Revista de Bioética y Derecho
Perspectivas Bioéticas
www.bioeticayderecho.ub.edu - ISSN 1886-5887

ARTÍCULO

The Quality and Veracity of Digital Data on Health: from Electronic Health Records to Big Data

La calidad y veracidad de los datos digitales en salud: de la historia clínica a los datos masivos

La qualitat i veracitat de les dades digitals en salut: de la història clínica a les dades massives

RAFFAELLA BRIGHI *

OBSERVATORI DE BIOÈTICA I DRET DE LA UNIVERSITAT DE BARCELONA

La Revista de Bioética y Derecho se creó en 2004 a iniciativa del Observatorio de Bioética y Derecho (OBD), con el soporte del Máster en Bioética y Derecho de la Universidad de Barcelona: www.bioeticayderecho.ub.edu/master. En 2016 la revista Perspectivas Bioéticas del Programa de Bioética de la Facultad Latinoamericana de Ciencias Sociales (FLACSO) se ha incorporado a la Revista de Bioética y Derecho.

* Raffaella Brighi. Centro Interdepartamental de Investigación en Historia del Derecho, Filosofía y Sociología del Derecho e Informática Jurídica (CIRSFID) de la Universitá di Bologna, Italia. E-mail: raffaella.brighi@unibo.it.

Abstract

The quality of health information online depends on our ability to assess whether it is accurate, whether we are making this assessment as citizens/patients or whether we are using predictive software tools. There is a vast literature on the quality of health data online, and it suggests that the various tools for ensuring such quality are not fully adequate. I propose to address this problem by getting technological, organizational, and legal tools to work synergistically together. Integral to this vision—across all three elements—is the training needed for professionals delivering healthcare services as well as for patients using and generating health information online.

Keywords: digital health data; big data; quality of health data; veracity of data; provenance; XML standard.

Resumen

La calidad de la información de salud que podemos encontrar en línea depende de nuestra capacidad para evaluar si ésta es precisa o no, de si estamos haciendo esta evaluación como ciudadanos/pacientes o de si estamos usando herramientas de software de predicción. Existe una amplia gama de literatura sobre la calidad de los datos de salud que podemos encontrar por internet, y ésta sugiere que las diversas herramientas para garantizar una alta calidad de la información no son totalmente adecuadas. Propongo abordar este problema obteniendo herramientas tecnológicas, organizativas y legales para trabajar juntos y generar sinergias. Integrada a esta visión, a través de los tres elementos, es necesaria la formación de los profesionales que prestan servicios de atención médica, así como de los pacientes que usan y generan información de salud en línea.

Palabras clave: datos digitales de salud; big data; calidad de los datos de salud; veracidad de los datos; procedencia; XML standard.

Resum

La qualitat de la informació de salut que podem trobar en línia depèn de la nostra capacitat per avaluar si aquesta és precisa o no, de si estem fent aquesta avaluació com a ciutadans/pacients o de si estem fent servir eines de software de predicció. Existeix una àmplia gama de literatura sobre la qualitat de les dades de salut que podem trobar per internet, i aquesta suggereix que les diverses eines per garantir una alta qualitat de la informació no són totalment adequades. Proposo abordar aquest problema obtenint eines tecnològiques, organitzatives i legals per treballar junts i generar sinèrgies. Integrada a esta visió, a través dels tres elements, és necessària la formació dels professionals que presten serveis d'atenció mèdica, així com dels pacients que fan servir i generen informació de salut en línia.

Paraules clau: dades digitals de salut; big data; qualitat de les dades de salut; veracitat de les dades; procedència; XML standard.

1. Digital Health Data

Significant changes in our way of accessing knowledge have resulted from our increasing use of mobile devices, coupled with widespread access to bandwidth and to Web 2.0 services, without needing much technical expertise to that end. Technological applications do more than give us ready access to a greater amount of information: they give us greater power to process that information by enabling us to interface with multiple possible worlds. This changes our understanding of reality, and with the ability to share, aggregate, and process online data —using data-mining techniques to do business analytics and build predictive systems— we also gain an essential decision-making tool.

