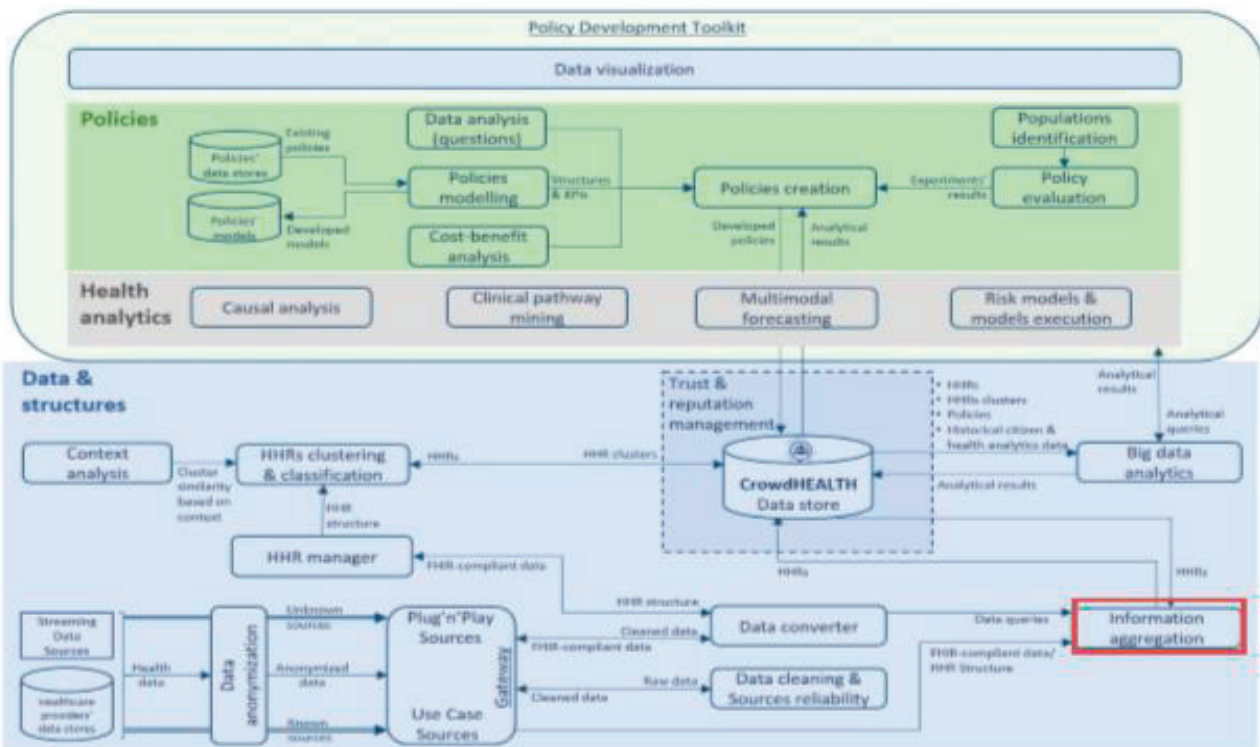


Figure 1. The Information Aggregation component in CrowdHEALTH architecture

simulations and decision making. The IA component is one of the main components of the CrowdHEALTH platform.

2. AIM

The purpose of this paper is to provide an overview of the CrowdHEALTH project and the technical architecture of the CrowdHEALTH platform in order to put the aforementioned IA mechanism in context. This is followed by the design details and initial specifications of the first prototype of IA component as well as its relationship with other components.

3. METHODS

The micro-services approach is used to perform information aggregation and to update HHRs in the CrowdHEALTH platform. Micro-services are a variant of the service-oriented architecture (SOA) where applications are structured as a collection of loosely coupled services with defined interfaces (5). In a micro-service based design approach, each service offers a specific functionality with relevant communication protocols as lightweight interfaces. Micro-services offer several benefits comparing to other application design approaches. The main benefit of the use of micro-services in the Information Aggregation component is the improved modularity of the component, which facilitates developing and testing the component in a distributed de-

velopment environment. The use of micro-service based design principles allow parallelizing the development of the Information Aggregation component in the CrowdHEALTH project. In terms of actual data aggregation operations, the use of Data Maps is being considered to integrate heterogeneous information in the HHRs. A Data Map is an instance of the process that involves creating data element mappings between two distinct data models. Data mapping can cover a wide variety of data integration tasks also including: Data transformation or data mediation between a data source and a destination. The Data Maps, developed using open source tools such as Talend Open Studio for Data Integration (6), can perform a specific aggregation operation based on the nature of the incoming data (7).

