Sustainable Cities and Society 39 (2018) 68–77

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Privacy-aware Big Data Analytics as a service for public health policies in smart cities

Marco Anisetti[a], Claudio Ardagna[a], Valerio Bellandi[a], Marco Cremonini[a], Fulvio Frati[a], Ernesto Damiani[b]

[a] Università Degli Studi di Milano, Via Bramante 65, Crema, Italy
[b] ETISALAT BT Innovation Center, Khalifa University, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:
Big Data
Privacy
Public policy making

ABSTRACT

Smart cities make use of a variety of technologies, protocols, and devices to support and improve the quality of everyday activities of their inhabitants. An important aspect for the development of smart cities are innovative public policies, represented by requirements, actions, and plans aimed at reaching a specific goal for improving the society's welfare. With the advent of Big Data, the definition of such policies could be improved and reach an unprecedented effectiveness on several dimensions, e.g., social or economic. On the other hand, however, the safeguard of the privacy of its citizens is part of the quality of life of a smart city. In this paper, we focus on balancing quality of life and privacy protection in smart cities by providing a new Big Data-assisted public policy making process implementing privacy-by-design. The proposed approach is based on a Big Data Analytics as a Service approach, which is driven by a Privacy Compliance Assessment derived from the European Union's GDPR, and discussed in the context of a public health policy making process.

1. Introduction

The recent diffusion of smart and powerful devices and sensors, wireless communication technologies and protocols, and Big Data technologies is paving the way to a new era of smart cities aiming to improve the quality of life of their citizens. Health is obviously one of the key aspects impacting the quality of life, especially in metropolitan areas (Boulos & Al-Shorbaji, 2014). In this context, smart cities can provide fundamental elements for a healthier environment and for improved well-being of city dwellers. The increasing diffusion of health-related wearable devices and the availability of IT interconnections between hospitals, supporting patient's medical data exchange, foster the development of a number of health-related services for smart cities.

In this scenario, several smart city projects focusing on healthcare have been funded, specifically regarding the continuous monitoring of the health condition of patients living at home. For instance, the SPHERE project (Sensor Platform for HEalthcare in a Residential Environment, 2013–2018)[1] adopted by Bristol City Council, UK, focus exactly on health monitoring. This trend of development of health-related services for smart cities also reflects the mission of the World Health Organization (WHO) Healthy Cities project, which started in 1987 and is currently at Phase V.[2] Starting from Phase IV, the importance of local public health policies, supported by evidence, has been clearly pointed out as a mandatory phase towards the fulfillment of the WHO long-term Healthy Cities project. Evidence is not just related to city dwellers' personal health data (e.g., medical data from hospitals and from wearable devices/sensors), but also includes other sources integrating the social dimension in the analysis of the personal health status. These aspects are considered of paramount importance also by the European Commission in the framework of the H2020 program.[3] The European Commission underlined the importance to include ethical aspects of data, the confidentiality and anonymity of data transfer, and the engagement of those collecting such data in its analysis and interpretation, to avoid misinterpretation and inappropriate conclusions.

A common practice to healthcare management is based on public health policies, which typically define a set of actions aimed at improving some key indicators of public utility. However, the policy making policy comes with some non-negligible requirements, which

E-mail addresses: marco.anisetti@unimi.it (M. Anisetti), claudio.ardagna@unimi.it (C. Ardagna), valerio.bellandi@unimi.it (V. Bellandi), marco.cremonini@unimi.it (M. Cremonini), fulvio.frati@unimi.it (F. Frati), ernesto.damiani@kustar.ac.ae (E. Damiani).

1 http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/K031910/1.
2 http://www.euro.who.int/en/health-topics/environment-and-health/urban-health/activities.
3 http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/calls/h2020-sc1-2016-2017.html.

https://doi.org/10.1016/j.scs.2017.12.019
Received 31 August 2017; Received in revised form 15 November 2017; Accepted 13 December 2017
Available online 06 February 2018
2210-6707/

call for careful consideration. First of all, public health policy definition requires a multi-step and iterative process that analyzes Big Data to tune features and thresholds/values in the policy. Moreover, management of private health data introduces strong privacy requirements that affect both policy definition and Big Data Analytics. This is made even more relevant by the advent of the new General Data Protection Regulation (GDPR) in EU (European Parliament, 2016), which provides stronger and more precisely defined requirements in terms of privacy.

Our work aims to suggest an improved way to develop public health policies for smart cities, by adopting an approach based on Big Data and supporting the design of privacy-compliant policies. This work extends the work of the EVOTION H2020 project,[4] which is focused on the definition of public health policies related to hearing loss (Anisetti, Bellandi, Cremonini, Damiani, & Maggesi, 2017), to a general public health scenario in the context of smart cities. More specifically, it extends a typical public health policy making process towards a Big Data-assisted process, driven by a privacy-aware analysis and processing of dwellers' personal data. Our approach also adopts TOREADOR Big Data Analytics framework (Ardagna et al., 2017) to support policy making and extend it to include a semi-automated compliance assessment of GDPR requirements. The TOREADOR framework, in fact, permits to handle privacy in an holistic and transparent way, as generic requirements associated with specific process steps. It is offered as a service, satisfying the fundamental need of simplicity of public policy makers, which are usually domain experts but often lack data science experience.

The contribution of this paper is therefore threefold: (i) to formalize a Privacy Compliance Assessment based on GDPR privacy requirements for public health policy making process, (ii) to improve on traditional public policy making processes towards Big Data-assisted public policy making, (iii) to define a privacy-aware Big Data campaign supporting policy makers in the definition of privacy-aware public policies. This work develops on our previous works (Anisetti et al., 2017; Ardagna et al., 2017) by providing (i) a refined definition of the Big Data-assisted public policy making process for smart cities, (ii) privacy-aware Big Data Analytics supporting the public policy making process, and (iii) a concrete scenario of application and a case study.

The remainder of the paper is organized as follows. After presenting some basic concepts including a public health policy making process and a Big Data engine in Section 2, Section 3 presents our Big Data-assisted public policy making process. In Section 4, we give an overview of Privacy Compliance Assessment based on the GDPR. Section 5 extends the Big Data-assisted process to include the Privacy Compliance Assessment, while Section 6 presents the related work. Finally, Section 7 gives our concluding remarks.

2. Basic concepts

In this section we describe two basic concepts underpinning this paper: (i) how a evidence-based public health policy making process can be modeled, and (ii) how a Big Data engine supporting the policy making process is structured.