Even science is changing its methods of inquiry in this "data society." The data-intensive e-science that has developed with the advent of this society brings together a range of theories, simulations, and experiences in a complex of processes and systems aimed at extracting knowledge from data. In this way scientists collect digital data that they process, manage, and put through statistical analysis —a method distinctive enough to have suggested to some that we are looking at a new epistemological paradigm of scientific inquiry—.[1]

Because we no longer access knowledge in the same way as in the past, it becomes crucial to be able to assess the quality of data, and so the quality of the information that can be extracted from such data. As much as the words *data* and *information* are often interchangeably, they actually have different meanings in computer science. A *datum* represents a fact, phenomenon, or event through a series of symbols that need to be processed in order to make sense. Information, for its part, is what we get once that data is processed, giving us actual knowledge of something. Data, therefore, is not yet information, which can only be obtained by working on data to make it intelligible, and which in turn becomes useful as knowledge once we can actually *do* something with it.

Information, however, is useful only to the extent that the data on which it is based can be ascertained to be factual, making sure that the automatic processing of the data does not yield false information —especially when it concerns people, and particularly when we are looking at *Big data*, the huge and growing mass of data which cannot easily be quantified but which nonetheless travels across digital communications systems, and whose analysis ideally makes it possible to deliver services specific to each person—.

[1] See Jim Gray on eScience: A Transformed Scientific Method (2007).

The healthcare sector has not remained untouched by this information revolution,[2] because information technology is making inroads in the delivery and management of care, and patients and physicians are increasingly interacting digitally. Healthcare information systems have become increasingly personalized over the years,[3] as new the methods have developed for collecting and organizing patient data. We have moved from the Electronic Medical Record (EMR) —which were collections of clinical information stored at a single healthcare facility using formats and procedures still tied to the age of paper filing— to the Electronic Health Record (EHR), making it possible to share health data among different persons across multiple facilities, and in such a way as to cover the arc of a patient's medical history. This is also where the Personal Health Record (PHR) comes into play, making it possible to store and share a spectrum of critical healthcare information about a patient, who becomes the focus of an integrated care practice which involves the larger welfare system, and which attends to the health of patients broadly by also taking their nutrition and lifestyle into account.

Our own health data and information is something we have traditionally confined ourselves to *collecting*, nor has there been a consistent standard for doing so, with different people using the most disparate methods based on their own criteria. But with the advent of ICT tools such as the EHR and the PHR, we can move beyond plain collection to active management, and even to *activated* management,[4] in such a way that when someone needs some information, they can have it in a timely way and in a manner and a format that makes the it intelligible and interoperable, and hence truly useful.

Next to the initiatives taken by government entities, we are also seeing people becoming increasingly computer-savvy when it comes to their own health care: mobile technology and the Internet are enabling us to digitally access our own health information much more easily than before, and while we cannot be our own doctors or have full access all the information we might want, we are learning to look for that information and make an active use of it in trying to understand our own symptoms before we even seek professional help, rather than confining

[2] FLORIDI, L.: *La rivoluzione dell'informazione*, Codice Edizioni, 2012.

[3] This transformation is analyzed in detail in MAIOLI, C. and SÁNCHEZ JORDÁN, E.: "Big Data e capacità informativa per l'autodeterminazione del paziente", *Strumenti, diritti, regole e nuove relazioni di cura*, 2015, 155–76.

[4] Patients are "activated" (www.informationweek.com/healthcare) when they are fully involved in diagnosing their own condition and and making decisions about their therapy and care, relating to their physician not like a child to a parent (who certainly knows best) but like one does to a partner.

ourselves to the role we would traditionally have been stuck in as passive receivers of medical information and instructions.[5]

Around the patient, therefore, a care network takes shape that is built on digital personal data (in the form of interoperable EHRs), and which uses the Web 2.0 to advantage to enable this information to be cooperatively shared. The growing demand for information about health has spawned a great many websites (we cannot say exactly how many, but google any disease, and millions of results will come out), but it has also engendered a whole range of noninstitutional tools pertaining to health: tools for online medical advice (dedicated websites, blogs, social networks, and even Twitter feeds); resources devoted to the training of physicians and medical personnel (YouTube videos, platforms for sharing research, and the like); patient forums and groups (some of them devoted to a specific subject, others of general interest); personal areas in which to track one's own activity (with the use of sensors, among other tools) as well as one's own health and habits.