4. RESULTS

Within the CrowdHEALTH architecture, the Information Aggregation component is situated between the Interoperability Layer and CrowdHEALTH Datastore. The Information Aggregation component performs the final processing and aggregation of interoperable data, before it is aggregated in the HHRs and stored in the big datastore of CrowdHEALTH platform.

4.1 Information Aggregation Overall View

Information aggregation can happen in four ways within the CrowdHEALTH platform: The

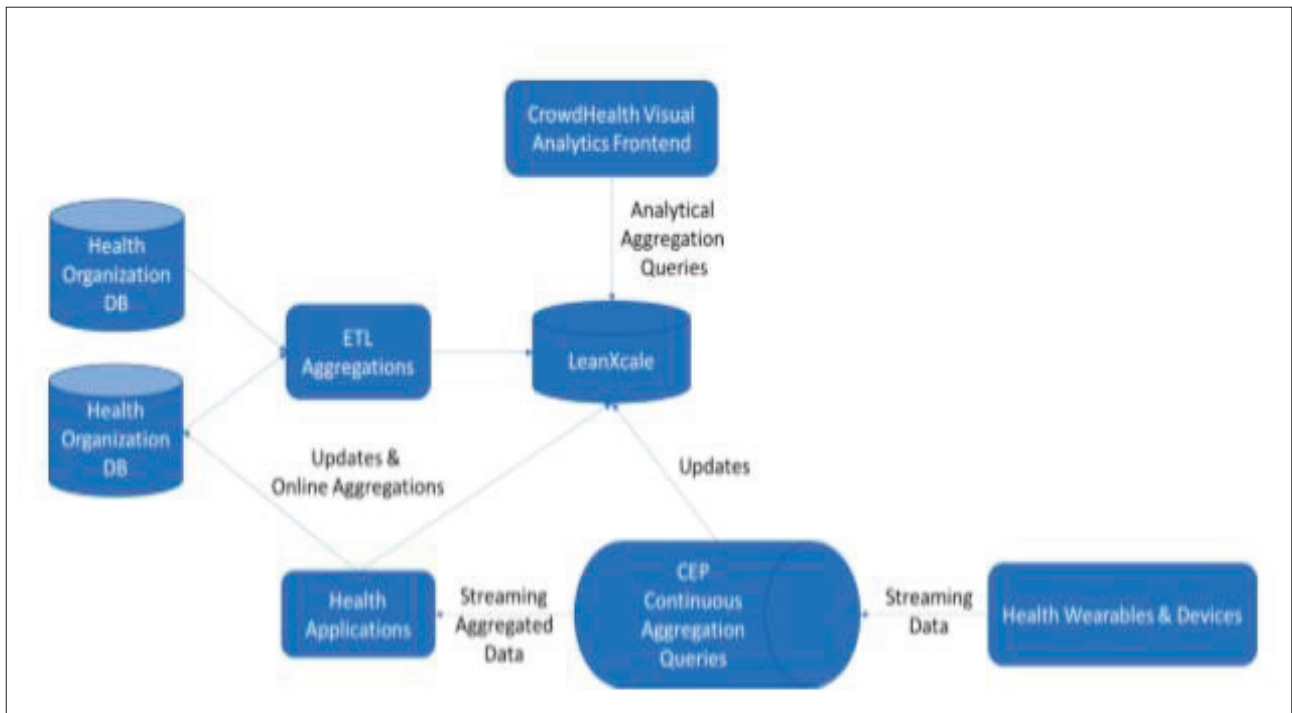


Figure 2. The four mechanisms to aggregate information in the CrowdHEALTH platform

data deriving from health systems (via internal components of CrowdHEALTH platform such as the Unified API component) can be aggregated into the HHRs with a micro-services based approach. The health systems or applications can add batch or streaming data to CrowdHEALTH Big Data platform. Aggregation analytic queries can update the existing data or HHRs within the CrowdHEALTH Big Data platform. High-frequency streaming data from wearables and other healthcare devices can be processed and stored within CrowdHEALTH using a Complex Event Processing (CEP) engine. Figure 2 presents the four mechanisms that can be used to aggregate information in the CrowdHEALTH platform. These mechanisms are designed after considering different information handling scenarios and the nature of data expected from the CrowdHEALTH use-case partners.