2.1. Public health policy making process

According to the World Health Organization (WHO): "*Health policy refers to decisions, plans, and actions that are undertaken to achieve specific health care goals within a society*." A health policy is qualified as public if it is made by public institutions for large groups of persons at regional, national or even international level. Loukis (2007) define an ontology for a generic public policy workflow and eight stages of a policy definition, each one with specific objectives and a corresponding sub-ontology. The approach in Loukis (2007) is generic but can be adapted to health policy generation based on evidence. This evidence supports an analyst in evaluating the policy effectiveness. An evidence is traditionally captured through clinical trials or epidemiological studies. According to Dunn (2015), a number of traditional methods are also adopted to assist the evaluation, like pseudo-evaluation methods, formal evaluation methods, and decision theoretic evaluation (e.g., graphic displays, tabular displays, index numbers, interrupted time-series analysis, control-series analysis, regression-discontinuity analysis) In this paper we consider a simplified evidence-based policy making process as a sequence of subsequent refinements based on three stages: *Situation analysis*, *Action plan*, and *Implementation evaluation, and monitoring*. The approach is derived from a simplification of Loukis (2007), where we merge stages and underline the evolution of the policy to be released through the subsequent stages:

- *Situation analysis* assesses the needs and gaps arising in connection with a situation that should be addressed by health policy. It implements an initial draft of the policy.
- *Action plan* sets the initial goals and activities, identifies the resources, and iteratively refines them. It iteratively converges to the final policy.
- *Implementation, evaluation, and monitoring* defines and evaluates recommendations, turns them to prescriptions, if needed, and monitors the whole policy life cycle.

This process is human-centric and driven by experts in the field. All decisions are taken based on longitudinal studies and literature, and often is region- or nation-wide. The intrinsic complexity and human-centric nature of the process are not well-suited for smart city environments, which are typically organized around dynamical and technology-oriented processes. Fig. 1 shows how this process is modeled. Situation analysis initializes the policy. It is executed once, ideally. Action plan is an iterative process aimed at refining the draft policy provided by the Situation analysis stage. Each iteration may take a great amount of time and may require the completion of a longitudinal study. Action plan can trigger additional Situation analyses if a refinement of the draft policy structure is needed (e.g., the goal is not addressable with the given structure of the policy). Implementation, evaluation, and monitoring step is aimed at evaluating the policy while in operation. It can trigger additional Action plan refinements if the policy presents any discrepancy with respect to the expectations (e.g., no positive effect on the population).

2.2. Big Data platform

A Big Data platform can be profitably used to support the policy making process described before, permitting the experts to define and execute analytic tasks. In general it allows three types of interaction: (i) inspect data using a simple query, (ii) analytics to execute processing task involving, for instance, machine learning approaches, (iii) open existing projects and monitor their status. We note that simple queries can be seen as an analytic task where no processing is requested. Fig. 2 shows an architectural view of our Big Data Engine. It is an abstract view divided into four layers: *Visualization*, *Data Processing*, *Data Acquisition*, and *Management*.

Visualization layer is the frontend for policy makers for the definition of analytic tasks to be executed in order to take a decision on a certain policy.

Data Processing layer is the core of the architecture. It is responsible to process data coming from Data Acquisition layer. It receives as input analytic task and configuration files to be evaluated, and gives as output feedback to policy makers allowing them to tune the policy. Our analytic pipeline specifies all processing activities needed as a set of consecutive steps:

- *Data Preparation* represents the operations to be performed on data

[4] http://h2020evotion.eu/.

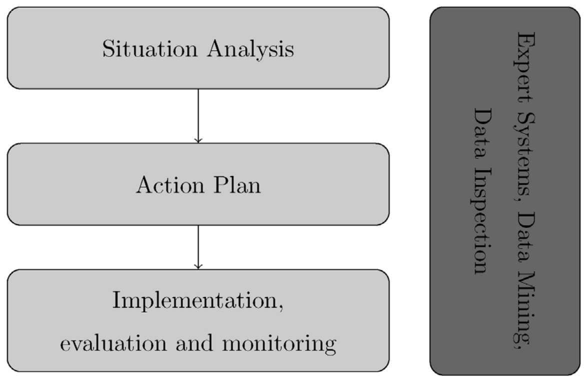

Fig. 1. Public health policy making process.

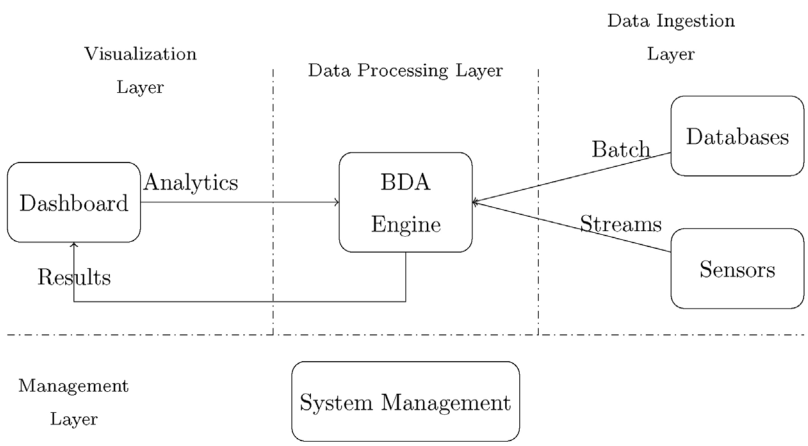

Fig. 2. Big Data platform supporting policy making process.

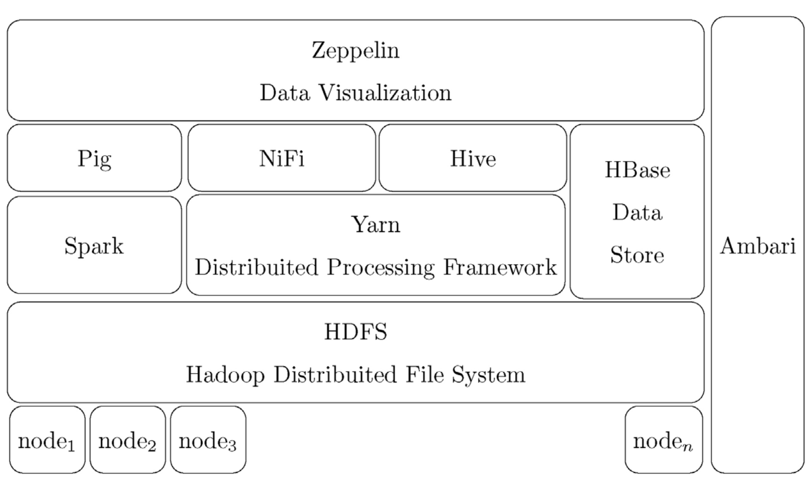

Fig. 3. Data platform technologies.

before the elaboration. It defines how to guarantee data owner privacy using anonymization (e.g., hashing, obfuscation, differential privacy, *k*-anonymity).

- *Data Representation* specifies how data are represented and expresses representation choices for each analysis process. For instance, it defines the data model (e.g., document-oriented, graph-based, relational column-oriented, extended relational, key value) and partitioning (e.g., clustering, sharding, memory caches).
- *Features Reduction/Selection* describes the features selection from the dataset. For instance, the most significant factors with respect to a specific objective are selected.
- *Data Analytics* specifies how data are represented and expresses representation choices for each analysis process. For instance, it defines the data model (e.g., document-oriented, graph-based, relational column-oriented, extended relational, key value) and partitioning (e.g., clustering, sharding, memory caches).
- *Data Reporting and Visualization* is the activity in charge of arranging results for visual analysis, **offering them to the visualization layer in a structured form.**

We note that there are dependencies between each stage of a pipeline, for instance Data Preparation depends on the classification algorithm adopted (e.g., transformation to numbers), features selection represents a constraint for the preparation.