Corresponding to this wide variety of tools and services is an equally varied pool of users, manufacturers, and consumers of information who have different kinds of expertise and are driven by different aims. Thus, for example, the information we find on the Web may come from professionals in the field, and thus be thoroughly and reliably sourced, or it may be information put out by an emotional support group formed by patients in a social network. Sometimes the information is purveyed for profit, for the purpose of promoting a specific kind of treatment or hospital network, and so it may not be wholly unbiased. There is also a fair share of false or misleading information, such as quack cures having no scientific basis.[6]

---

[5] Many are the studies that have been done on our behaviour as online patients/consumers. See, for example, SQUILINI, R.: "Surfing the Internet for Health Information: An Italian Survey on Use and Population Choice," *BMC Med Inform Decis Mark*, 2011, 11ff.; BELL R.A. et al.: "Lingering Questions and Doubts: Online Information-Seeking of Support Forum Members Following Their Medical Visit", *Patient Educ Couns*, 2011, 525ff.

[6] One example, among many, involved the case of a Chinese student who trusted a paid advertisement promoted by the search engine Baidu. As a result of the incident, the Chinese government ordered the search engine to reduce by at least 30 percent the advertisements that show us as search results on each of its pages, and to completely do away with paid advertising bearing any connection to health.

2. Quality and *Veracity* of Digital Data

If on the one hand the copious supply of health services on the Internet *empowers* the patient, on the other hand this very abundance makes it more difficult to sift through all the information that comes up and to separate the good from the bad.

There is much research that has been devoted to this topic, raising doubts about the quality of the information that consumers of information (professionals or otherwise) will find on the Internet.[7] The healthcare sector is particularly critical in this respect, because erroneous information can clearly to great harm to people. Many medical studies have assessed the way health information shapes the way we choose to care for ourselves and the way we relate to physicians. Others have looked at how reliable the information is in specific areas, and for the most part the findings have not been encouraging.

The quality of information is therefore critical, and not only for citizens and patients: the health information we find on the Web and in the databases maintained by healthcare providers are a precious resource for public health.[8] The transition from paper records to EHRs facilitates the task of creating large databases of health data whose algorithmic analysis and processing (in keeping with data privacy regulations) can advance scientific research and even streamline the healthcare system as a spillover effect.[9]

Patients may not have the same needs as professionals, and, accordingly, different data-analytics systems will be aimed at different purposes, but regardless of the purposes of such consumers of information, the value of health information online is directly dependent on our ability to determine the *Veracity of information*, a criterion forming part of the broader measure of the *quality of data*, which tells us how reliable the information we have gathered is. Even so, the great volume, variety, and speed of big data can prevent us from selecting the data before we

---

[7] See PAOLINO L. et al.: *The Web-Surfing Bariatric Patient: The Role of the Internet in the Decision-Making Process* (Spinger, 2012); ASLANI A. et al.: "Web-Site Evaluation Tools: A Case Study in Reproductive Health Information", *e-Health for Continuity of Care*, IOS Press, 2014; HESSE, B. W. "Trust and Sources of Health Information: The Impact of the Internet and Its Implications for Health Care Providers", *Arch Intern Med*, 2005; and LAWRENTSCHUK, N. et al.: "Oncology Health Information Quality on the Internet," *Ann Surg Oncol*, 2012.

[8] See BRIGHI, R. and VIRONE, M. G.: "EHR and Usability of Health Data to Benefit Patients and Public Health," in *E-Health for Continuity of Care*, IOS PRESS, 2014.

[9] For a discussion of the legal issues involved, see HOFFMAN, S. and PODGURSKI, A.: "The Use and Misuse of Biomedical Data: Is Bigger Really Better?", *American Journal of Law & Medicine*, 2013.

analyse it and make decisions on that basis —all of which makes even more prominent the question of the *trust* that we can place in data—.

The problem of obtaining quality data is complex and cross-disciplinary. Over the years, several standards organizations have contributed to defining the quality of various products and services and identifying ways of measuring such quality.

