

OPEN DATA: APPLICATIONS IN SOCIAL CARE AND HEALTH

Carlotta Galeone¹, Rossella Bonzi

Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy

Abstract. *The recent emergence of the open data movement and its rapid spread across the world are creating new challenges in the health and pharmaceutical field. The present study aimed at providing an overview of open data and their application in social and healthcare. We provided a definition of open data, and we discussed the current state of play in the world. Finally, we focused on the Italian situation pertaining to open health data.*

Keywords: *open data, linked data, open government.*

1. INTRODUCTION

The rapid technological progress that has characterised the latest decades has had a significant influence on humanity in general, on the economy and society, acting on lifestyles and irreparably changing people's daily lives.

This unprecedented technology development has led public institutions to promote open access to endless amounts of data. While once data were collected and kept protected in specialised data repositories protected by passwords and closed to public access because of legitimate concerns over privacy or national security, they are now considered a fundamental source of knowledge for citizens, as well as private and public organisations. Indeed, the synergy between free access to data and technological capabilities may boost the potential of huge databases to promote a massive social, political and economic change (Jetzek, 2016). This aim could be achieved through wide dissemination of social values such as transparency, sharing and teamwork.

Open data were initially introduced in the software implementation field but have now become fundamental for medical and pharmaceutical research.

1.1 DEFINING OPEN DATA

Even though there is no single and universally accepted definition of open data, a commonly accepted way to define this concept is the one proposed by Verhulst and colleagues:

¹ Corresponding author Carlotta Galeone, email: carlotta.galeone@gmail.com

“Open data is publicly available data that can be universally and readily accessed, used, and redistributed free of charge. Open data is released in ways that protects private, personal or proprietary information. It is structured for usability and computability” (Verhulst et al., 2014, p.12).

When defining open data, several key criteria must be respected. Some of these criteria were generally defined for ensuring the quality in statistics (Eurostat, 2003):

- *Relevance*: open data should be relevant, i.e. they should meet current and potential users' needs.
- *Accuracy*: open data should be accurate so that there is a high correspondence between the information they convey and the reality.
- *Timeliness*: open data must be characterised by punctuality in disseminating results in order to provide users with up-to-date information.
- *Accessibility and clarity of the information*: open data must be easily obtained and should be easily accessible and understandable to all users. Appropriate metadata and illustrations should be made available.
- *Comparability*: this concept is related to the comparison of open data across space and time; in other words, differences in data referring to different times or geographical locations must be due to real variations rather than to different statistical concepts, measures or procedures.

In addition to these criteria, open data should exhibit the following features:

- *Reliability*: the reliability of open data may be defined as the closeness of the original information to the subsequent data (see Ceolin et al. 2013 for the computation of the reliability of open data).
- *Integrability*: it is important that open data allow the integration of supporting information coming from different sources.
- *Periodicity*: they must periodically update the information they make available to different users.
- *Punctuality*: this feature refers to the period between the date of data release and the date scheduled by a calendar, by regulation or by prior agreement between partners.
- *Consistency*: it measures the adequacy of the statistics to be differently combined and for different use.
- *Reconciliation*: this is the characteristic open data must have so that the statistics within the same source – related to different variables, calculated on different domains, from different sources or processes with different periodicity – can be assimilated to each other.

Besides, open data must be accompanied by an information provision, necessary, in terms of documentation, to evaluate the quality of the data provided, and consequently improve it, and for full and correct exploitation of all the potential information. Users must be enabled to choose the most suitable sources to satisfy their cognitive needs and to understand to what extent the available data can be used for further processing. A detailed description of the statistical methodologies used, together with measurements of sampling and non-sampling errors, presents a high information content both for the users of the data themselves and for the producers.

Open data may be considered as an actual product, with defined features and a user license concerning its field of application. Since open data can be freely used, reused and distributed, they must undergo a license for accessibility to be accessible without legal or technological restrictions, require authorisation for use and reuse and redistribution without constraints and the possibility to quote the reference source.

It would also be useful to consider open data as a service and focus on the needs of the data user. In this way, open data would be effectively usable and qualifiable as a data source (Gurstein, 2013).

1.2 RAW DATA, LINKED DATA AND OPEN DATA

The term “raw data” refers to data in their original or atomic state, namely disorganised and disaggregated data. Raw data are available in lots of different formats and represent an incredible resource because of their vast unlocked potential. It would be useful to make “raw data” public to support the process of checking and re-analysing data.