The data aggregation component is a central part of the architecture of CrowdHEALTH, it is responsible for merging the relevant information from different sources into an aggregated HHR per citizen. In the CrowdHEALTH platform, information aggregation can be obtained in a scalable, seamless and reliable manner using all the approaches illustrated in Figure 2. The aggregated information provides the basis for querying and analytic functions that serve as decision support tools for policymakers. For example, based on the availability of aggregated information in HHRs,

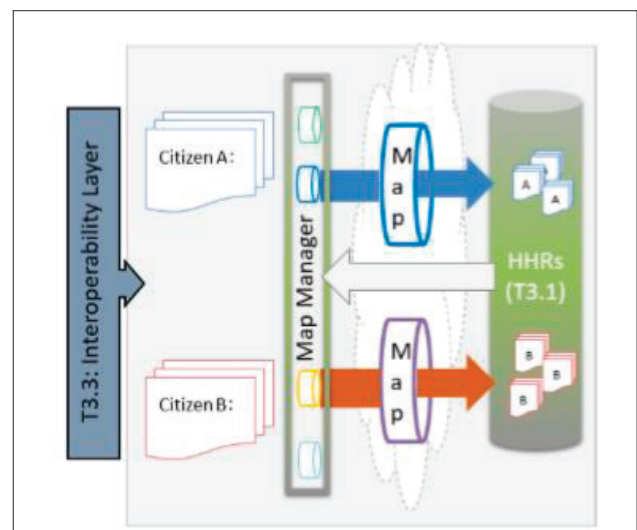


Figure 3. An overview of the data map

the policymakers may identify the number of patients for different illnesses at the level of the health center, district, city, region and country. Whenever a new patient record is added into the HHR, the underlying big datastore in the CrowdHEALTH platform increments the relevant HHR with a seamless database transaction and makes the latest information available for querying and analytic functions. In a regular database, this approach would cause many write-write conflicts causing most transactions to abort. In the CrowdHEALTH's LeanXscale datastore, the HHR aggregation functions implement an innovative semantic concurrency control that prevents these conflicts and aborts occurring. The advantage of this approach is that now computing an aggrega-

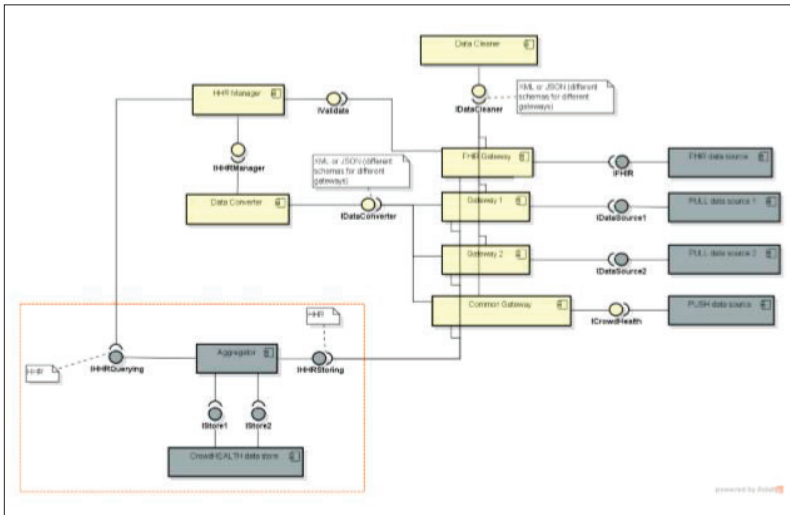


Figure 4. The flow of information to and from Information Aggregation component

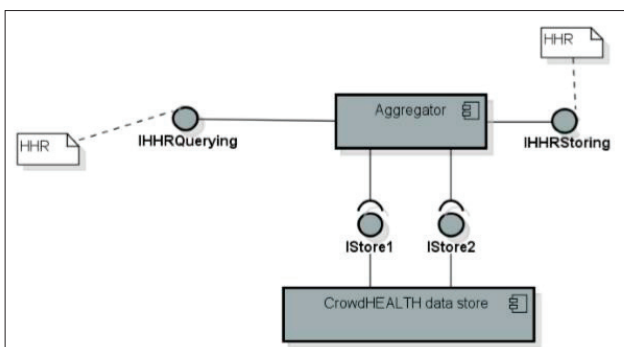


Figure 5. Interaction of the information Aggregation component with CrowdHEALTH data store

tion does simply imply to read a single row from the corresponding aggregation table. Among the many ways in which the LeanXscale datastore supports the information aggregation and retrieval function, one way is to leverage the OLAP (Online Analytical Processing) query engine to process arbitrary aggregation or analytical queries. In this way, aggregations can be obtained as and when needed within a reasonable response time. As shown in Figure 3, each Data Map can be specific to data type or data source, with the main function to receive the data, perform necessary transformation operations and aggregate the data in the relevant HHR.