One such measure is the ISO/IEC 9000:2015 Standard,[10] issued under the name Quality Management Systems: Fundamentals and Vocabulary: it lays out the basic quality concepts and language, defining *quality* itself as the "degree to which a set of inherent characteristics of an object fulfils requirements," where *requirement* is in turn defined as a "need or expectation that is stated, generally implied or obligatory."

We thus have a set of yardsticks that we can use to make a quality assessment, and specifically, where we are concerned, to measure the quality of data.

There are two kinds of indicators: *core indicators* apply to the data itself, and we can use them to measure whether the it is accurate, up to date, complete, and consistent, among other attributes, while *proxy indicators* apply to the source of the data and to its aims (whether commercial or for dissemination, for example), or to its readability, among other attributes.

Various proposals have been made for criteria on which basis to assess the quality of health data. The European Commission, for example, has set out six such criteria to serve as guidelines for all Member States and all EU bodies that publish health data. We thus ask: (1) Is the data being provided *transparently* and *honestly*? (2) Is its source *authoritative*? (3) Have *privacy* and *data protection* safeguards been put in place in giving access to the data? (4) Is the data being regularly *updated*? (5) Is the data provider *accountable* to its users? And (6) is the data easily *accessible* (is it easy to find, understand, and use)?[11]

These indicators measure in general the quality of the data itself and its sources. But no less important is the quality of the *model* used to represent the structure the data, as well as the *formats* in which the data is contained, which need to be standardized and interoperable, making it possible as well to source the data.

---

[10] The ISO 9000 series covers various quality management areas and contains some of the best-known standards. The standards in this series provide guidelines and tools for companies and organizations seeking to certify that their products and meet customer demands and that their quality is constantly improving.

[11] COM(2002) 667 del 29/11/2012, Quality Criteria for Health Related Websites.

3. Tools for Assessing Health Data

There are essentially two ways to go about reducing the risk of unreliable information: one is to teach users to judge the resources they find on the Web and filter out those that can't be trusted; the other is to use technology that will automatically validate the quality of the data. An impressive number of initiatives are being taken on both fronts, and just as numerous are the studies that have been carried out to measure the reliability of the assessment tools used in specific clinical areas.

What these initiatives all have in common is that they rely on codes of ethics and conduct that set out standards for putting out health information, a prominent example being the e-Health Code of Ethics,[12] whose focus is on making sure that health information is transparent and can easily be understood by users. (While the standards are in place, however, compliance with them is proving to be a challenge.)[13]

The tools available to date can be grouped into four types as follows:14

- *Self-regulation and self-governance codes*. These provide uniform rules and guidelines for those who put out health information. Adherence to these codes is often signalled by a label or service mark.[15]
- *Rating systems*. These tools help users assess health information on the basis of a questionnaire that yields a numerical measure of quality.[16]

[12] This code was developed in 2000 by a coalition of organizations, among which are the Health on the Net Foundation and Hi-Ethics.

[13] On this question see ZULLO, S. and DE PANFILIS, L.: "Aspetti etici delle applicazioni di eHealth", *Strumenti, diritti, regole e nuove relazioni di cura*, Giappichelli, 2015, 55–67.

[14] On this question see HANIFE, F. et al.: "The Role of Quality Tools in Assessing the Realibility of the Internet for Health Information", *Informatics for Health & Social Care,* 34(4), 2009, 231ff; FAHY, E. et al., "Quality of Patients' Health Information on the Internet: Reviewing a Complex and Evolving Landscape" *AJM,* 7(1), 2014, 24ff.

[15] Undoubtedly the most established of these is the HON Code of the Health on the Net Foundation (https://www.healthonnet.org/), which for twenty years has been setting standards for those who put out medical information. A website's compliance with the code can be checked automatically using a toolbar you can install on your Internet browser. Another well-known example is the code drafted by the *Journal of the American Medical Association* (JAMA), available on the website of the American Medical Association (ama-assn.org).

[16] An example is DISCERN, created in 1998 by the Division of Public Health and Primary Care at the University of Oxford. Another example, in Italy, is a 2008 initiative by the Ministry of Health called *Misurasiti*.

