A Review of Big Data Quality and an Assessment Method and features of Data Quality for Public Health Information Systems

Author's Details:
[1] Paolo Pietro Biancone [2] Silvana Secinaro [3] Valerio Brescia

Abstract: *Data governance Refers to the management of public health information systems and data. The article tries to give an updated definition of big data quality through a review and systematic approach. We Identified publications by searching several eletronic bibliographic databases. The articles were confined to Inglese and italian leanguage. We performed the litterature advanced between January 2015 and December 2016. The group of study would propose a method for handle Big Data Quality System and the features That the system must have to be a management tool. The study investigates and provides some thoughts and ideas on the future of big data quality and their use in the Governance in Health. The paper highlights the usefulness of big data quality both in medical choices in information systems and infrastructure useful to policy decisions and organizational control*

Key words: *big data, big data quality, data quality management, healtcare, healthcare information systems*

1. Introduction

Data governance refers to the management of public health information systems and data. It is an infrastructure designed to "mediate divergent interests, solicit involvement from and communicate effectively with diverse constituent groups, and provide direction in the coordinated development of public health information systems that support the health improvement efforts of communities" (Shabo and all. 2016). With the emergence of more and more sophisticated informatics technologies, the growth of data available, and push for data-sharing, it is increasingly important that public health practitioners responsible for the development and management of their information systems be fully informed when making decisions on system design, data integration and use (June 2016). In the healthcare sector, lack of data quality has far-reaching effects. Planning and delivery of services rely heavily on data from clinical, administrative and management sources. For example, evidence-based practice (Strauss et al., 2005) requires access to extensive research data, collated and presented in a way that a clinician can use at the time of diagnosis or in other decision-making situations. The higher the quality of the data, the better will be the patient outcomes. Similarly, quality data, particularly with regard to timeliness and accuracy, are needed for administrative purposes such as hospital bed-rostering and for planning services to ensure that they are cost-effective. These different but interlocking data requirements and decisions ensure that health care organisations and their relationships are inherently complex and demanding (Gendron & D'Onofrio, 2001). Health care data are items of knowledge about an individual patient or a group of patients. In health care, data are captured about a patient in paper or electronic format during his or her attendance at an outpatient clinic, community health centre, primary health care provider, or his or her admission to a hospital (Davis and LaCour, 2002). To ensure data quality, two key principles are data accuracy and data validity. To communicate effectively, data must be valid and conform to an expected range of values. To be useful, data must be accurate. Providers of health care services need information not only at the point of service but also at the point of decision making in a format that maximizes the decision-making process. Correct and up-to-date information is critical, not only for the provision of high-quality clinical care, but also for continuing health care, maintaining health care at an optimal level, clinical and health service research, and planning and management of health systems. The collection and use of information should not impose a burden on the health system. It should be collected as a routine by-product of the health care process. As coded health care data are being increasingly used in the health care environment, it is important to ensure that the original source data are accurate and timely, which in turn, will produce reliable and useful information. (World Health Organization 2003). The healthcare industry historically has generated large amounts of data, driven by record keeping, compliance, and regulatory requirements, and patient care (Raghupathi 2010) While most data is stored in hard copy form, the current trend is toward rapid digitization of these large amounts of data. Driven by mandatory requirements and the potential to improve the quality of healthcare delivery meanwhile reducing the costs, these massive quantities of data (known as 'big data') hold the promise of supporting a wide range of medical and healthcare functions, including among others clinical decision support, disease surveillance, and population health management (Burghard 2012, Fernandes 2012). The Institute for Health Technology Transformation defines big data as “a term that describes large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.” Big data goes beyond size and volume to encompass such characteristics as variety, velocity, and, with respect specifically to health care, veracity. Big data can be said to comprise five different categories, or streams, of information: 1. Web and social media data: Clickstream and interaction data from social media such as Facebook, Twitter, Linkedin, and blogs. It can also include health plan websites, smartphone apps, etc. 2. Machine-to-machine data: Readings from sensors, meters, and other devices. 3. Big transaction data: Health care claims and other billing records increasingly available in semi-structured and unstructured formats. 4. Biometric data: Fingerprints, genetics, handwriting, retinal scans, and similar types of data. This would also include X-rays and other medical images, blood pressure, pulse and pulse-oximetry readings, and other similar types of data. 5. Human-generated data: Unstructured and semi-structured data such as electronic medical records (EMRs), physicians’ notes, email, and paper documents. In recent years, it has become increasingly apparent that multiple streams of data like these can be leveraged with the powerful new collection, aggregation, and analytics technologies and techniques to improve the delivery of health care at the level of individual patients as well as at the levels of disease- and condition-specific populations. The expansion and the use of big data quality part in the 90, and is often linked to investments in information technology and information system. In recent years they have several problems and deficiencies arising from the non-use of big data in information techlology and strategy planning: business opportunities are missed; the business may even be disadventaged by the information system or digital developments, application and infrastructure investments do not support the business objectives and may even become a constraint to business development, lack of integration of application and ineffective information management produce duplication of effort and inaccurate and inadequate for managing the business, priorities are not based on business needs, resources levels are not optimal and investment plans are consistently changed. Business performance does not improve, costs are high, solutions are of poor quality and value for money is low, technology strategy is incoherent, incompatible options are selected and large sums of money are wasted attempting to fit things together retrospectively. Lack of understanding and agreed direction between users, senior management and the information system and information technology specialists leads to conflict, inappropriate solutions and a misure of resources (Ward et all. 2016).