Conversely, the term “linked data” refers to organised and connected data. These data are machine-readable, have a well-defined meaning, are linked to external data sets and are published on the Web (Bizer, Heath, and Berners-Lee, 2009).

A star rating classification was created to describe linked data: the basic level, which is represented by one star, refers to data available on the web in whatever format but with an open licence; two stars data are available as machine-readable structured data; three-star data require a non-proprietary format; four-star data use open standards from W3C to identify things; finally, five-star data are interconnected (Berners-Lee, 2009).

The Linking Open Data project is the most significant example of the enactment of the linked data principles (W3C SWEO Community Project, 2017).

It must be noted that “linked data” and “open data” are not equivalent terms. Even though linked data may be open, and vice-versa, it is also possible that linked data include restrictions such as a licence that prevents them being considered “open”. Similarly, open data may not meet all of Berners-Lee's linking rules (Miller, 2010).

2. OPEN DATA IN THE WORLD

Open data are necessary for governments because of their potential to empower important decisions at each level of society in several ways. Firstly, they may provide a concrete aid in fighting government corruption, through an enhanced accountability and government services; secondly, they may render governments transparent and participatory; thirdly, they may also foster journalism against the spread of fake news; last but not least, they may launch profitable businesses based on public sector data (Kundra, 2012). The first steps in this direction were taken by the Obama government that introduced the concept of Open Government, which is now a fundamental principle of government initiatives for about 70 countries worldwide. Open Government assumes that all government and administrative activities are open and available to boost practical actions and ensure widespread control on citizens' management of public assets. Transparency of information, participation and accountability are the cornerstones of this government line (Jetzek, 2016).

In the long-term, Open Government is likely to lead to an interactive and collaborative administrative model promoting citizens active participation and Public Administration's evaluation.

2.1 THE G8 OPEN DATA CHARTER

Open data are a crucial international issue because of their potential to create significant social and economic benefits around the world. For this reason, governments need to delineate shared principles in support of open data, to boost thriving and upstanding societies. Hence, in July 2013, G8 leaders² signed the G8 Open Data Charter (The International Open Data Charter, 2015), which outlined a set of five core open data principles:

- 1) *Open Data by Default*: this principle means all government data, namely those by all kinds of government bodies, or by the wider public sector, must be published openly and automatically.
- 2) *Quality and quantity*: this principle refers to high-quality, timely, comprehensive, and accurate open data. In other words, data should be published in their original form and fully described, and information must be written in plain, clear language to be understood by all.
- 3) *Universal accessibility*: this principle aims to release as much data as possible in open formats, ensuring that the data are available to the broadest range of users and for the most comprehensive variety of purposes.

² Canada, France, Germany, Italy, Japan, Russia, the UK and the United States of America.

- 4) *Releasing Data for Improved Governance*: this principle outlines the importance of sharing technical skills and experience among countries, and transparency on collection methods, standardisation and publishing processes.
- 5) *Use of open data to promote innovation*: this principle encourages the re-use of open data for commercial and non-commercial purposes, in order to support open data literacy.

Then, several areas of intervention were identified; among those, the healthcare sector plays a significant role. In this field, prescription data and performance data should be openly available.

2.2 OPEN DATA PORTALS

The most famous and well-structured open data portal is American (The U.S. General Services Administration), and was created in 2009 following Obama's "open government initiative".

In January 2010, Tim Berners-Lee launched British government open data website where the central government, local authorities and public bodies have been publishing a vast amount of public-sector data, for private or commercial reuse (The GOV.UK team, 2010).

At the end of 2012, the beta version of the open data portal of the European Community was launched (Unione Europea, 2018), and it currently provides access to more than 10,000 open datasets published by EU institutions and bodies. "My country in a bubble" is an interesting data visualisation within the EU Open Data Portal (Eurostat). It allows comparing data of different European countries for about 50 indicators, ranging from economy and finance, population and social conditions, transport, environment and energy, and many others.

Another noteworthy portal is the Base Adresse Nationale (BAN), i.e., French National Address Database (Direction interministérielle du numérique, 2017), which contains over 25 million freely accessible and downloadable geocoded addresses (with no personally identifiable data). It represents a thriving private-public partnership and a clear example of how open data can boost government efficiency.

2.3 THE OPEN DATA BAROMETER

The Open Data Barometer provides a global measure of the government use and publication of open data for accountability, innovation and social impact, through a user-friendly and information-rich website. (World Wide Web Foundation, 2008). In 2016, the 4th and latest Edition contained data from 115 countries.