- *Expert audits*. These are carries out by third-party experts (physicians, burses, pharmacists, and the like) who make an independent assessment of the quality of information.[17]
- *Quality certifications*. These certification systems are managed by independent parties who will certify a provider of health information (usually for a fee) by looking at how well it complies with a set of well-defined standards.[18]

As a survey of the literature will reveal, however, these evaluation tools have not done much to improve the practice on the ground. Many of the programmes and initiatives have been short-lived, nor it is clear that they can accurately evaluate whether the information at issue is actually reliable, considering, too, that they often base their evaluation exclusively on proxy rather than core indicators. What is more, these tools are better suited to a *static* Web, such as Web 1.0 was, and are ill-equipped to deal with the dynamic interactivity of the current Web 2.0. Some of the studies that have been carried out focus on the tools for assessing websites devoted to specific diseases and medical conditions,[19] and even here the results have been disappointing.

For the big picture, however, we cannot neglect to also take into account the tools that users most commonly refer to when looking for information on the Web: Google and Wikipedia.

The order in which Google ranks the webpages in its search results shapes the way the user accesses information. Most users typically only look at the first search results, and some studies suggest that Google's page ranking does not correlate with quality of information.[20] Other studies, by contrast, have looked at Wikipedia —the world's most widely read online encyclopaedia, based on an open-editing model— finding that the accuracy and completeness of the information contained in it can be compared to that of any professionally edited encyclopaedia.[21] This finding is quite encouraging, for it suggests that even if a resource is not peer-reviewed, the user-generated content it makes available on the participatory model of the Web 2.0 can deliver a high standard of health information.

---

[17] These experts will even review health-related databases maintained by universities and research and nonprofit organizations.

[18] Among these are MEDCERTAIN (MedPICS Certification and Rating of Trustworthy Health Information on the Net) and OMNI (Organised Medical Network Information). Neither of them, however, is giving certifications any longer. Still active, by contrast, is the accreditation programme maintained by URAC (www.urac.org).

[19] See ASLANI et al., cit. ; LAWRENTSCHUK et al., cit.

[20] See FAHY et al., cit.

[21] RAJAGOLOPALAN, M. S. et al.: "Patient-Oriented Cancer Information on the Internet: A Comparison of Wikipedia and a Professionally Mainted Database", *J Oncol Pract,* 7(5), 2011.

What the literature suggests, all told, is that filtering tools, codes of ethics, and criteria used to either evaluate health information once it's already on the Web or to ensure a standard of practice in making the information available have not quite lived up to their promise, in part owing the sheer speed and variety of data that is being generated by the use of computer tools. If we want better-quality health data that is accurate and reliable, we should probably turn to another set of tools based on another model.

4. Trust in Data and Services

Given how pervasive and decentralized the Internet is, it is generally challenging to exercise any effective governance or oversight over the production of the information that winds up in it.[22]

Trust is also a concern in information technology, with the sharing of data in service-oriented systems,[23] as well as in the law,[24] and although existing frameworks are not fully suited to deal with the kind of information at issue, some principles can be laid out as jumping-off points.

Specifically, it is understood that healthcare cannot be treated as only a moral or an organizational problem but needs to be approached synergistically on different levels, turning to advantage the ability of Web 2.0 to facilitate cooperation and exchange.

The first level is that of technology, requiring tools and models with which to capture the formal characteristics of data and services on which basis to *automatically* assess their reliability —and to this end we can rely on the semantic Web—. The technological solutions need to be coupled with organizational ones on which basis to certify the data, in such a way as to increase the trust that can be placed in health information and provide legal solutions to the problem of technologically identifying digital health data and securing its authenticity and integrity.

---

[22] Interesting in this regard is the analysis contained in HANMEI, F. et al.: "How trust is formed in online Health Communities: A process Perspective", Communications of the Association for Information Systems, 34, 2014, advancing a dynamic model of trust on which we first assess the credibility of the information itself and then the credibility of the source, and in the process trust builds up between providers and consumers of information.

[23] See TOWNEND, P. et al.: *A Framework for Improving Trust in Dynamic Service-Oriented Systems*, IEEE, 2012, 136ff.

[24] See SARTOR, G.: "Privacy, Reputation, and Trust: Some Implications for Data Protection", *Trust Management*, LNCS, Springer, 2006.