1.1 Big data lifecycle

The processing of large data is through a life cycle now known in the literature, in particular, and divided into 5 steps. Data collection: it involves the collection of data from various sources and storing. Data can be anything such as casa history, medical images, social logs, senord data etc. Data cleaning: it involves the process of verifyng whether there is any junk data that have missing values. Such data needs to be removed. Data classification: it involves the filtering of data based on their structure. Big data consists of mosly unstructured data such as hand written physician notes. Structured, semi-structured and unstructured data should be classified in order to perform meaningful analysis. Data modelling: it involves performing analysis on the classified data. First it has to classify the data based on the specific location, need to trigger the health report of children, need to identify the children whose family are under the poverty line and these data should be processed. Data delivery: it involves the generation of a report based on the data modelling done. At the all the stages of Big data lifecycle it required data storage, data integrity and fata access control (Lynch 2008, Li 2015, Demchenko 2012).

2. Methods

2.1 Literature Search

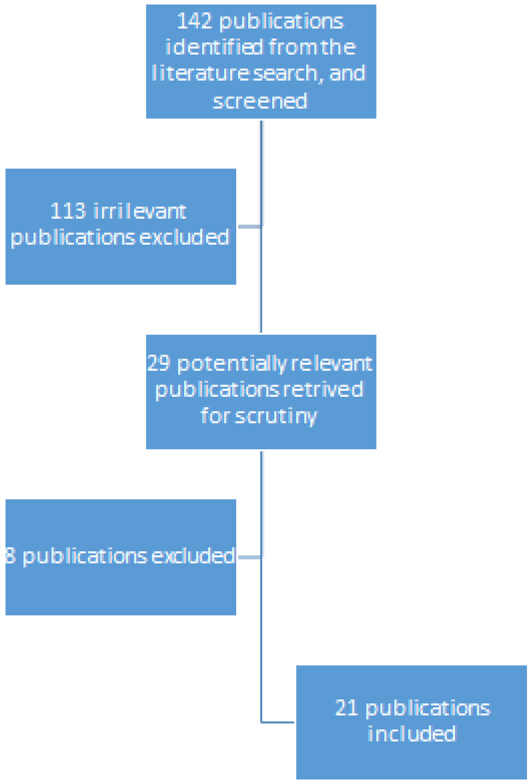
We identified publications by searching several eletronic bibliographic databases. These included Scopus, MEDLINE/PudMed Web of Science, PudMed Central, CrossRef, National Library of Sweden, BMJ Publishing Group, Oxford University Press, SAGE Journals, Elsevier. The following words and MeSH heading were used individually or in combination: “big data quality healthcare” and “data quality”. The articles were confined to english and italian leanguage. We performed the litterature serch between January

2015 and December 2016. The inclusion criteria were peer-refereed. The exclusion criteria were an expert opinion, correspondence and commentaries in the topic area. To increase coverage, a manual serach of the literature was conducted to identify papers references by other publications, papers and well-known authors, and papers from personal databases. The group of study would propose a method for handle Big Data Quality System and the features that the system must have to be a management tool.

2.2 Selection of Publications

Citations identified in the literature search were screened by title and abstract for a decision about inclusion or exclusion in this review. If there was uncertly about the relevance of citation, the full-text was retrieved and checked. A total of 142 publications were identified and were manually screened. If there was uncertaly about whether to include publication, its relevance was checked by the authors. Finally, publications that met the inclusion criteria were selected. The screening processi s summarized in Figure 1.

Figure 1. Publication search process.



2.3 Data abstraction

The selected publications were stored in an EndNote library. Data extracted from the publications includes author, year of of publication, aim of data quality assessment, country and context of the study, function and scope, definition of big data quality, methods for data quality assessment, study design, data collection methods, data collected, research procedure, methods for data analysis, key finding, conclusion and limits. Only 21 selected state publications and are assessed as suitable for the purposes of our analysis.

3 Results

Of the 21 publications review, 21 were peer refereed research paper. The publications selected are listed in Table 1. The number of publications between 2015 and 2016 is in line with the previous three years. The themes are similar to the previous three years even if the application of big data quality in health care is applied in the last two years.