The Leader Edition (World Wide Web Foundation, 2018) offers information about the 30 governments that have adopted the Open Data Charter and those that

have adhered to G20 Anti-Corruption Open Data. The participating countries were ranked from high to low-performance and grouped accordingly as “Champions”, “Contenders”, and “Stragglers”.

The “Champions” governments showed the highest total scores and a balance among scores on the sub-indexes of open data readiness, implementation, and impact. Canada and the UK were tied for first place with a total score of 76 out of 100. The other governments in this group were Australia, France, South Korea, Japan and New Zealand.

The USA, Germany, Uruguay, Colombia, Russia, Brazil, Italy, India, Argentina, Ukraine and the Philippines were included in the “Contenders” group. Finally, Chile, Indonesia, South Africa, Paraguay, China, Costa Rica, Turkey, Panama, Guatemala, Saudi Arabia and Sierra Leone were considered as “Stragglers”.

2.4 OPEN DATA IN HEALTHCARE AND PHARMACEUTICS

The G8 Open Data Charter strategic principles, which require the online and automatic publication of all administrative data according to specific quality and quantity criteria, aim to improve Governance and promote innovation. These principles may also be applied within the healthcare field.

The amount of health data freely available to the public has risen considerably in recent years. Wang and colleagues (Wang et al., 2019) conducted an exploratory study to investigate the current status of open health data. They examined 1942 datasets from the health catalogue on Data.gov and proposed to improve the overall usage of open data through some interactive functions and feedback mechanisms.

Bonacina (2016) conducted a literature review of Linked Open Data (LOD) in healthcare and subsequently examined the extent of the usage of LOD within the biomedical literature (Bonacina, 2019). The results of this study showed that an important concept related to open health data is the Resource Description Framework (RDF; Miller, 1998).

Furthermore, it is vital to ensure the interoperability of different resources to enhance health open data. For instance, Banda (2019) created a linked data version of the Observational Health Data Science and Informatics (OHDSI) vocabulary to connect it to other pre-existing linked resources. Several biodiversity data management platforms provide a substantial infrastructure, which is essential for the research community. LOD may be used as best practices for standardised data submissions and data citations to improve their interoperability (Droegge et al. 2019).

LOD may also be used in cancer research, as illustrated by Zehra and colleagues (2019): they provided an overview of three different cancer genomics

datasets and introduced a new data space within the LOD cloud intending to create a network of coherently linked datasets containing cancer genomics data and other bio-medical resources. The ReVeLD (a Real-time Visual Explorer and Aggregator of Linked Data) is another remarkable application coming from the cancer research setting. It is a user-centred visual analytics platform, which was devised to help researchers to flexibly mine the Linked Biomedical Data Sources (LBDS), and increase intuitive interaction with data from distributed sources (Kamdar et al., 2014).

It is worth mentioning the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT), which is a collaborative European project, coordinated by the European Medicines Agency (EMA) and GlaxoSmithKline Research and Development LTD (GSK), and consists of 34 public and private partners. It was created to strengthen the monitoring of the benefit-risk of medicines in Europe. The PROTECT website (PROTECT Project Management Support Team, 2017) gathers regularly updated drug consumption databases.

Another example of open data in the pharmaceutical field is represented by the Drugle (KI/Seibo Software Studios, 2009). The Drugle is a semantic search engine, which lists and analyses information on medicines available on the Internet through a simple query. It can be combined with other healthcare applications and systems. LOD may help researchers facing challenges related to the need for aggregated data for drug discovery, drug repurposing and drug safety (Kamdar et al., 2019).

3. OPEN DATA: THE ITALIAN SITUATION

The Open Knowledge Network runs the Global Open Data Index (GODI) measures and benchmarks the openness of government data around the world. It offers valuable insights for government's data publishers to detect potential areas of data gaps. In 2016, Italy ranked 32nd against other places in the Index, and obtained a score of 47% in the ranking score, based on weighted questions, displayed as a percentage of the maximum possible score and 26% fully open datasets (according to the Open Definition, Open Knowledge Foundation).

Some thematic areas such as Procurement, National Maps, Company Register, Locations, Water Quality and Government Spending lagged behind and scored 0%, meaning that no dataset was completely open. The methodology used in the Global Open Data Index changed significantly between 2015 and 2016; therefore, the results are not directly comparable, and it would be pointless to monitor the Italian performance over time. However, it is reasonable to assume that the 2016 score may

improve in the future. Indeed, according to the latest “Open Data Maturity” survey (whose data refer to the period between July 2016 and June 2017) of the European Data Portal, Italy was ranked among the Trendsetters, while the previous year it was among the followers. This means that Italy was able to implement an advanced and well-coordinated open data policy among all the players operating on the national territory and which have a national open data portal with advanced features. In particular, Italy was ranked 7th thanks to 56 open data portals, before Germany and the Netherlands (Mastroianni, 2018).