Data Ingestion layer is the layer where data are ingested depending on the needs. Data sources, in general, include screening/monitoring activities (e.g., through sensors), or clinical databases. These types of data are critical and need to be managed in accordance with national and international laws and regulations.

Management layer is the layer that permits to manage the cluster (composed by n nodes). It is responsible to allocate resources and permits to administrate the hardware and the software services. It is responsible to secure communication between BDA layers and if needed to implement specific privacy related deployment for instance implementing storing or processing tenant isolation.

Nowadays, there are a number of Big Data frameworks covering all functionalities of the above layers; some of them are offered as a service in the cloud as commercial products. In this paper, we adopt the open source Apache Foundation framework, which is composed of a number of tools/services that constitute a complete and powerful ecosystem. Fig. 3 shows the main components of our Big Data platform. Even if most of the tools/services that we adopted are able to cover functionalities required by more than one layer, we installed and configured them to exploit their peculiarities specifically related to one platform layer only (mentioned in square brackets below).

Hadoop–YARN [Data Ingestion layer] Hadoop is a tool for data-intensive distributed applications, based on YARN programming model and a distributed file system called Hadoop Distributed Filesystem (HDFS). Hadoop allows writing applications that rapidly process large amounts of data on large clusters of compute nodes. A YARN tool permits to divide the input dataset into independent subsets that are processed in parallel.

HBase [Data Ingestion layer] A database engine built on Hadoop and modeled after Google's Big Table. HBase is optimized for real-time data access to large tables in the billions of rows. Among other features, it offers support for interfacing Hive. The Pig API features a storage function for loading data from an HBase data base.

NiFI [Data Ingestion layer] It is an integrated data logistics platform for automating the movement of data between disparate systems. Apache NiFI provides real-time control that permits to manage the movement of data between different type of source and different type of destination. It support disparate and distributed sources of differing formats, schemas, protocols, speeds and sizes such as machines, geolocation devices, click streams, files, log files etc.

Hive [Data Processing layer] A data warehousing infrastructure, which runs on top of Hadoop. Hive provides a language called Hive QL to organize, aggregate, and run queries on dataset. Hive QL is similar to SQL, it uses a declarative programming model and results are described in one big query. HQL queries can be broken down by Hive to communicate to MapReduce jobs executed across a Hadoop cluster.

Spark [Data Processing layer] A general purpose cluster computing engine providing APIs to various programming languages such as Java, Python, or Scala. Spark is specialized at making data analysis faster, supports in-memory computing that enables it to query data much faster compared to disk-based engines such as Hadoop, and also it offers a general execution model that can optimize arbitrary operator graph. Generally speaking, Spark is advance and highly capable upgrade to Hadoop aimed at enhancing Hadoop ability of cutting edge analysis. Spark also offer several tools, such as machine learning tool M-Lib, structured data processing, Spark SQL, graph processing tool Graph X, stream processing engine called Spark Streaming, and Shark for fast interactive question device.

Pig [Data Processing layer] It is a scripting platform for processing and analyzing large data sets. Apache Pig permits to write complex MapReduce transformations using a scripting language called Pig Latin. Pig translates the Pig Latin script into MapReduce so that it can be executed within YARN for access to a single dataset stored in the (HDFS).

Zeppelin [Data Visualization layer] A web-based and multi-purpose notebook that enables interactive data analytics. The notebook is the place for data ingestion, discovery, analytics, visualization and collaboration. It can make data driven, interactive and collaborative documents with Scala and more. Apache Zeppelin Interpreter allows any language/data-processing-backend to be plugged into Zeppelin.

Ambari [Management layer] It is a management platform for provisioning, managing, monitoring and securing Apache Hadoop clusters. Apache Ambari takes the guesswork out of operating Hadoop.

3. The Big Data-assisted public policy making process

Our Big Data-assisted public policy making process extends the process in Section 2.1 by (i) better detailing the three stages of policy definition, (ii) formally representing the policy to allow semi-automatic processing, (iii) integrating a Model-Based Big Data Analytics as a Service (MBDAaaS) defined for the platform in Section 2.2 to allow model-driven definition of Big Data campaigns that collect evidences for the definition of a policy.

A formal description of the policy makes it readable and processable, and support the semi-automatic triggering of Big Data campaigns without ambiguity. Compared to a traditional public policy making process, the adoption of MBDAaaS allows to (i) do quicker iterations on each stage of the policy making process to inspect multiple solutions or refine them, (ii) produce fine-grained models that support a better holistic view needed for smart cities' public health policies.

Our Big Data-assisted public policy making process works on two set of data. On one side, *retrospective data* collected prior to the definition of the policy, useful for simulations and predictions. On the other hand, *perspective data* collected while policy is in place; they are used to evaluate the effect of the policy after its adoption. We assume that relevant data needed for reaching a decision is available in the *retrospective data set*, and that this data set is comparable to those adopted in longitudinal studies.

In the following, we first introduce our formal representation of a given public health policy considering a relevant scenario. We then describe our Big Data-assisted public policy making process. Our process extends the one described in Section 2.1. Then we finally introduce the Model-Based Big Data Analytics as a Service (MBDAaaS) to define and execute a Big Data campaign required by the policy making process.

3.1. Public health policy

The output of the public policy making process of Section 2.1 is a policy described in a natural but rigorous language. We show how to gradually transform it into a formal model instance of the policy, which will be then adopted in the rest of the paper. Let us consider the following example as the output of the policy making process of Section 2.1, adapted for presentation purpose from studies on hospital readmissions of patients (Dungan, 2012; Dungan et al., 2014; Rubin, 2015):

Example 1 (*Natural language*). It is recommended for patients with diabetes having more than *65 years* of age, living *alone* in *deprived areas* of the city, and already hospitalized at least *twice* to receive a post-discharge visit by a diabetes nurse *every month* with the aim of *reducing the readmission rate by K%*.

It is well known that living in metropolitan areas increase the risk of diabetic disease (Health City Institute – Italian Barometer Diabetes Observatory (Ibdo) Foundation) due to stress, wrong habits and societal frictions dictated by the city life, to name but a few. In turn, the specific area within the same city (i.e., affluent or deprived areas), as well as the family conditions (e.g., living alone), may have a sensible impact on the evolution of the disease (Cities Changing Diabetes – Atlas 2017). In the example, the age and the area of living are well-known risk factors for readmission (Rubin, 2015). The condition on the number of hospitalizations discriminates patients below or above a threshold of abnormal number of readmissions. Example 1 can be rewritten in a more abstract form by referring to predicates, which in turn refer to features and corresponding values.