Table 1 Data quality assessment publications and DOI references

Title	DOI
The promising future of healthcare services: When big data analytics meets wearable technology	https://doi.org/10.1016/j.im.2016.07.003
Technology Roadmap Development for Big Data Healthcare Applications	https://doi.org/10.1016/j.im.2016.07.003
A Semantic Big Data Platform for Integrating Heterogeneous Wearable Data in Healthcare	https://doi.org/10.1007/s10916-015-0344-x
Impact of Processing and Analyzing Healthcare Big Data on Cloud Computing Environment by Implementing Hadoop Cluster	https://doi.org/10.1016/j.procs.2016.05.171
A novel intelligent approach for predicting atherosclerotic individuals from big data for healthcare	https://doi.org/10.1080/00207543.2015.1087655
Big data as a new approach to emergency medicine research	https://doi.org/10.1016/j.joad.2015.04.003
Big Data: transforming drug development and health policy decision making	https://doi.org/10.1007/s10742-016-0144-x
Indian Health Care Analysis using Big Data Programming Tool	https://doi.org/10.1016/j.procs.2016.06.101
Big Data and Predictive Analytics - Applications in the Care of Children	https://doi.org/10.1016/j.pcl.2015.12.007
A Systems Approach Using Big Data to Improve Safety and Quality in Radiation Oncology	https://doi.org/10.1016/j.ijrobp.2015.10.024
A Framework to Assess Healthcare Data Quality	https://doi.org/10.15405/ejsbs.156
Big Data Analytics in Healthcare	https://doi.org/10.1155/2015/370194
Improving Therapeutic Effectiveness and Safety Through Big Healthcare Data	https://doi.org/10.1002/cpt.316
Big Data and Healthcare: Building an Augmented World	https://doi.org/10.4258/hir.2016.22.3.153
A Survey Of Big Data Analytics in Healthcare and Government	https://doi.org/10.1016/j.procs.2015.04.021
Toward a Literature-Driven Definition of Big Data in Healthcare	https://doi.org/10.1155/2015/639021
Big Data in Israeli healthcare: hopes and challenges report of an international workshop	https://doi.org/10.1186/s13584-015-0057-0
Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data	https://doi.org/10.1186/s13742-016-0117-6
Big Data Analysis Framework for Healthcare and Social Sectors in Korea	https://doi.org/10.4258/hir.2015.21.1.3
A Data Quality in Use model for Big Data	https://doi.org/10.1007/978-3-319-12256-4_7
Design of QoS-Aware Multi-Level MAC-Layer for Wireless Body Area Network	https://doi.org/10.1007/s10916-015-0336-x

3.1 Contex and Scope of the Studies

Several studies and publications in the field of big data in health care products in the past decade. But still it has not been given a clear definition of big data quality and characteristics of the system and the data life cycle in the health system. The study seeks to identify the characteristics of the health sector in big data, by identifying what criteria must meet the big data quality. The health sector, however, can be identified, as we shall see, in different sectors. The analysis sometimes involves the scope of generalized policies and the use of data in this context, very often the application and individual sectors or medical research fields. The analysis was carried out to check if it is possible to consider the big data quality as a tool for planning and control and risk management. We have an approach based on the reformulated theory that highlights in the process the characteristics that must possess a big data quality to respond to potential problems and limitations identified by the literature in the light of the latest publications in the health field.

3.2 Methods for Data Quality Assessment in the review

The analysis methods for assessing the quality of data In the reviewed publications and presented in the base of different criteria: Data, data collection or use of process data sector. Six points of view are examined, including the university quality attributes for any size, greater measurement indicators for each attribute design of the studio / evaluation method, the data Collection methods, the methods of data analysis, the contributions and limitations.

3.2.1 Methods fos Assessment of the Dimension of Data of the review

In this section, the concept of data quality is a narrow one, meaning the quality of the dimension of date. All of the publications, a total of 21 publications, conducted an assessment of the quality of data. At the end of our evaluation, we used the analysis of quality criteria for data used by Spepherd (Shepherd 2012) .

3.3 Quality attributes of data and corresponding measures

They analyzed the attributes major measures used in the advanced publications. The data quality is used in various areas and they are more criteria attributes. Only one publication does not identify any adjective or distinctive to identify and clarify the quality of big data. In particular the quality and big data have different impacts summarized in Table 2. The study also identified some of the possible problems that may arise.

Table 2 Attributes major measures and possible problems

Impact	Adjectives	Possible problems
strategies, implementation of healthcare, reduce healthcare cost, impact on quality and efficiency, sharing data, accessibility of digitalized patinet, unprecedent challenge opportunities to customize the care, use processing application to expedite the diagnosis, identify ineficiencies in care delivery and variation from standard practice, expanding services, maximize efficiency with limited resources, peer benchmarking, support research, maximizing quality and patient safety, test hypotheses, increase confidence in validity, collect and custumize different pieces of information, it used to combine the results of different devices, merge data with different formats	volume, variety, veracity, value	quality, safety, size, speed, type, complexity, standardization, variability, formats, different rates of speed, heterogeneity

Almost all of the studies identified in the use of big data a useful tool to increase the quality of service, reduce costs and support the decisions both in the management field and in the medical field through the customization of the treatment or prevention of disease.

The characteristic adjectives big data are identified in the 4 V: volume, variety, veracity and value. Volume is voluminous of data or ammount of data. Veracity refers to the trustworthiness of data. Variety is the range of data types and sources. Value refers to techiques of deriving value from data.

Instead are several problems that may arise related to the characteristics and needs of processing and use of big data; in particular, the ability to process data with different dimensions, data is often heterogeneous, of different types and different quality or data that are not part of the same set.

3.4 Study design in the review

Ten studies using literature review, most often refers to the framework of big data, other times in systematic studies of the system and of analytics of big data analysis. In the majority of the articles analyzed, 12 using qualitative models of analysis, this immediately brings attention to the fact that quantitative studies are reduced and often refer to small samples. In most homes the analyzed data studies are referred to national contexts or in single cases of pathologies and processing. This figure is related to the number of studies that apply a quantitative methodology. Large is the number of studies that apply a comparison of methodological approaches and features Included that the system must have in the life cycle of the data in the big data system. In almost all instances the data collected refer to a maximum period of 3 years and have a large number of samples but in specific contexts. This never allows generalizing the findings obtained by taking account of the territory and local social and epidemiological characteristics.