Conversely, the Open Data Barometer-Leaders Edition (World Wide Web Foundation, 2018) highlighted some weaknesses: Italy obtained a total score of 50 (out of 100) and was grouped within the “Stragglers”. Indeed, the Italian government and citizens should collaborate in a more proficient way to make open data work for people. Despite some shortcomings, open data initiatives in Italy are growing.

Since October 2011, public administration (PA) data in Italy can be found on the portal www.dati.gov.it, which was created to organise the PAs data-sharing process and foster innovation and transparency. To date, this website gathers more than 20,000 datasets from 76 different administrations, including free downloadable geographical data. It must be noted that this portal is not comprehensive of all the Italian available open data: for example, there are only 8 health datasets from Lombardy, even though 52 datasets can be found on the Lombardy open data portal (Regione Lombardia). Open data guidelines were outlined to standardise methods for publishing data from various administrations and maintain a national portal (Formez PA, 2011).

The vademecum is composed of two macro areas:

- Part I) PA and open data: it introduces the Open Government and Open Data concepts, and provides insights on the regulatory framework;
- Part II) How to proceed in opening PA data: it describes several practical, organisational and legal aspects to consider before publishing such data.

It must be noted that PAs, in line with the principles of the digital administration's new code, are responsible for making their data available openly and digitally. In order to promote the spread of open data and make the re-use of public information easy and immediate, a specific licence agreement, the Italian Open Data License (IODL) v 2.0, has been designed (Agenzia per l'Italia Digitale, 2017).

At a regional level, several initiatives have been taken over the last few years. To date, the largest number of regional downloadable dataset comes from Tuscany (over 1,400), followed by Lombardy (over 1,200). The Piedmont Region was the first Italian region to launch an open data portal (Regione Piemonte) with more than 800 datasets from single municipalities and provinces. The Emilia-Romagna

Region followed its example, making available an open data catalogue containing 390 datasets (Regione Emilia Romagna).

Despite legislation supporting the free use, reuse and distribution of data, open data circulation in Italy is still limited. Only 37% of Italian Municipalities publish data in an open format. The main challenge seems to be overcoming the perception of open data as a regulatory obligation rather than a real opportunity. 80% of Italian Municipalities cannot see any positive impact in publishing open data, and 55% of them consider open data of little use for the growth of the business sector. Low-quality data, access barriers, and the lack of unitarian national usage hamper the open data culture (Osservatori Digital Innovation, 2018).

3.1 OPEN HEALTH DATA IN ITALY

The Italian Ministry of Health created a dedicated data-store (Ministero della Salute, 2011) containing 44 national datasets related to healthcare. However, 259 datasets are freely accessible in the health section of the national open data portal, including the regional or municipal data.

Among the most downloaded datasets, it is possible to find:

- the complete list of pharmacies;
- the complete list of parapharmacies that are allowed to sell medicines to the public (Article 5 of Decree-Law 223/2006 - commercial establishments);
- the complete list of medical devices registered in the Directory of the Ministry of Health;
- the complete list of authorised logistic sites to distribute medicinal products for human and animal use on the national territory (art.100 D.Lgs. 219/2006, already Legislative Decree 538/92).

In addition, many noteworthy health datasets come from structures such as the Local Health Authority (ASL: Azienda Sanitaria Locale), the ASL correspondence with municipalities, family counselling centres, hospitals, university hospitals and “Scientific Institutes for Research, Hospitalisation and Healthcare” (IRCCS: Istituto di Ricovero e Cura a Carattere Scientifico); datasets concerning health professionals such as health managers and staff with a flexible working relationship between ASL and hospitals, divided by role.

The “Health Advisor” application was born to create a virtual community to allow users to evaluate their personal experience with the National Health Service. It must be pointed out that even though the idea is innovative, the lack of related open data makes it challenging to work appropriately. The performance of the Health Service can also be assessed through the available National Outcomes Program open data: it offers comparative evaluations at a national level based on

multiple indicators of the safety, efficiency and quality of the treatments provided by the National Health System.

As the number of open data on healthcare is still unsatisfactory, it is clear that the undertaken digitisation process will contribute to the improvement of information capacity through the integration of data. However, Italy is still struggling to start a phase of open data massive use.