Example 2 (*Parametric form*). It is recommended for people with more than < *age* > and living < *living style* > and in city < *type of the area* > and already hospitalized < *number of readmissions* > to receive a post-discharge visit by a diabetes nurse < *nurse visit frequency* > for < *readmission rate reduction* > . Where:

- *normative feature*: "age", "living style", "type of area", and "number of readmission"; *normative value*: "65", "alone", "suburbans deprived area", and "twice";

- *objective feature*: "nurse visit frequency"; *objective value*: "once per month";

- *goal feature*: "readmission rate reduction"; *goal value*: "K%".

With respect to the definition of predicates, goal predicates are typically defined at situation analysis stage, for both features and values, because they represent the outcome that makes the policy effective, either in terms of economic sustainability or improved welfare. Objective predicates depend both on constraints defined in the situation analysis stage, for example technical or economic constraints, and possibly on analysis carried out during the action plan stage. Finally, normative predicates mostly depend on the analysis and correlations carried out during the action plan. Furthermore, to fully formalize the policy, in this paper, we consider a simplified *deontic logic* form (Balbiani, 2008), which let us express concepts like "recommended" or "obligatory".

Definition 1 (*Deontic logic-based Policy***).** Let us consider the following policy:

$$\theta = \begin{cases} \text{Policy:} & P_1 \rightarrow M(P_2) \\ \text{Goal:} & P_3 \end{cases}$$

It defines a policy where P_1, P_2, P_3 are first order logic preposition expressed in terms of normative, objective, and goal predicates, respectively, M is a modal operator $\in\{O, R\}$, where O express obligation and R express recommendation (Balbiani, 2008).

The expressive power of the deontic logic form is lower than natural language, but still sufficient for many applications and, being machine readable, fully exploitable by our Big Data-assisted policy making engine. In the following, we consider public health policies defined using the deontic logic form in Definition 1.

Before proceeding to discuss the typical stages of a public policy making process – situation analysis, action plan, implementation, evaluation, and monitoring – it should be noted that policy makers typically need assistance by several domain experts. Clinicians, data analysts, and data processors are likely to be involved in the process. Domain experts should assist policy makers in the definition of the health policy by bringing their experience in: medicine (e.g., helping in the definition of normative features tightly dependent on a pathology), data analysis (e.g., helping policy makers in the definition of analytic tasks), and data processing (e.g., operating the technical infrastructure and executing analytics and policy simulations). For instance, in *situation analysis*, clinicians could be heavily involved in supporting policy makers, while data analysts and data processors are limited at basic assistance, like in the definition and execution of queries. In *action plan and implementation, evaluation, and monitoring*, the role of domain experts changes. In action plan and implementation, data analysts and data processors are likely more active than clinicians in supporting policy makers, especially when advanced analytics need to be defined and executed. Clinicians become again important in evaluation and monitoring of the policy outcome, real or simulated, when the effectiveness of the policy with respect to the goal is considered.

The same differences between actors in the different stages of the policy making processes may suggest different types of data access, in particular considering the strong privacy requirements that health data require. Data access can be described in terms of privacy level and scope:

- *Policy maker*: a policy maker is typically interested in aggregating data from all data sources and feeds. Therefore, he/she wants to have the broader view, but no information on specific individuals (scope: all data; privacy: anonymized).
- *Clinician*: typically, a clinician, before processing, has access to patients individual information, for instance, those cared by him/her or by his/her hospital (scope: limited; privacy: not anonymized). After processing, he/she might have access to aggregate data to support policy makers (scope: all data if permitted by policy makers; privacy: anonymized).
- *Data Analyst*: a data analyst, in general, does not need access to any individual information and, needs to aggregate data only if a policy maker requires it. A data analyst should know data types, relations, structure, and data analytic techniques (scope: all data if permitted by policy makers; privacy: anonymized).
- *Data Processor*: a data processor may have full access to individual data if in charge of providing data anonymization (scope: limited; privacy: not anonymized) and to aggregate data for supporting the execution of analytics (scope: all data; privacy: anonymized).

3.2. Situation analysis

This stage is mostly concerned with policy makers and domain experts (e.g., clinicians, data analysts, data processors), which specify the objectives and constraints of a public health policy. Policy makers decide which type of intervention should be needed (e.g., periodic screening, home medical assistance, education, lifestyle monitoring) and the objective of the intervention (e.g., frequency of periodical visits, medication, or medical tests).

The first phase, Situation analysis, takes as input all available data already collected (retrospective data) and returns as output a draft version of the policy in a form that will be further refined in subsequent stages. Big Data campaigns are adopted mainly to execute simple querying on the retrospective data to extract basic knowledge for setting up the draft policy. Policy makers and domain experts can do the following activities:

- *Data Exploration*: Querying the available retrospective data to retrieve basic information on their characteristics, for instance, the list of available features, the average age of all patients in the data set, the percentage of the patients wearing specific sensors, to name but a few. It is useful to have an idea on the ranges of each available data in the data set.[5]
- *Data Crop (optional)*: Crop retrospective data deciding which are the data used for the action plan stage. For instance, given a specific time frame the data can be cropped accordingly. The rest of the data can be just removed from the process or associated with the perspective data, if needed.
- *Draft Policy Initialization*: Given the data exploration achievements, the policy makers can define the draft policy where at least the objective and goal are identified. The structure of the policy can also be defined, as well as some of the normative features and values, if needed. This occurs when experts want to drive the definition of the policy since the beginning. For instance, considering Example 2, an expert can fix the normative features "age" at 65, since it is a threshold well-known from literature.

Example 3 (*Draft Public Health Policy***).** Considering Example 2 and the formal Definition 1, the draft Public Health Policy, with patient p belonging to the population P of hospitalized diabetes patients, can be defined as follows:

[5] At this stage it is possible to evaluate whether the retrospective data are representative enough for the situation under analysis.

$$\vartheta = \begin{cases} \text{Policy:} & \forall \ \ p \in P\text{:} \ \ a_1 \odot \ a_2 \odot \cdots a_n \\ & \rightarrow R(\text{assistanceFreq}(p, x)) \\ \text{Goal:} & \text{readmissionReduct}(P, K) \end{cases}$$

with

- $\odot$ as a logic operator $\in \{\wedge, \vee, \neg\}$;
- R(*assistanceFreq*(p, x)) the modal operator expressing the recommendation applied to the predicate representing the diabetes nurse visiting patient p with frequency x;
- *readmissionReduct*(P, K) the predicate representing the reduction of K% in readmission rate of population P.

In the draft version of the public health policy, we see the goal predicate *readmissionReduct*(P, K) specified with respect to the reference population P and a certain (minimum) outcome K%, reflecting the general goal of improving social welfare and hospital management. The objective predicate *assistanceFreq*(p, x) is defined with respect to each patient p selected for obtaining assistance and with value x representing the intensity, possibly variable, of the support offered by the health policy. Normative predicates are not necessarily specified at this stage and will be subject to action plan stage.