3.5 Data collection methods in the review

In almost all studies of the data are collected through computer and electronic systems and media. These instruments for the collection is complemented by the available software. In some cases, the collection tool to highlight theoretical models is comparative across tables. Rarely they are proposed mathematical tools to explain the information collection methods. In only a case we found a semi structured questionnaire survey.

3.6 Data analysis methods in the review

Data analysis methods were determined by the purpose of the study and the types of data collected. In most cases we have identified a prevalence of statistical methods of analysis and data processing. Almost always there is a correlation between the methods collected data. When there is the presence of numerical data we have a quantitative analysis and a correlation also in this case between the variables identified or in some cases with adopted or adoptable policies. In several cases it is possible to identify the construction of roadmap data and elaboration of the big data. A few times we have theoretical models of public economic fallout that highlight the use of big data in the market.

3.7 Methods for assessment of the dimension of data use

The studies were Concerned with the assessment of data use and the factors influencing data use. The other included assessment of data use, but this was not always highlighted in most cases.

3.8 Conclusion of review

The conclusions of several studies show there's big data are necessary for the evolution of the health care system. In fact, it is clear from the findings that allow big data according to the type of speleologia. Located price savings and linked to the service quality. The big data make it possible to prevent some diseases through the construction of scenarios based on the collected data. They are essential as it related to new technologies is identified in telemedicine that equipment used daily as smartphones, social media and digital TV networks. Implement the decisions in terms of medical care and priority of care cases; in this sense they increase the safety of patients and allow a greater diagnosis of diseases based on the symptoms and to search keys. Great emphasis is given to two large expenditure management issues with increased efficiency, these are the drugs and care personalization. In both cases it is avoided excessive waste or unnecessary dispersion of performance. Major medical research areas where they are currently used big data analysis are three: medical image analysis, Physiologica signal processing, and genome data processing. In all cases it is evident that the technological advancement also allowed by greater access and processing of data within the system, correlated where possible with an external system, allows greater control of the entire company system is becoming an important attachment for the risk assessment. Nearly all papers identify the use of big data as a tool for planning and control, alongside the medical aspect with the economic one. Therefore, the common area can be identified in public health but also in management. Almost always, the working group identified in the analysis consists of doctors with different specialties, IT and corporatists/economists.

3.9 Limitations of review

The search limits as mentioned earlier, are often identified within limits offered by samples associated with individual pathologies, cases or territories. Very often the results are not generalizable being different epidemiological and social context in which the world, but each continent is located. As for the review and the discursive analysis it is not possible to find evidence of a generalization of the different territories of the models. Most often the time limit or the low investment on big data does not allow a real confirmation of schedules and proposed approaches.

4 Discussions

For several years we witness in computerising internal communication and delivery of services (De Carvalho, 2016). The recipients of public information are, in general, all the stakeholders and the users of the balance sheet. It is important to distinguish between the subjects of “internal communication” and the ones of “external communication”. The subjects of the former are: the community, the other public entities, the lenders, the providers, the third; the recipients of the latter are: the managers, the political and administrative bodies, employees (Puddu 2011). In internal information systems and external becomes necessary to make the system processing of accounting and other data in order to ensure a total quality through the administration ration public company. This also allows you to meet the needs related to IPSAS accounting standards related to transparency (Biancone et all., 2016). Governance has now needed, especially in large amounts of information and data structures that can only be synthesized and analyzed through the use of big data quality (Chergui, 2016). Without these risk management tools to governance it is not found in to account to be able to justify their actions. This as seen above also holds for the lower decision levels as for example in patient care. The privatization of health facilities in Europe and the increasing size of public companies in order to streamline costs no longer allows executives to see the big picture together without tools enabling real way an immediate reading (Andre et all. 2016). In this situation it is essential to have an information system capable of providing the data immediately without making mistakes and allowing comparison between governance standards and quality of service (Torchia, 2016). At this both it supports the need for accounting harmonization in order to better plan and schedule activities in a perspective of analytical accounts of public companies (Puddu et. All, 2016).

From the literature and analyzed the evidence gathered to date suggest an approach to quality management that comprises several phases, these follow the approach of rational accounting and the total quality system (table 3):

- Application: the purpose for which the data are collected

- Collection: process by which the data is accumulated

- Storage: process and systems used to store data and journal data

- Analysis: the data transformation process information used for an application

Table 3 quality management approach and phases

Characteristic	Application	Collection	Warehousing	Analysis
Data accuracy Data are the correct values and are valid.	To help exactness, fix the practice's function, the interrogation to be replied, or the intention of assembling the information component.	Guaranteeing exactness requires proper coaching and practice and opportune and correct transmission of information descriptions to those who gather information. For instance, information exactness will help guarantee that if a case's sex is female, it is faithfully reported as female and not male.	To store information, correct revisions should be in place to guarantee exactness. For instance, mistake notes should be created for unstable standards such as an examination unsuitable for age or gender. Peculiarity or mistake notes should be created and adjustments should be produced.	To carefully examine information, guarantee that the algorithms, principles, and translation procedures are equitable. For instance, guarantee that the encoder gives right ciphers and that the exact DRG is given for the ciphers entered. Also, make sure that every report or access to the database is true.
Data accessibility Data items should be obtainable and legal to collect.	The practice and lawful, monetary operation, and other margins establish which information to assemble. Make sure that assembled information is lawful to assemble for the practice. For instance, reporting the age and race in medical reports may be correct. Nevertheless, it may be illicit to assemble these data in human resources departments.	When progressing the information assemblage tool, analyse approaches to enter the required information and make sure that the best, least expensive approach is chosen. The entirety of available information may be expanded through procedure interfaces and incorporation of procedures. For instance, the best and simplest approach to acquiring demographic data may be to acquire it from a living procedure. Another approach may be to give information assemblage by the expertness of every team member. For instance, the entry personnel gathers	Technology and hardware hit approachability. Organize information possession and guidelines for who may enter information and/or procedures. Schedule information to simplify entry.	Entry to entire, contemporary information will better guarantee faithful examination. Otherwise products and consequences may be incorrect or unsuitable. For instance, employment of the Medicare case mix index (CMI) alone does not properly display total hospital CMI. Therefore, calculated planning based merely on Medicare CMI may not be exact.