Using a threaded implementation methodology, several Data Maps should be able to process multiple aggregation jobs simultaneously in order to deliver performance and scalability benefits while processing large volumes of data. The flow of information to and from the Information Ag-

gregation component is depicted in Figure 4 showing that the Information Aggregation component should be able to receive data from different data sources and aggregate the data to the existing HHRs stored in the CrowdHEALTH Datastore.

Figure 5 shows the relationships of the IA component with other components in the CrowdHEALTH architecture. It shows that the data enters the CrowdHEALTH platform from heterogeneous sources through the Gateways. From there, the data is cleaned and converted for use in the CrowdHEALTH platform.

The cleaned and converted data passes through the IA component before being stored in the CrowdHEALTH datastore.

The above figure (Figure 5) provides an overview of the information flow between the IA component and the final destination of the data in the CrowdHEALTH datastore.

4.2 Internal Mechanism of Micro-Services based Aggregation Approach

The internal mechanism of the Information Aggregation component breaks down the envisioned functionality into several sub-components. The sub-components are interlinked with interfaces (e.g. RESTful endpoints) (Figure 6) to allow interaction within and outside the Information Aggregation component. The overall Information Aggregation component consists of a series of interfaces, allowing other components and data sources (e.g. healthcare systems) to post updated information towards the CrowdHEALTH

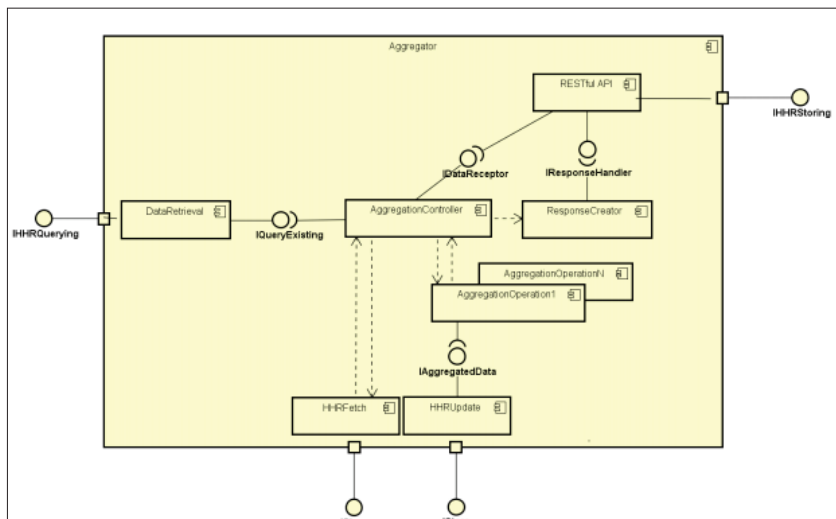


Figure 6. Internal architecture of the information aggregation component



datastore or for the HHR manager to check and retrieve existing data. It then has an interface to the CrowdHEALTH datastore for storage and retrieval of information to/from HHRs. Internally it has an Aggregation Controller subcomponent to select the appropriate action or micro-service to deal with incoming data. The aggregation functions are implemented as a series of micro-services that handle merging different types of existing data with incoming data in the HHRs.

4.3 Overall Functional Characterisation

The Information Aggregator is based on requirements derived from the CrowdHEALTH State of the Art and Requirements Analysis document [10] as well as from the internal architecture compiled during the project life cycle. The Information Aggregation component communicates by providing and accessing RESTful APIs, passing and receiving data in a standard agreed between different components of the CrowdHEALTH architecture. Internally aggregation operations are controlled by a sub-component called Aggregation Controller, which handles the internal flow of information within the Information Aggregation component. This controller hands off specific aggregation operations to individual aggregation

engines, implemented as micro-services.