Situation analysis can be re-executed in case of problems at implementation evaluation and monitoring stage, or Action plan stage. In that case, the feedback for the following phases are used to change the draft policy in terms of goal or structure.

3.3. Action plan

This stage is mostly focused on tuning the draft policy by evaluating its efficacy on the (cropped) retrospective data. This is an iterative phase with policy makers, assisted by data analysts and data processors, asking to Big Data platform multiple executions of analytics on available data until a satisfying result is obtained. The scope is to fulfill the policy defining all predicates and parameters. Action plan takes as input the retrospective data selected at Situation analysis stage and returns as output a final version of the policy in a form that can be verified by the subsequent Implementation, evaluation and monitoring stage. The Big Data platform is adopted to do different analytic tasks aimed at finalizing the policy. Policy makers and clinicians can do the following activities:

- *Data Partitioning*: The Big Data platform supports policy makers in defining a suitable partitioning of retrospective data and inspecting the results. For instance, data could be grouped according to personal traits like ethnicity (a well-known risk factor for diabetes) or re-hospitalization history. Statistical analysis can also be performed to evaluate the representativeness of data partitioning (e.g., PCA analysis). Partitioning also permits to separate training set from test set if needed by the analytics. We note that at situation analysis stage the data set is cropped, while in this case the data set is divided into portions useful for different analysis. This stage include Data Preparation, if needed.
- *Feature Selection*: A set of relevant n features,[6] with respect to the policy goal, should be selected on the partitioned retrospective data by means of an analytic approach (e.g., Rough Set). Iteratively, policy makers evaluate the n features and decide whether to further modify them, change the analytic algorithm, or accept the setting.
- *Values Selection*: Given the selection of normative features the relative normative values should be defined. In this case, a suitable classification approach should be selected (e.g., regression, clustering, random forest). With normative features and values, portion of the retrospective data can be processed as a training set. The result is a classification that policy makers should evaluate with respect to the policy goal. The decision could be to adopt a different classification approach, modify the normative value selection, or confirm both values and classification approach. Normative features and values could be further evaluated by policy makers, after the definition of normative predicates. If needed the policy maker can restart the process from the data Partitioning phase.
- *Voting*: Given a set of different evaluations in parallel, the final results for each of them is presented to the policy makers. The policy maker can decide to choose one of the results or let Big Data engine do a voting system to decide which are the most suitable set of values depending on the approaches adopted. Our engine supports a number of different voting approaches, from the simple majority to prioritized/weighted majority and fusion.
- *Final Policy Definition*: Thanks to the analysis on the data the policy makers can finalize the Public Health Policy, that it is then passed to the next stage. The policy makers can also decide that the original goal and objective are not well defined and then instead of refining the final policy, re-execute the situation analysis with a modified draft policy.

Example 4 (*Refined Public Health Policy***)**. Considering Examples 2 and 3, one possible refined policy θ is as follows.

$$\theta = \begin{cases} \text{Policy:} & \forall \ \ p \in P\text{:} \ \ \text{ageGT}(p, 65) \wedge \\ & \text{Lifestyle}(p, \text{alone}) \wedge \text{LivingArea}(p, \text{suburbs}) \\ & \quad \wedge \ \text{hospitalGE}(p, 2) \\ & \rightarrow R(\text{assistanceFreq}(p, \text{month})) \\ \text{Goal:} & \text{readmissionReduct}(P, K) \end{cases}$$

with predicate *ageGT* meaning the age greater than a value, *Lifestyle* for the lifestyle of a specific type, *LivingArea* for live in a particular area, and *hospitalGE* for the number of hospitalizations greater or equal to a value.

We note that Big Data improves the action plan stage by allowing the execution of multiple analytics in a given time frame; in addition, voting approach permits to execute parallel analytic evaluations and combines them to have a holistic view made by different predictors, statistical evaluation or classifications.

3.4. Implementation, evaluation and monitoring

This stage is focused on deploying the policy in production and monitoring how it works. Implementation, evaluation and monitoring takes as input the perspective data as well as the final policy and returns as output the final policy deployed in production. The Big Data platform is adopted to execute simulations on the suitability of the policy with respect to the perspective data. Policy makers and clinicians can do the following activities:

- *Simulation*: Simulate the deployment in production before the real deployment. The Big Data platform permits to continuously execute the classification and the comparison with normative and objective features in order to evaluate the efficacy of the policy, while it is executed in production.
- *Final Policy Release*: If simulations are positive enough, the policy can be released in production. The policy makers can also decide at this time to modify modal operator in the policy, for instance, switching from recommendation to obligation.

We note that Simulation is a continuous process that can be executed even while the final policy is released (i.e., monitoring). It this case, it monitors the policy and permits a fast reaction to societal or environmental changes, which may require a re-definition of the policy. For instance, if people are getting more used to have healthy food, the impact of the disease is getting lower and some of the normative features/ values can be updated lowering the need of costly early

[6] The number n of features can be defined or not in the Draft policy.

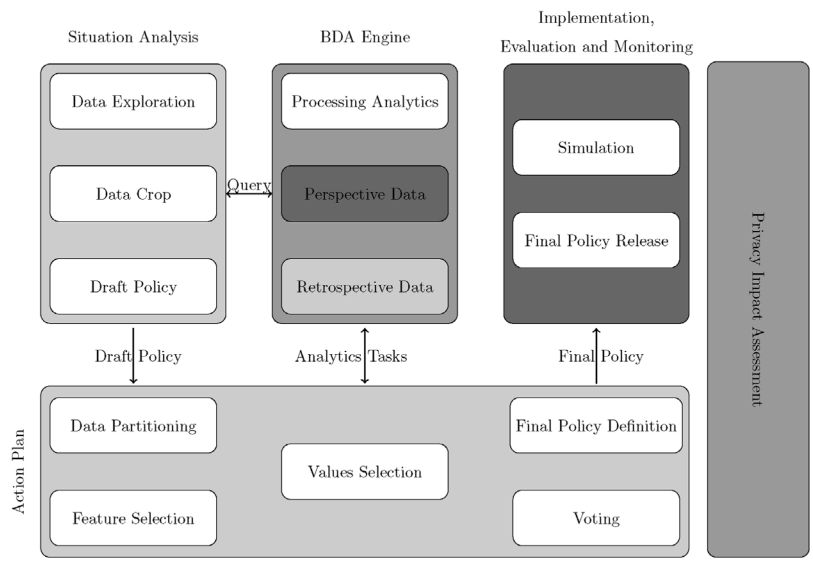

Fig. 4. Big Data-assisted public health policy making process.

screening.

From the perspective of policy makers, the support provided by the Big Data approach let them collect real-time evidence of the policy outcome and trends in the observed population of patients. This could be extremely valuable in situations like the re-hospitalization of diabetes patients, whose incidence depends on a wide range of factors, many of them correlated with patient's lifestyle, alimentary habits, cultural background, or environmental factors. Providing individual post-discharge assistance and education could sensibly reduce the rate of readmission for patients that lack the willingness for autonomously adopting healthy practices, however, the resources (nurses, diabetes specialists) are limited and should be employed for individuals whose readmission risk is very high.