		demographic information, the nursing personnel gather symptoms and the HIM personnel give ciphers. Team members should be given properly.		
Data comprehensiveness All required data items are included. Ensure that the entire scope of the data is collected and document intentional limitations.	Explain how the information will be employed and catalogue end-users to guarantee all information is assembled for the practice. Incorporate a problem relation and cost aid or affect study when assembled information is expanded. For instance, additionally to result it may be relevant to assemble information that affects results.	Cost-operative all-inclusive information assemblage may be carried out via an interface to or download from other automated procedures. Information description and information accuracy affect all-inclusive information assemblage (see these properties below).	Storage incorporates managing associations of information holders, information acquirers and information end-users to guarantee that everybody is conscious of the accessible information in the schedule and approachable procedures. This also helps to cut out extra information assemblage.	Make sure that the whole appropriate information affecting the procedure is examined in collaboration.
Data consistency The value of the data should be reliable and the same across applications.	Information is important when the value of the information is the same through practices and procedures such as, the case's medical report number. Moreover, linked information pieces should comply. For instance, information is important when it is reported that a male case has had a hysterectomy.	The usage of information descriptions, comprehensive coaching, institutionalized information assemblage (methods, precepts, edits and operation) and combined/interfaced procedures simplify correspondence.	Storage requires edits or conversion boards to guarantee correspondence. Correlate edits and boards with information description substitutions or information description distinctions through procedures. Certify edits and boards.	Examine information under reproducible conditions by employing accepted prescriptions, mathematical equivalence, divergence calculations and other approaches. Correlate "apples to apples".

Data currency The data should be up to date. A date value is up to date if it is current at a specific point in time. It is outdated if it was current at some preceding time yet incorrect at a later time.	The properness or importance of practice differs through time. For instance, established quality assurance practices are progressively being substituted by those with the more contemporary practice of efficiency amelioration.	Information descriptions differ or are altered through time. These should be reported so that contemporary and coming users know what the information indicates. These modifications should be conveyed in an opportune way to those gathering information and to the end-users.	To make sure contemporary information is accessible, storage requires constantly modernizing procedures, boards and databases. The dates of storage events should be reported.	The accessibility of contemporary information affects the examination of information. For instance, to analyse the degree of illnesses or operations, ICD-9-CM ciphers may be employed. Ciphering methods or the verified cipher for an illness or methods may differ through time. This should be taken into account when examining tendencies.
Data definitions Clear definitions should be provided so that current and future data users will know what the data mean. Each data element should have a clear meaning and acceptable values.	The practice's function, the interrogation to be replied, or the intention of assembling the information component must be explained to guarantee exact and full information descriptions.	Comprehensible, brief information descriptions simplify proper information assemblage. For instance, the description of case disposal may be "the case's anticipated position or status succeeding liberation or exoneration". Admissible values for this information component should also be described. The tool of assemblage should comprehend information descriptions and make sure that information coherence peculiarities are controlled.	Storage comprehends recording documentation and information. As a result, information possession documentation and descriptions should be kept through time. Schedule keeping exercises (axing, updates and others), aim for gathering information, assemblage programmes, data administration programmes and information origins should be kept through time too.	For proper examination, demonstration information requires mirroring the aim for which the information was assembled. This is described by the practice. Proper correlations, associations and links require being displayed.
Data granularity The attributes and	A unique practice may need	Gather information at the proper position of	Storage information at the	Proper examination

value of data should be defined at the correct level of detail.	changing positions of particular or granularity. For instance, census statistics may be used day by day, by the week or monthly according to the practice. Census is required day by day to guarantee proper staffing and food utility. Nevertheless, the monthly tendency is required for long-range planning.	particular or granularity. For instance, the temperature of 100° may be registered. The granularity for reporting outside temperatures is distinct from reporting case temperature. If case Jane Doe's temperature is 100°, does that signify 99.6° or 100.4°? Proper granularity for this practice suggests that the information requires being registered to the first decimal point while proper granularity for reporting outside temperatures may not need it.	proper position of particular or granularity. For instance, inconsistency or mistake records indicate granularity based on the practice. A spike (inconsistency) in the day-by-day census may display little or no effect on the month-to-date or monthly records.	displays the position of particular or granularity of the information assembled. For instance, a spike (inconsistency) in the day-by-day census happening in instant activity to guarantee exact food utility and staffing may have had no effect on examination of the census for long-range planning.
Data precision Data value should be just large enough to support the application or process.	The practice's function, the interrogation to be replied, or the intention of assembling the information component must be explained to guarantee information accuracy.	To gather information accurate enough for the practice, describe adequate values or value fields for every information detail. For instance, contain values for gender to male, female and unknown; or gather data by age fields.		
Data relevancy The data are meaningful to the performance of the process or application for which they are collected.	The practice's function, the interrogation to be replied, or the intention of assembling the information component must be explained to guarantee appropriate information.	To better guarantee appropriateness, fill a pilot of the information assemblage tool to certify its usage. A "parallel" test may be proper too, filling the new or checked tool and the contemporary procedure at the same time. Report results to those assembling information and to	Decide proper retention calendars to guarantee accessibility of important information. Importance is described by the practice.	For proper examination, show information to indicate the aim for which the information was assembled. This is described by the practice. Display proper comparabilities, associations and links.