5. The Provenance of Data on the Semantic Web

In parallel to the great advances that have been made in the technological ability to exploit the huge mass of unstructured data found on the Web, the researchers and organizations whose job it is to develop the Web are working on standards and models with which to formally express the *semantics of data*, or the meaning conveyed by the data, so as to make it possible to share, access, and integrate information otherwise broken up across a patchwork of different platforms.

We started out with Tim Berners Lee's semantic Web and his pioneering work in creating a Web whose *human understandable data* becomes *machine understandable*, and now our ability to share knowledge and ensure transparent data has advanced through the development of interoperable systems and machines capable of making reasoned decisions.[25]

If we are to establish trust in online information and services —in a context of growing amounts of information and increasingly distributed service applications— it becomes essential to be able to find out where the information comes from (its *provenance*) and how it has been produced.[26]

Provenance can be established by attributes such as who or what created the data, what its history is, who has modified it, and what its place and date is:

> *Digital Provenance: documentation of processes in a Digital Object's life cycle. Digital Provenance typically describes Agents responsible for the custody and stewardship of Digital Objects, key Events that occur over the course of the Digital Object's life cycle, and other information associated with the Digital Object's creation, management, and preservation.*[27]

Anything pertaining to provenance is classified as metadata, meaning data about the data itself. In representing such data and metadata we can rely on shared interoperable models and standards that make it easy to exchange data, in such a way that the (human or software) users of data can analyse it and decide how trustworthy it is.

---

[25] Semantic interoperability means that everyone needs to be able to understand the data in the same way regardless of how it is being processed.

[26] On this topic see MOREAU, L. et al., "The provenance of electronic data", *Comunications of the ACM*, 2008 51(4), 52–58; BERTINO E. et al.: *The challenge of assuring data trustworthiness*, Springer-Verlag, 2008.

[27] Data Dictionary for Preservation Metadata: PREMIS working group version 2.0, 2008.

Models and standards for codifying the provenance of data and make it shareable have been developed by several working groups, notably the pioneering Dublin Core,[28] with its work on the description of digital resources. Several initiatives have been launched,[29] and in 2013 they made it possible for W3C to define a set of models and standards grouped under the PROV Framework,[30] making it easier for heterogeneous systems to exchange information on the provenance of data. The PROV standard makes it possible to associate a resource with any kind of structured and detailed data describing the process through which the resource was generated and its links to other resources, bearing in mind that different people can have different needs and perspectives depending on their role (e.g., doctor and patient). Then, too, we can use semantic Web standards —among which XML, RDF, and OWL— to integrate PROV metadata and make it interoperable. These schemes can be processed by way of software that guides the use of content.

6. Computer Certification, Authentication, and Identification

Provenance provides an indirect certification of the quality of data. There are scenarios, especially in healthcare, that require a greater degree of trust. Consider, for example, remote medical consultation services or telemedicine tools for home healthcare sending information to healthcare providers, or again data collection for medical research. In none of these cases can technology alone deliver the trust needed for these transactions: the technology needs to work synergistically with other types of solutions providing stronger security guarantees.

On such solution lies in the organizational apparatus with which to *certify* data and metadata. Certification is a process for making sure that the data satisfies a specific set of criteria, with a certification body acting as an impartial third party. The quality of a certificate depends on the authoritativeness of the certifying body, which may in turn be accredited by an independent organism attesting to its credibility —or this credibility may rest on its reputation—. Examples of certification processes abound: some are mandatory, others voluntary (e.g., ISO certifications).

[28] http://dublincore.org/.

[29] In addition to the ones already mentioned, we have Provenir ontology, Provenance Vocabulary, Proof Markup Language, WOT Schema, SWAN Provenance Ontology, Semantic Web Publishing Vocabulary, and Changeset Vocabulary. See the PROV specifics document.

[30] W3C, PROV Overview, An Overview of the PROV Family of Documents, 2013.

Reliability in third-party medical certification depends on whether the certifying organism is qualified, but this requires investment and resources —for which reason the initiatives out there are few and short-lived—.