Fig. 4 shows our Big Data-assisted public policy making process.

3.5. Model-Based Big Data Analytics as a Service (MBDAaaS)

Establishing a sound and straightforward link between regulations, in terms of policies and policy makers, and Big Data campaigns, in terms of analytics and Big Data technologists/scientists, is fundamental to make Big Data Analytics effective also in those public domains like health where Big Data Analytics are less effective. Current approaches and technologies for Big Data Analytics in fact suffer from two main drawbacks: (i) they are often seen as black boxes that hide workflows and computations in complex services, (ii) the availability of many different and non-standard solutions makes Big Data management difficult for not.expert users.

Our goal is therefore to provide an approach that addresses the above drawbacks. To this aim, we rely on the Model-Based Big Data Analytics as a Service (MBDAaaS) approach in Ardagna et al. (2017), where the policy makers, with the assistance of Big Data consultants, declaratively define the goals to be achieved by Big Data campaigns and such that they accomplish the defined public policies, while smarter engines manage Big Data platform deployment and analytics execution. MBDAaaS builds on a model-driven engineering paradigm (Schmidt, 2006) and decouples high-level goals of a Big Data campaign from low-level details of the Big Data architecture. It is based on three models defined as follows.

Declarative model. It is a computation-independent model describing the requirements in terms of goals to be achieved by a Big Data campaign. Each requirement is a triple specifying a *goal* to be achieved, an *indicator* on how to assess the goal, and an *objective* representing the threshold/class for the indicator. Requirements can be categorized in five different areas concerning data preparation, data representation, data analytics, data processing, and data visualization and reporting. For example, a typical requirement in area data preparation might require *Anonymization* (Goal) by means of an *Anonymization_Technique* (Indicator) based on *k*-anonymity (Objective). Each requirement can be further specified with constraints, such as the value for *k* used by the *k*-anonymity technique.

Procedural model. It is a platform-independent model describing what the Big Data Analytics should achieve and how to achieve objectives in the declarative model. The procedural model can be implemented as a simple direct acyclic graph where each node is a Big Data component with corresponding configurations (possibly restricted by constraints in the declarative model) and each edge is a call to the component. For example, a procedural model deriving from a declarative model specifying goal anonymization based on *k*-anonymity will include a component for *k*-anonymity.

Deployment model. It is a platform-dependent model instantiating the procedural model on the basis of the target Big Data engine. Each node in the procedural model is then replaced by one or more nodes in the deployment model, each referring to real Big Data components. For example, in the deployment model, we refer to a real component implementing *k*-anonymity.

MBDAaaS process for policy making is currently composed of three main phases as follows.

- *Declarative model specification*. The Big Data policy maker produces a declarative model specifying the goals of a Big Data campaign.
- *Declarative to procedural model transformation*. Abstract Big Data services compatible with goals in the declarative model are first selected. They (or a subset thereof) are then composed to address the defined public policies. The resulting service composition represents our procedural model.
- *Procedural to deployment model transformation*. Abstract services in the procedural model are instantiated in real services available on the Big Data engine. In other words, the procedural model is transformed in a platform-dependent workflow, which is executed on the target Big Data engine. This workflow represents the deployment model.

MBDAaaS, being based on MDA paradigm, increases the transparency of the Big Data campaigns by providing access to detailed execution workflow and Big Data computations. This approach makes the verification of adherence to defined policies a simpler process that can be achieved by traditional certification and compliance solutions. Moreover, it provides analytics automation reducing the complexity of the design and management of Big Data campaigns.

4. Privacy Compliance Assessment

The high privacy standards defined for the compliance with the GDPR will necessarily require public health initiatives based on Big Data to adequately assess privacy constraints since their design phase. This clearly applies also to MBDAaaS approaches like the one in Section 3.5. *Privacy-by-design* is now a formal requirement for regulatory compliance that should be comprehensively addressed in a structured and efficient way in projects dealing with large amounts of citizens' health data.

In this section, we recall the main topics of a *Privacy Compliance Assessment* with respect to the GDPR requirements that should be considered and suggest an analysis template of requirements in a semi-machine readable format, forming the basis for a privacy-by-design approach mixing automatic checks and an analysis carried out by privacy experts. Privacy Compliance Assessment and corresponding requirements are used in Section 5 to implement a privacy-aware Big Data-assisted public policy making process.

Establish a role. The first step in the privacy-by-design task is to identify its own formal role with respect to data protection. Public health initiatives aim to establish policies based on the feedbacks obtained from Big Data Analytics, means that large amount of personal data are under the control and are processed. Typically, three formal roles are the most relevant: *Data Controller*, *Data Processor*, and *Data Subject*. The first is the entity who determines the purpose of personal data processing, the second is the entity entrusted, by the Data Controller, with specific data processing tasks, while the third is the subject of a public health policy (i.e., citizens, patients, etc.).

Data status. The notion of personal data is crucial in a privacy assessment, representing the asset to protect from misuse. Public health initiatives based on Big Data very likely process personal data (i.e., this is true every time data from individuals are collected, it might not be the case for those projects processing aggregate or statistical data only). Once established that personal data is processed, other constraints for the applicability of the GDPR, mostly based on the EU membership nature of the data controller/processor/subject and the location of data processing, must be considered.

Transparency, fairness, and lawfulness. These are fundamental principles of data protection in the EU that should be enforced by all projects processing personal data. Citizens whose personal data are collected in the context of a public health project should be fully informed about the whole process. They should be aware of which personal data is collected, the location and duration of the storage, and which type of processing is carried out. The amount of personal data should be minimized with respect to the specific goal they are serving. This means that it is not possible to collect data for generic or weakly defined analyses.

Purpose limitation and consent. The principle enforced by the GDPR is clear: personal data can only be used for the specific purposes a Data Subject has been made aware of and has given explicit consent. It must be also noted that, being the acquisition of individual consent an activity with a significant cost, the exact purposes of personal data and processing types need to be defined in advance and with great precision. In addition, Data Subjects are granted with rights over their respective personal data: they can ask to access their data, ask the rectification whether they find incorrect data, and require the cancellation of all of them. In short, the enforced principle is that the ownership of personal data remains to the Data Subject and with it its natural rights. Traditionally, public health initiatives have never been used to consider personal data that way, for this reason procedures to grant Data Subjects' rights over personal data should be carefully planned.

Responsibility of the controller and of the data processor. It is the duty of Data Controllers to implement appropriate technical and organizational measures to comply with the GDPR. The guiding principle to follow is of *proportionality*: data protection policies and technical solutions need be proportionate to processing activities, adhere to approved codes of conducts or certification mechanisms, and take into account the nature, scope, context and purpose of processing as well as any risks for the rights and freedoms of natural persons. In other words, a *risk-oriented approach* is required for evaluating severity and likelihood of threats to personal data and adopt appropriate risk mitigation strategies with respect to costs and complexity of solutions.