		the end-users. Simplify or arrange distinctions as required by fields or users.		
Data timeliness Timeliness is determinated by how the data are being used and their context.	Timeliness is described by the practice. For instance, case census is required day-by-day performances staffing, such as nursing and food utility. Nevertheless, yearly or monthly case census information is required for the resources decisive planning.	Opportune information assemblage is an exercise of the procedure and assemblage tool.	Storage makes sure that information is accessible per data administration code and retention programmes.	Opportune information examination leads to the introduction of activity to prevent conflicting effects. For some practices, opportune may be seconds. For others it may be years.

5 Results

In our view there is no single definition scientifically proven and generalizable to big data quality. Almost all the findings and analyzes identify certain characteristics and some problems that may occur in the use of big data. The identifying adjectives big data and big data quality are still the same and they are also the cause of the problems that arise most frequently in data management. For this reason we tried to provide an approach for managing big data quality in order to remove the limitations and expand the application in health care in light of the latest publications, discoveries and thematic areas. The literature shows that they are still few investments amplicare big data quality and IT technology tools that can really assist the governance and responsible choices (WHO, 2003). This is also why the few found evidence unique to each territory or medical areas. The use of high quality data sector and closely related strategic planning information system. The benefit of data quality and information systems strategic planes includes: effective management of an expensive and critical asset organization, improving communication and the relationship between the business and information system organization, aligning the information system direction and priorities to the business direction and priorities, identifying opportunities to use technology for a competitive advantage and increase the value to the business, planning the flow of information and process, efficently and effectively allocating information system resources, reducing the effort and money required throughout the life cycle of systems. The integration of big data lifecycle, and the proposed model improves the highlighted problems in information system and to avoid the problems encountered in the literature on the development and analysis of data. The integration of data between devices and sharing of facilities will be the basis of new medicine and new treatment installation path. A field still being explored that gave the first results just in preventive medicine and public health with regard to patient management and care pathways. We are confident that the big data quality is able to give essential information and useful, but only if they meet few features required in the approach. The data from processing errors and the collection will be deleted and the next few years will provide healthcare organizations the ability to provide information both internally and externally. The big data quality according to the literature thus guarantees financial savings, greater efficiency and effectiveness. The future analysis will be focused on what the big data quality is able to assist the comparison and equality in the quality of services offered at the European level between

Member States (Ferrara ,2008; Pasini 2011). The possibility of harmonizing the accounting systems and to compare the quantitative and qualitative results of the services offered by the different countries in Europe will also through the use of big data quality a real reduction of the differences present in spite of the common constitutional objective in all member countries (Condor 2016; Pontoppidand et all. 2016). The limit of this survey is considered the health sector as a whole. Necessary data quality on an analysis for individual sectors of health activities. The depth degree focused by sector would be more accurate and therefore more significant. Future work will focus on application areas.

References:

i. *Abdelhak M, Grostick S, Hankin MA, Jacobs (1996). E. Health Information: Management of a Strategic Resource. Philadelphia, WB Saunders Company. doi:10.4337/9780857932006.00025*

ii. *Alemayehu, D., & Berger, M. L. Big Data: transforming drug development and health policy decision making. Health Services and Outcomes Research Methodology, 1-11. doi:10.1007/s10742-016-0144-x*

iii. *André, C., Batifoulier, P., & Jansen-Ferreira, M. (2016). Health care privatization processes in Europe: Theoretical justifications and empirical classification. International Social Security Review, 69(1), 3-23. doi:10.1111/issr.12092*

iv. *Archenaa, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. Procedia Computer Science, 50, 408-413. doi:10.1016/j.procs.2015.04.021*

v. *Belle, A., Thiagarajan, R., Soroushmehr, S. M., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big data analytics in healthcare. BioMed research international, 2015. doi:10.1155/2015/370194*

vi. *Baro, E., Degoul, S., Beuscart, R., & Chazard, E. (2015). Toward a literature-driven definition of big data in healthcare. BioMed research international, 2015. doi:10.1155/2015/639021*

vii. *Beniger, J. (2009). The control revolution: Technological and economic origins of the information society. Harvard university press. doi:10.2307/2578757*

viii. *Biancone, P., Secinaro, S., & Brescia, V. (2016). Popular Report and Consolidated Financial Statements in Public Utilities. Different Tools to Inform the Citizens, a Long Journey of the Transparency. International Journal of Business and Social Science Vol. 7, No. 1; January 2016 doi:10.5539/ijbm.v11n11p115*

ix. *Burghard C: Big Data and Analytics Key to Accountable Care Success. IDC Health Insights; 2012.*

x. *Chang, H., & Choi, M. (2016). Big data and healthcare: building an augmented world. Healthcare Informatics Research, 22(3), 153-155. doi:10.4258/hir.2016.22.3.153*