4.4 Information Aggregation Component Workflow

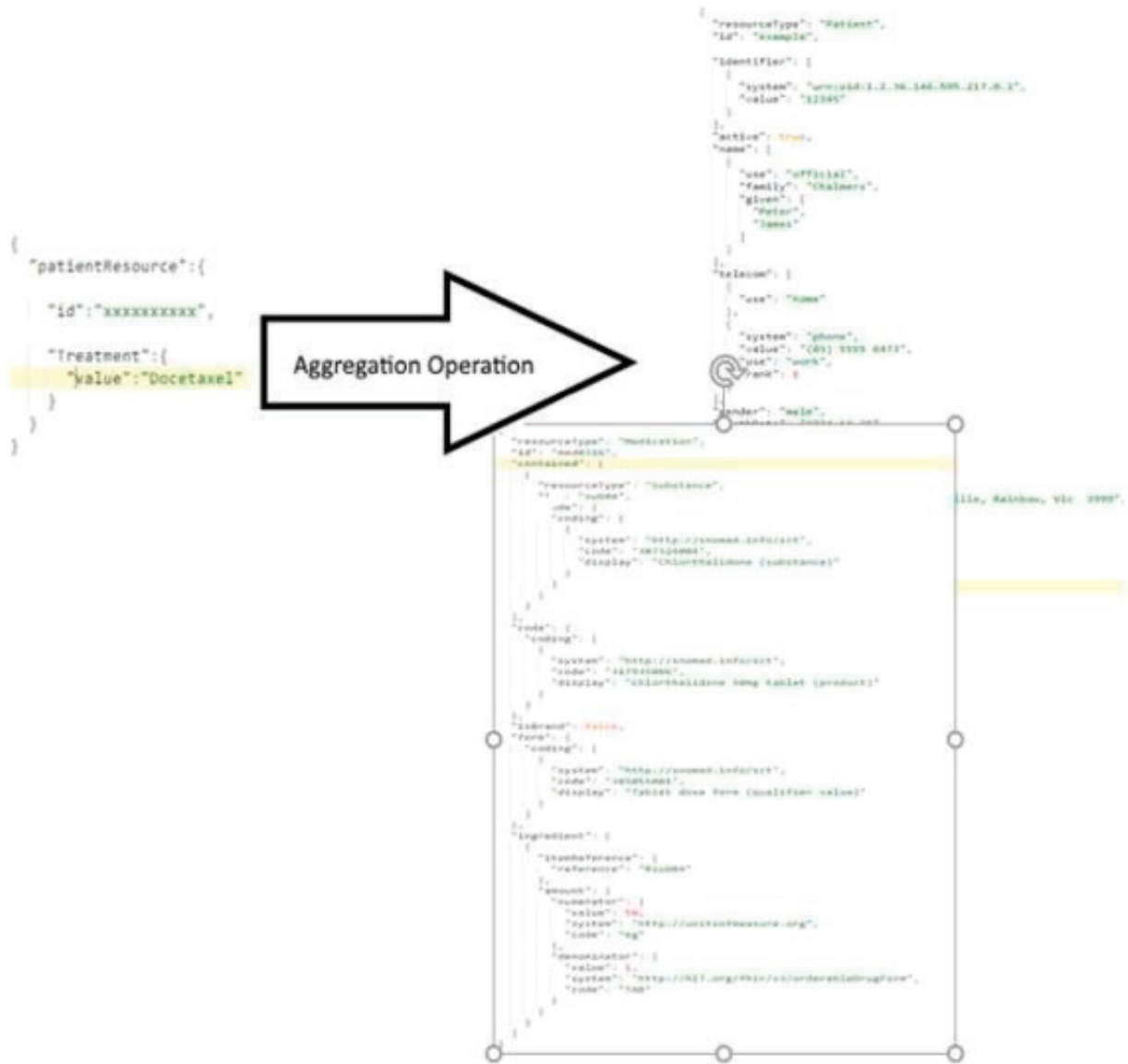
Receive Incoming Data from the Interoperability Layer

- Operation: Receive data from the Interoperability Layer through an open interface;
- Outcome: Data received by Information Aggregation component in JavaScript Object Notation (JSON) or Extensible Markup Language (XML) format – sample snippet of incoming data is shown below:

Retrieve Existing HHR from the Datastore

- Operation: Parse incoming data from the Interoperability Layer to extract the key identifier of the data.
- Operation: Connect to the underlying LeanX-scale Datastore and begin a transaction.
- Operation: Retrieve an existing HHR record (or tuple of information) related to the key identifier of incoming data – as shown in the following snippet.
- Outcome: Existing record or HHR of person/patient is identified in the CrowdHEALTH datastore.

Choose Relevant Instance of Information Ag-



gregator

The project will develop multiple instances of Aggregators capable of handling / processing different types of data e.g. batch, streaming etc.