Integrity and confidentiality. Data processors must ensure security of personal data, protecting them against unlawful processing, theft, accidental loss, destruction or damage. The principle informing decisions about which technical solutions should be adopted is still that of *proportionality*: technical solutions should be chosen according to the analysis performed in a risk assessment activity.

Data protection impact assessment (DPIA). In the particular case that the data processing is "likely to result in a high risk to the rights and freedoms of natural persons" (*ref. Art. 35* in European Parliament, 2016), then the Data Controller should carry out a preliminary assessment of the impact of the operation on the protection of personal data. This additional check is called *Data Protection Impact Assessment* (DPIA) and, given the special status of health data, considered among the most critical with respect to privacy, is likely to be required to public health initiatives based on Big Data.

Data protection officer. A data protection officer should be designated by data controllers and processors when processing is carried out by a public authority or body, typically the case for Public health initiatives.

Concluding, a Public Health initiative should account for carrying out a Privacy Compliance Assessment. This task should be associated with a policy making process in the sense that it regulates all data processing while defining a public health policy.

Table 1 shows a summary of the main controls for compatibility with the GDPR,[7] which we assume to be part of the semi-automated process of policy generation. In operational form, the privacy checks should be transformed into requirements to be satisfied within the Big Data-assisted Public Policy Making Process, in line with the principle of privacy-by-design.

The transformation into requirements are in general associated with each question in Table 1.

Example 5 (*Personal data processing requirements*). Let us consider for simplicity the question "Personal data are limited to what necessary in relation to the purposes?" related to Personal data processing in Table 1. This may trigger a number of requirements that can be defined by the process owner in the process of verifying compliance to the question. In the following we list some of them.

- Data must be anonymized before the evaluation.
- Data must be anonymized while it is processed.
- Data protection before the evaluation.
- Retention of personal data in plain form while it is processed must not exceed a specific period of time.
- Sensitive data related to a specific subject must be limited to the strictly needed amount for a specific processing.

These requirements should be further detailed in more technical terms, when processing activities are about to be triggered.

5. Privacy-aware Big Data-assisted public policy making process

In this section we extend the approach of Section 3 to include requirements derived from the Privacy Compliance Assessment

[7] The list of checks presented here is not meant to be fully comprehensive of the whole set of requirements needed to grant the legal compliance with the GDPR.

Table 1
Main checks for privacy compliance.

Role
Can you identify which role (Controller, Processor, Subject) you play in your Big Data project?
Data status
Does your project process any personal data?
Does your processing of personal data dissatisfy any condition for exclusion from the GDPR?
Personal data processing
Personal data are processed lawfully, fairly, and in a transparent manner?
Personal data are collected for specified, explicit, and legitimate purposes?
Personal data are limited to what necessary in relation to the purposes?
Personal data are accurate and kept up to date?
Personal data permits identification of data subjects for no longer than necessary for the purposes?
Appropriate security (e.g., confidentiality, integrity) of personal data is granted?
Purpose limitation and consent
Has data subject given written consent to processing his/her personal data for specific purposes or any other clause for exclusion of consent applies?
Was the request for consent clearly distinguishable from other matters and easily understandable?
Did you explicitly request consent for processing special categories of personal data (including health)?
Did you provide data subject with the prescribed information regarding the data controller and processor?
Did you inform data subject of his/her rights on personal data (access, rectification, and erasure) and on consent withdrawal?
Responsibility of data controller and processor
Was a risk analysis performed for the threats to natural person posed by your data processing?
Were “data protection by design” and “data protection by default” principles enforced?
Was access to personal data restricted to a limited and authorized number persons?
Were processing activities recorded according to the prescriptions of the GDPR?
Security
Do you employ state-of-the-art security processes and controls to protect personal data?
Did you perform a risk analysis to select security processes and controls?
Did you rely on assurance checks and audits?
Did you have an incident response procedure to identify data breaches and to notify your country's Supervisor Authority?
DPIA
Did you carry out a Data Protection Impact Assessment (DPIA), if the conditions defined by the GDPR are met?
Data protection officer
Was a data processor officer designated, if the conditions defined by the GDPR are met?

introduced in Section 4.

The Privacy Compliance Assessment is meant to define the core compliance checks that must be satisfied by a Big Data-assisted public policy making process to comply with the GDPR. We assume that the owner of the public policy making process has evaluated the requirements associated with the compliance checks as in Example 5 for all of the relevant checks, before including them into the policy making process. It is during the *Situation Analysis* stage that a policy makers, guided by Privacy Compliance Assessment, annotates the draft policy with constraints about the data to be processed and the requirements to be enforced on them (e.g., confidentiality at rest if data are pre-processed and temporally stored). For instance, if the requirements is “Data must be anonymized before the evaluation” like in Example 5, then it is needed to identify which are the data to be protected among all available data. This is obtained thanks to the knowledge extracted via *Data Exploration* step. We note that a situation analysis stage can be itself subject to Privacy Compliance Assessment while accessing the data.

Definition 2 (*Annotation***).** Let us consider a policy θ as in Definition 1, an annotated policy θ^{λ} is a policy associated with a set of requirements λ related to normative objectives and goal predicates (e.g., the sensibility of the data referred in the normative features), and, more in general, to the data processing required to meet the policy goal (e.g., requirement “Data must be anonymized while it is processed” in Example 5).

Example 6 (*Annotated Policy***).** Let us consider the policy in Example 4 and requirement “Data must be anonymized before the evaluation”. The annotation λ refers to all the predicates stating that they refer to a personal data and that any processing activities involving them must be anticipated with a preparation based on *Anonymization*.

Privacy Compliance Assessment annotations on the policy drive MBDAaaS. They mostly insist on phase *declarative model specification* by refining declarative model goals and corresponding goal domains, accordingly. In particular, declarative model specification is modified in a 2-step process as follows.

Privacy-Aware Declarative Goal Filtering. The first step receives privacy metadata λ of policy θ as input and returns a view v on the set of goals in the declarative model satisfying Privacy Compliance Assessment requirements as output. Goals can be manipulated at three levels of granularity as follows: (i) goals incompatible with Privacy Compliance Assessment requirements are removed, (ii) indicators incompatible with Privacy Compliance Assessment requirements are removed keeping the corresponding goal, (iii) objective domains are restricted to accomplish Privacy Compliance Assessment requirements.

Privacy-Aware Definition of Declarative Models. The second step is driven by public policies and aims to semi-automatically define the filtered declarative goals. Ad hoc rules, which depend on the specific scenario, can be defined to provide a link between conditions in a policy and goals in the declarative model.

Upon declarative model specification, the two transformations in the original MBDAaaS are applied and the Big Data campaign modeled by the deployment model executed on the target platform.

6. Related work

Public health policies, in the form of laws, regulations, and guidelines, have a profound effect on public health. However, there is a considerable gap between what research shows as effective and the policies that are enacted and enforced. Research is most likely to influence policy development through an extended process of communication and interaction (Brownson, Chriqui, & Stamatakis, 2009).