xi. *Chergui, M., Chakir, A., Medromi, H., & Radoui, M. (2017). A New Approach for Modeling Strategic IT Governance Workflow. In Advances in Ubiquitous Networking 2 (pp. 285-298). Springer Singapore. doi:10.1007/978-981-10-1627-1_22*

xii. *Condor, V. (2016). Public Sector Accounting and Auditing in Europe. The Challenge of Harmonization. doi:10.1080/17449480.2016.1251601*

xiii. *Cusack CM, H. G. (2012). The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. Journal of the American Medical Informatics Association, 1-7. doi: 10.1136/amiajnl-2012-001093*

xiv. *Dayal, M., & Singh, N. (2016). Indian Health Care Analysis using Big Data Programming Tool. Procedia Computer Science, 89, 521-527. doi: 10.1016/j.procs.2016.06.101*

xv. *Davis N, LaCour M. Introduction to Information Technology. Philadelphia, WB Saunders Company, 2002.*

xvi. *De Carvalho, J. V., Rocha, Á., & de Vasconcelos, J. B. (2016). Maturity Models for Hospital Information Systems Management: Are They Mature?. In Innovation in Medicine and Healthcare 2015 (pp. 541-552). Springer International Publishing. doi: 10.1007/978-3-319-23024-5_49*

xvii. *Demchenko, Y., Zhao, Z., Grosso, P., Wibisono, A., & De Laat, C. (2012, December). Addressing big data challenges for scientific data infrastructure. In Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on (pp. 614-617). IEEE. doi: 10.1109/cloudcom.2012.6427494*

xviii. *Dinov, I. D. (2016). Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. Gigascience, 5(1), 1. doi: 10.1186/s13742-016-0117-6*

xix. *English, L. P. (1999). Improving data warehouse and business information quality: methods for reducing costs and increasing profits. John Wiley & Sons, Inc..*

xx. *European Union (2012). The management of health systems in EU Member States The role of local and regional. European Union 2012 ISBN: 978-92-895-0720-2*

xxi. *Fernandes L, O'Connor M, Weaver V: Big data, bigger outcomes. J AHIMA 2012:38–42.*

xxii. *Ferrera, M. (2008). Dal welfare state alle welfare regions: la riconfigurazione spaziale della protezione sociale in Europa. La Rivista delle politiche sociali, 3(2008), 17-49. doi: 10.3280/qu2014-004002*

xxiii. *Groves P. et all. (2013). Center for US Health System Reform Business Technology Office. The big data revolution in healthcare. McKinsey & Company, Gennuary 2013.*

xxiv. *Hu, L., Zhang, Y., Feng, D., Hassan, M. M., Alelaiwi, A., & Alamri, A. (2015). Design of QoS-Aware Multi-Level MAC-Layer for Wireless Body Area Network. Journal of Medical Systems, 39(12). doi: 10.1007/s10916-015-0336-x*

xxv. *Jarvenpaa, S. L., & Ives, B. (1990). Information technology and corporate strategy: a view from the top. Information Systems Research, 1(4), 351-376. doi: doi.org/10.1287/isre.1.4.351*

xxvi. *Lynch, C. (2008). Big data: How do your data grow?. Nature, 455(7209), 28-29. doi: doi.org/10.1038/455028a*

xxvii. *Li, J., Tao, F., Cheng, Y., & Zhao, L. (2015). Big Data in product lifecycle management. The International Journal of Advanced Manufacturing Technology, 81(1-4), 667-684. doi: 10.1007/s00170-015-7151-x*

xxviii. *Lovis, C., & Gamzu, R. (2015). Big Data in Israeli healthcare: hopes and challenges report of an international workshop. Israel Journal of Health Policy Research, 4(1), 1.*

xxix. *Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2015). A Data Quality in Use model for Big Data. Future Generation Computer Systems. doi: 10.1186/s13584-015-0057-0*

xxx. *Mezghani, E., Exposito, E., Drira, K., Da Silveira, M., & Pruski, C. (2015). A Semantic Big Data Platform for Integrating Heterogeneous Wearable Data in Healthcare. Journal of medical systems, 39(12), 1-8. doi:10.1007/s10916-015-0344-x*

xxxi. *Pasini, N. (Ed.). (2011). Confini irregolari. Cittadinanza sanitaria in prospettiva comparata e multilivello: Cittadinanza sanitaria in prospettiva comparata e multilivello. FrancoAngeli.*

xxxii. *Pearlson, K. E., Saunders, C. S., & Galletta, D. F. (2016). Managing and Using Information Systems, Binder Ready Version: A Strategic Approach. John Wiley & Sons.*

xxxiii. *Pontoppidan, C. A., & Brusca, I. (2016). The first steps towards harmonizing public sector accounting for European Union member states: strategies and perspectives. Public Money & Management, 36(3), 181-188. doi:10.1080/09540962.2016.1133970*

xxxiv. *Potters, L., Ford, E., Evans, S., Pawlicki, T., & Mutic, S. (2015). A Systems Approach using Big Data to Improve Safety and Quality in Radiation Oncology. International Journal of Radiation Oncology* Biology* Physics. doi:10.1016/j.ijrobp.2015.10.024*

xxxv. *Puddu, L. (Ed.). (2011). Elementi essenziali per la predisposzione e la certificazione del bilancio delle aziende sanitarie. Giuffrè Editore.*