Next to traditional certification systems, there is also underway an effort to develop new paradigms based on the Web 2.0 principles of transparency and participative collaboration. A case in point is the ODI *Open Data Certificate*,[31] which could be tweaked to work effectively in healthcare.

The Open Data Certificate has been developed with a view to making the use, publication, and distribution of open data more rigorous and reliable. This is a three-stage process in which the open data publisher first obtains an auto-certification published on its website; then it can choose to have auto-certification validated by the scientific community; and then it can seek a third level of certification by the Open Data Institute.

The process thus uses the cooperation mechanisms typical of Web 2.0, and here we may have an initial solution for assessing the quality of health data: a multilevel certification involving Web users themselves (on the Wikipedia model), coupled with semantic structures describing the provenance of the data itself.

However, care must be taken to make sure that the certification process is rigorous, for a certificate may give consumers of information a false sense of trustworthiness.[32]

Furthemore, while a certification satisfies criteria of transparency, testifying to the quality of the data, it also places on the data publisher a responsibility it would not have if the data were published anonymously, and this may be another reason why these tools are struggling to gain traction in the healthcare sector.

Anonymity means that you cannot identify online the person who creates or modifies the information in question, and this is certainly a stumbling block in remote-access scenarios, where one is accessing information without any clear sense of who is making it available.

Digital identity is conspicuously a legal problem, for it forms the legal and technological basis of eGovernment services, where it is essential to be able to establish the identity of those to whom a service is delivered.

---

[31] This certification is promoted by the Open Data Institute, an independent nonprofit organization out of an initiative by Tim Berners-Lee and Nigel Shadbolt.

[32] BURKELL, J.: "Health Information Seals of Approval: What Do They Signify?", *Information, Communication & Society*, 7, 4, 2004.

In 2014 the EU issued Regulation (EU) No. 910/2014 —or eIDAS (short for electronic IDentification Authentication and Signature)— looking to establish a common platform through which citizens, businesses, and government could digitally interact securely in the EU. To this end a legal framework was set up for electronic signatures, electronic seals, electronic time stamps, electronic documents, electronic registered delivery services and certificate services for website authentication so as to promote trust in online services. Electronic identification tools are already in wide use to deliver government heathcare services in many countries, and the eIDAS framework is designed to also make cooperation possible across national borders. So, too, any tool by which to identify the persons behind a digitally transacted service would increase our trust in the information provided and would consequently improve the service —and that goes double in the healthcare sector, with online medical consultation and wherever a service involves an expert—.

The process of digitizing healthcare —through the use and transfer of electronic health records, digital medical reports and diagnostic images, and suchlike— depends for its legal validity on the use of electronic signatures and time-stamping to sign and file digital documents. So, for example, in the ministry guidelines issued in Italy on the basis of the eIDAS Regulation, different levels of legal validity are defined for health records.

Considering that health data can be put to *secondary uses* —for scientific research, for example, or to streamline healthcare services, including on an open data model— it is essential that the tools with which the data is processed can guarantee its integrity and provenance, even in keeping with criteria of legal validity.

7. Conclusions

The many attempts made in the medical community to provide guidelines and tools with which to ensure and assess the quality of online health data and services are no longer adequate in a society that spawns 2.5 quintillion bytes of data every day. The complexity of the online world —making it possible for different people with different interests to interact— eludes effective governance. A case in point is big data, for on the one hand big data carries the potential to transform healthcare by making it possible to integrate traditional health data with data that patients collect on their own, thus enabling the latter to play a more active role in their own care, but at the same time the data produced by healthcare providers is often perceived as a by-product of healthcare delivery rather than a tool for streamlining the same process.

As much as the problem is difficult to solve, some possibilities are open.

For one thing, we can work on the technological level, with tools and models making it possible to automatically validate health data and services through their formally expressed features —and the semantic Web can certainly do a lot here (especially by standardizing provenance metadata under the W3C PROV Framework)—.

But we can also work on an organizational level, developing data-certification models making for greater trust in health information, to this end also exploiting the cooperative principles of Web 2.0 (as in the case of the Open Data Certificate).

Finally, we need to work on a legal level in solving the problem of *digital identification* and of the *authenticity* and *integrity* of digitized health data.