In particular, new technologies like for instance Big Data Analytics, give rise to a set of new opportunities for science, government, and citizens. Because of the novelty of these technologies, policymakers aimed at regulating such data-driven innovation, will attempt to draft new laws using existing paradigms and schema. But the evolution of technology always outpaces the legislative process, in particular in this context, paving the way to a re-definition of this process to establish a flexible, forward-looking policy-making procedure and reshape existing legislation to support the technological change (Hemerly, 2013).

Big data can expand the capacity to generate new knowledge and support the generation of health policy. The cost of answering many clinical questions prospectively, through the collection of structured data is high and somehow prohibitive. Moreover, Big Data-based analysis techniques of unstructured data within health report using computational techniques (like for instance the processing of natural language to extract medical concepts from free-text documents) allows the automatic acquisition of knowledge that can be used in the definition of health policies. Big data techniques can help with the dissemination of knowledge acquired using traditional or innovative systems. In fact, the digitization of medical knowledge and policies can greatly improved the access and the enforcement of better medical treatments (Murdoch & Detsky, 2013).

Big Data related techniques may help translate personalized medicine initiatives into clinical practice by offering the opportunity to use analytical capabilities that can integrate systems biology (eg, genomics) with personal medical and social data. In that case, medical records should be stored with patients and improved by linking traditional health-related data (eg, medication list and family history) to user personal data (e.g., income, education, neighborhood, military service, diet habits, sport activity, entertainment) (Murdoch & Detsky, 2013). Also. Big Data Analytics enables the capture of insights from data gathered from research, clinical care settings, and operational settings to build evidence for improved care delivery. As indicated in the Institute of Medicine (IoM) report, there are some open problems (The Global Use of Medicines, 2016) (e.g., how to manage the data coming from the IoT devices). There is a significant opportunity to improve the efficiencies in the healthcare industry by using an evidence-based learning model, which can in turn be powered by Big Data Analytics (The Global Use of Medicines, 2016).

The problem of the definition of health policy basing on Big Data Analytics has been considered of primary importance also by the European Commission in the framework of the H2020 program.[8] In their vision, data sources can be represented not only by new e-Health personal solutions, but can be extended also to more generic and commercial instruments, like mobile apps for health and well-being, and social networks, to integrate the social dimension in the analysis of the personal health scenario and, in an holitic view, in the definition of local health policies. In that case, it is important to assure ethical aspects of data, confidentiality, anonymity of data transfer and engagement of those who collect/ code such data in its analysis and interpretation, to avoid misinterpretation and inappropriate conclusions. The monitoring the combined effects of the previous factors, enables an early identification of effects, having large impacts the provision of healthcare services. The link between users and the provision of that services can be represented by the definition of general and personal health policies.

In that context, a number of methods and techniques can assist analysts in evaluating health policy like pseudo-evaluation, formal evaluation, and decision theoretic evaluation (Dunn, 2015). Their usage requires data sciences knowledge and expertise in the field of application, in addition their integration into a usable policy making-specific framework is far from being realized due to interdependency and heterogeneity of both techniques and programming languages.

Overall, McKinsey & Company estimates that $300 billion to $450 billion can be saved in the healthcare industry from Big Data Analytics (Peter Groves, Kayyali, & Kuiken, 2013a). As the focus shifts from cure to preventive health and as new technologies such as wearable sensors evolve as part of the Internet of Things (IoT), the volume of data in healthcare is expected to grow significantly and can provide a wealth of actionable information. The combined power of information from real-time devices, people, clinical systems, and historical population data makes Big Data a very helpful tool in improving the healthcare system (Peter Groves, Kayyali, & Kuiken, 2013b).

7. Conclusions

The definition of effective public health policies has been identified as one of the most wanted research achievement for smart cities environment. Big Data permits to obtain timely responses allowing frequent interactions with policy makers, the execution of several data mining tasks in parallel, to support the automatic monitoring and evaluation of a given evidence-based policy. With the advent of GDPR, Big Data processing as well as the policy making process as a whole, need to mandatory show compliance to privacy requirements, especially when treating medical sensible data. Thanks to the approach proposed in this paper privacy-aware Big Data executions can be simply modeled and triggered as a service supporting for the definition of the policy is straightforward way without requiring advanced data science background.

Acknowledgements

This work was partly supported by EU-funded project EVOTION (contract no. H2020-727521) and by the program “piano sostegno alla ricerca 2015-17” funded by Università degli Studi di Milano.

References

Anisetti, M., Bellandi, V., Cremonini, M., Damiani, E., & Maggesi, J. (2017). Big data platform for public health policies. In: *Proc. of IEEE SmartWorld 2017*.
Ardagna, C. A., Bellandi, V., Bezzi, M., Ceravolo, P., Damiani, E., & Hebert, C. (2017). A model-driven methodology for Big Data Analytics-as-a-service. In: *Proc. of IEEE congress on Big Data*.
Balbiani, P. (2008). Logical approaches to deontic reasoning: From basic questions to dynamic solutions. *International Journal of Intelligent Systems, 23*(10), 1021–1045.
Boulos, M. N. K., & Al-Shorbaji, N. M. (2014). On the internet of things, smart cities and the who healthy cities. *International Journal of Health Geographics, 13*(1), 10.
Brownson, R. C., Chriqui, J. F., & Stamatakis, K. A. (2009). Understanding evidence-based public health policy. *American Journal of Public Health, 99*(9), 1576–1583.
Dungan, K., Lyons, S., Manu, K., Kulkarni, M., Ebrahim, K., Grantier, C., et al. (2014). An individualized inpatient diabetes education and hospital transition program for poorly controlled hospitalized patients with diabetes. *Endocrine Practice*.
Dungan, K. M. (2012). *The effect of diabetes on hospital readmissions*.
Dunn, W. N. (2015). *Public policy analysis.* Routledge.
Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European UnionL119*, 1–88. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC.
Hemerly, J. (2013). Public policy considerations for data-driven innovation. *Computer, 46*(6), 25–31.
Loukis, E. N. (2007). An ontology for g2g collaboration in public policy making, implementation and evaluation. *Artificial Intelligence and Law, 15*(1), 19–48.
Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Journal of the American Medical Association, 309*(13), 1351–1352.
Peter Groves, D. K., Kayyali, B., & Kuiken, S. V. (2013a). *The Big Data revolution in healthcare: Accelerating value and innovation.* Center for US Health System Reform.
Peter Groves, D. K., Kayyali, B., & Kuiken, S. V. (2013b). *The Big Data revolution in healthcare: Center of US health system.* Center for US Health System Reform.
Rubin, D. J. (2015). Hospital readmission of patients with diabetes. *Current Diabetes Reports, 15*(4), 1–9.
Schmidt, D. (2006). Guest editor's introduction: Model-driven engineering. *Computer, 39*(2), 25–31.
The Global Use of Medicines: Outlook Through 2016. MS Institute.

[8] http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/calls/h2020-sc1-2016-2017.html.