xxxvi. *Puddu, Luigi; Rainero, Christian; Tradori, Vania; Secinaro, Silvana; Indelicato, Alessandra; Migliavacca, Alessandro; Brescia, Valerio. (2016). Risk management and healthcare: "separation" of revenues and expenditure. Risk management: perspectives and open issues. A multi-disciplinary approach. McGraw-Hill Education 248-265.*

xxxvii. *Priya, M., & Ranjith Kumar, P. (2015). A novel intelligent approach for predicting atherosclerotic individuals from big data for healthcare. International Journal of Production Research, 53(24), 7517-7532. doi:10.1080/00207543.2015.1087655*

xxxviii. *Raghupathi W: Data Mining in Health Care. In Healthcare Informatics: Improving Efficiency and Productivity. Edited by Kudyba S. Taylor & Francis; 2010:211–223. doi:10.1201/9781439809792-c11*

xxxix. *Rallapalli, S., Gondkar, R. R., & Ketavarapu, U. P. K. (2016). Impact of Processing and Analyzing Healthcare Big Data on Cloud Computing Environment by Implementing Hadoop Cluster. Procedia Computer Science, 85, 16-22. doi:10.1016/j.procs.2016.05.171*

xl. *Rockart, J. F., & Morton, M. S. (1984). Implications of changes in information technology for corporate strategy. Interfaces, 14(1), 84-95. doi:10.1287/inte.14.1.84* xli. *Sahama, T. R., Hofdijk, J., Hafen, E., Bignens, S., Goossen, W., Yasnoff, W., & Ball, M. (2016). IMIA working group on health record banking-How complex systems can cooperate towards having individual's consolidated data.*

xlii. *Schneeweiss, S. (2016). Improving therapeutic effectiveness and safety through big healthcare data. Clinical Pharmacology & Therapeutics, 99(3), 262-265. doi:10.1002/cpt.316*

xliii. *Shepherd, A. M., Laurens, K. R., Matheson, S. L., Carr, V. J., & Green, M. J. (2012). Systematic meta-review and quality assessment of the structural brain alterations in schizophrenia. Neuroscience & Biobehavioral Reviews, 36(4), 1342-1356. doi:10.1016/j.neubiorev.2011.12.015*

xliv. *Schick, A. G., Gordon, L. A., & Haka, S. (1990). Information overload: A temporal approach. Accounting, Organizations and Society, 15(3), 199-220. doi:10.1016/0361-3682(90)90005-f*

xlv. *Schmitt, J., Arnold, K., Druschke, D., Swart, E., Grählert, X., Maywald, U., ... Rüdiger, M. (2016). Early comprehensive care of preterm infants—effects on quality of life, childhood development, and healthcare utilization: study protocol for a cohort study linking administrative healthcare data with patient reported primary data. BMC Pediatrics, 16(1). doi:10.1186/s12887-016-0640-8*

xlvi. *Song, T. M., & Ryu, S. (2015). Big data analysis framework for healthcare and social sectors in Korea. Healthcare informatics research, 21(1), 3-9. doi:10.4258/hir.2015.21.1.3*

xlvii. *Strassmann, P. A. (1997). The squandered computer: evaluating the business alignment of information technologies. Strassmann, Inc.. doi:10.2307/20048812*

xlviii. *Suresh, S. (2016). Big Data and Predictive Analytics: Applications in the Care of Children. Pediatric clinics of North America, 63(2), 357-366. doi:10.1016/j.pcl.2015.12.007*

xlix. *Torchia, M., & Calabrò, A. (2016). Increasing the Governance Standards of Public-Private Partnerships in Healthcare. Evidence from Italy. Public Organization Review, 1-18. doi:10.1007/s11115-016-0363-1*

l. *Ward, J., & Peppard, J. (2016). The Strategic Management of Information Systems: Building a Digital Strategy. John Wiley & Sons.*

li. *Warwicka, W., Johnsona, S., Bonda, J., Fletchera, G., & Kanellakisa, P. (2015). A Framework to Assess Healthcare Data Quality. European Journal of Social & Behavioural Sciences, The, 13(2), 1730. doi:10.15405/ejsbs.156*

lii. *Wharton, M. K. (2016, June). Data Governance Structures for Integrated Disease Information Systems in Local Health Departments. In 2016 CSTE Annual Conference. Cste.*

liii. *World Health Organization, 2003, Improving data quality: a guide for developing countries, ISBN 92 9061 050 6*

liv. *Wu, J., Li, H., Cheng, S., & Lin, Z. (2016). The promising future of healthcare services: When big data analytics meets wearable technology. Information & Management, 53(8), 1020-1033. doi:10.1016/j.im.2016.07.003*

lv. *Wong, H. T., Yin, Q., Guo, Y. Q., Murray, K., Zhou, D. H., & Slade, D. (2015). Big data as a new approach in emergency medicine research. Journal of Acute Disease, 4(3), 178-179. doi:10.1016/j.joad.2015.04.003*

lvi. *Yang, C. C., & Veltri, P. (2015). Intelligent healthcare informatics in big data era. Artificial intelligence in medicine, 65(2), 75-77. doi:10.1016/j.artmed.2015.08.002*

lvii. *Zillner, S., & Neururer, S. (2015). Technology Roadmap Development for Big Data Healthcare Applications. KI-Künstliche Intelligenz, 29(2), 131-141. doi:10.1007/s13218-014-0335-y*