- Operation: Design and implementation of Data Maps that can aggregate incoming data into the existing HHR while preserving the integrity of the aggregated HHR.
- Operation: Aggregation Controller selects a relevant instance of the Aggregator, based on the analysis of following factors:
 - o Incoming data type.
 - o Existing data type.
 - o Available resources or aggregators.

The selection of an Aggregator is performed using a scoring mechanism. The Aggregator Controller analyses the incoming datasets and assigns a score to each incoming dataset that contains a specific type of data. The score signifies how well the incoming dataset matched with an

existing dataset. In this respect, the scoring of datasets can be seen as a part of a sort of pipeline specification that maps the incoming and already existing datasets. The outcome of the scoring mechanism provides information on the processing of the incoming dataset using a specific Aggregator.

- Operation: Expose and access Aggregators as micro-services with REST interfaces.
- Outcome: Relevant Aggregator is selected and the aggregation operation is launched.

Aggregation Operation (Simple)

- Operation: Execute a Data Map to perform an aggregation operation e.g. add a tuple of information into an existing HHR.
- Operation: Commit a transaction in the Datastore.
- Outcome: Aggregated record updated in the datastore.

Aggregation (Complex)

- Operation: Execute a relevant Data Map to perform Aggregation into multiple tables. This may include the creation or updating of multiple records/tables.

- Operation: Commit a transaction in Datastore.

- Outcome: Aggregated record updated in the datastore.

5. CONCLUSION

The Information Aggregation component ensures incoming data is correctly aggregated with existing data to develop Holistic Health Records. The Information Aggregation component manages health-related data coming from various sources (systems, devices) and different velocities (stream or batch). The component includes a number of aggregation operations to perform specific types of aggregation functions, depending on the nature and type of incoming data. It also provides an API and set of interfaces to communicate with data sources. The implementation of aggregation functions is based on the micro-services approach and also on the enhanced aggregation functionality offered by the underlying LeanXcale datastore. The aggregation functions will be encoded using big data management techniques, with enhancing the state of the art in specific areas such as the use of micro-services to perform synchronous aggregation operations on heterogeneous datasets or the use of an ultra-scalable, polyglot and full ACID big-data platform to enable efficiency in data handling, aggregation and analytic operations. Although this is an initial version of the IA component, the specifications and implementation level details will be further updated during the project's course.

- **Acknowledgements:** CrowdHEALTH project is co-funded by the Horizon 2020 Programme of the European Commission Grant Agreement number: 727560 – Collective wisdom driving public health policies.
- **Author's contribution:** Each author gave substantial contribu-

tion in acquisition, analysis and data interpretation. Each author had a part in preparing article for drafting and revising it critically for important intellectual content. Each author gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

- **Conflict of interest:** None declared.

REFERENCES

1. Mena LJ, Felix VG, Ostos R, Gonzalez JA, Cervantes A, Ochoa A. et al. Mobile personal health system for ambulatory blood pressure monitoring. *Computational and mathematical methods in medicine*. 2013; 13. doi: 10.1155/2013/598196.
2. Chawla V, Davis DA. Bringing Big Data to Personalized Healthcare: A Patient-centered Framework. *Journal of General Internal Medicine*. 2013; 28(3): 660-665. doi: 10.1007/s11606-013-2455-8
3. Kyriazis D, Autexier S, Brondino I, Boniface M, Donat L, Engen V. et al. CrowdHEALTH: Holistic Health Records and Big Data Analytics for Health Policy Making and Personalized Health, *Stud Health Technol Inform*. 2017; 238: 19-23. doi: 10.3233/978-161499-781-8-19
4. Montandon L, Kyriazis D, Ramon V Z, Llatas F, Traver V. CrowdHEALTH- Collective wisdom driving public health policies. *IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. 2019; 1-3. doi:10.1109/CBMS.2019.00010.
5. Newman S. *Building Microservices: Designing Fine-Grained Systems*. O'Reilly Media, Inc. 2015; Available at: <http://ce.sharif.edu/courses/96-97/1/ce924-1/resources/root/Books/building-microservices-designing-fine-grained-systems.pdf>
6. Talend Data Mapper–User Guide. Available at: <https://help.talend.com/reader/8P5AdZmUvft63yHLVqkRsw/zOczPWm3KuUqQw78e5doMw>
7. Mantas J. Future trends in Health Informatics - theoretical and practical. *Studies in health technology and informatics*. 2004; 109: 114-127.