There is one more point that needs to be stressed: the *training* needed for a cognizant use of the technologies in question, so that users can grasp their potential and appreciate their limitations, with political and social programmes designed to help patients develop the skills by which to look for, understand, and assess the online health information and put that knowledge to use in solving their medical issues.

Indeed, the apparent ease with which some technologies can be used today may lead some to overlook some critical aspects, especially where health data is concerned. Examples are the need to (a) protect personal data, often stored in public databases without appreciating the risks involved; (b) implement security policies for storing data and accessing services; and (c) understand when an interlocutor's stated identity is real.

But the training also needs to be done at every level: physicians need to be trained to use the technology properly, but no less important is the training needed for healthcare providers and administrative personnel, as well as for citizens and patients who use and sometimes generate content.

Bibliography

- ASLANI, A. et al.: "Web-Site Evaluation Tools: A Case Study in Reproductive Health Information", *e-Health for Continuity of Care*, IOS Press, 2014.
- BELL R.A. et al.: "Lingering Questions and Doubts: Online Information-Seeking of Support Forum Members Following Their Medical Visit", *Patient Educ Couns*, 2011, 525ff.
- BERTINO E. et al.: *The challenge of assuring data trustworthiness,* Springer-Verlag, 2008.

- BRIGHI, R. and VIRONE, M. G.: "EHR and Usability of Health Data to Benefit Patients and Public Health," in *E-Health for Continuity of Care*, IOS PRESS, 2014.
- BURKELL, J.: "Health Information Seals of Approval: What Do They Signify?", *Information, Communication & Society*, 7(4), 2004.
- FAHY, E. et al., "Quality of Patients' Health Information on the Internet: Reviewing a Complex and Evolving Landscape" *AJM,* 7(1), 2014, 24ff.
- FLORIDI, L.: *La rivoluzione dell'informazione*, Codice Edizioni, 2012.
- HANIFE, F. et al.: "The Role of Quality Tools in Assessing the Realibility of the Internet for Health Information", *Informatics for Health & Social Care,* 34(4), 2009, 231ff.
- HANMEI, F. et al.: "How trust is formed in online Health Communities: A process Perspective", *Communications of the Association for Information Systems*, 34, 2014.
- HESSE, B. W.: "Trust and Sources of Health Information: The Impact of the Internet and Its Implications for Health Care Providers", *Arch Intern Med*, 2005.
- HOFFMAN, S. and PODGURSKI, A.: "The Use and Misuse of Biomedical Data: Is Bigger Really Better?", *American Journal of Law & Medicine*, 2013.
- LAWRENTSCHUK, N. et al.: "Oncology Health Information Quality on the Internet," *Ann Surg Oncol*, 2012.
- MAIOLI, C. and SÁNCHEZ JORDÁN, E.: "Big Data e capacità informativa per l'autodeterminazione del paziente", *Strumenti, diritti, regole e nuove relazioni di cura*, Giappichelli, 2015, 155–76.
- MOREAU et al.: "The provenance of electronic data", *Comunications of the ACM*, 2008, 51(4), 52–58.
- PAOLINO L. et al.: *The Web-Surfing Bariatric Patient: The Role of the Internet in the Decision-Making Process*, Spinger, 2012.
- RAJAGOLOPALAN, M. S. et al.: "Patient-Oriented Cancer Information on the Internet: A Comparison of Wikipedia and a Professionally Mainted Database", *J Oncol Pract,* 7(5), 2011.
- SARTOR, G.: "Privacy, Reputation, and Trust: Some Implications for Data Protection", *Trust Management*, LNCS, Springer, 2006.
- SQUILINI, R.: "Surfing the Internet for Health Information: An Italian Survey on Use and Population Choice," *BMC Med Inform Decis Mark*, 2011, 11ff.

- TOWNEND, P. et al.: *A Framework for Improving Trust in Dynamic Service-Oriented Systems*, IEEE, 2012.
- ZULLO, S., DE PANFILIS, L.: "Aspetti etici delle applicazioni di eHealth", *Strumenti, diritti, regole e nuove relazioni di cura*, Giappichelli, 2015.

Fecha de recepción: 30 de julio de 2017
Fecha de aceptación: 10 de septiembre de 2017