Other interesting examples are the list of the 50 best sellers non-prescription drugs of the last semester and data related to the available accredited beds of the National Health System, within public hospitals, care centres and hospices or private and certified nursing homes.

Regarding health applications of open data, several Italian initiatives may be listed. Firstly, the “Pronto Soccorso Lazio Ospedali” app, was created using data from the Open Data Lazio portal (Regione Lazio) and provides data from various first aid units. The hospitals within the Lazio Region are displayed on a map, and the real-time number of people inside the Emergency Room, as well as their triage colour code, is shown for each hospital.

Similar applications are available for other Italian areas. In Veneto, there is an app that distinguishes the type of first aid unit (e.g. generic, paediatric, gynaecological).

The introduction of the electronic medical prescription bolstered the automatic transmission of prescription data between general practitioners and pharmacies, leading to the dematerialisation of more than 29 million prescriptions (Federfarma).

In Sicily, despite the lack of an institutional open data portal (but see the non-official one: nonportaleopendatasicilia.it), approximately 87% of medical prescriptions were dematerialised.

The autonomous province of Bolzano and Calabria are in the last position.

In order to deepen the knowledge of the “drug prescription” phenomenon, the Videofar website was launched (EpiCentro). This computerised drug prescription monitoring system offers the possibility to analyse the trends over time of different classes of drugs and to differentiate among regions (Raschetti, 1991).

Despite these encouraging results, the health sector remains the weak point in the open data context, and much has still to be improved (World Wide Web Foundation, 2008).

It must be stressed that not all the Italian regions have embraced the open data culture, and the lack of concerted actions adversely affects the open data networking. Considering the case of people living in region border areas, if First Aid real-time data are available and inter-connected but for a limited number of regions, the usefulness of the data is insufficient and makes health services ineffectual. Hence,

a well-coordinated national policy on open data is needed. Public awareness is necessary to acknowledge the potential of information, especially in the healthcare field.

However, in 2012 the Health Ministry website was launched to support the publication of health system data (Agenzia per l'Italia Digitale, 2012). This was a good omen, and a significant number of downloads of the available datasets was registered. Though, to date health data represent a minimum part of approximately 23,000 Italian open datasets.

The currently insufficient spending commitment encumbers the sector digitisation.

Federsanità highlighted that the primary care management software, which is present in almost all the General Medicine and Pediatricians studies, is not integrated with the information systems of the hospitals and the territory, in which the spread of the electronic medical record is still minimal. Although the aim was to activate the Individual Health Card (FSE: Fascicolo Sanitario Elettronico) for at least 70% of Italian citizens within 2020, to date, only 11 Regions have successfully passed the interoperability tests with the central platform, where the FSE services provided are between 90 and 100% (Anitec-Assinform, 2018).

Finally, the “Spaghetti Open Data” mailing list provides constant updates about this dynamic topic.

The challenge is to transform information into knowledge in the healthcare field. The aim is to interconnect as many open data coming from different sources as possible and summarise this new information.

The new evidence is not likely to replace the knowledge gained through ad hoc studies and cannot produce data with the same level of precision and accuracy. However, this complementary evidence may be useful to fill in some gaps of complex and multifaceted public health theme. The integration of data obtained thanks to the spread of open data culture may contribute to the improvement of professional databases.

In our opinion, there are at least three reasons for the distinction to remain: firstly, open data are mainly collected for administrative purposes rather than for research purposes; secondly, there is still no professional or business logic to orientate the use of open data; thirdly, most of the open data criteria are not respected.

Furthermore, some professional databases that today benefit from the difficulty of accessing administrative data are meant to disappear or be transformed through adequate integrations with original and research data.

Ultimately, the growing importance and usefulness of open data are

indisputable, and it is plausible drawing future scenarios that describe an ever-increasing interest in them by companies, institutions and citizens. The availability of public sources has started to revolutionise the world of data, their analysis and their interpretation (Kitchin, 2014). In fact, according to the “Open Data Charter”, which comprises Italy, even prescription data could and should become open, with consequences on the pharmaceutical industry and research. However, there are obvious limitations such as the possibility that the data structure changes every time an update is published; an update rate that cannot keep up with the pace of professional use of the data; the purpose of collecting different information that barely coincides with users' needs; the difficulty in finding unique and common encodings between different open data and resulting connection issues.

4. CONCLUSIONS

The present review provided an overview of the current situation of open data around the world, with a particular focus on Italy. Overall, there has been a considerable rise in the amount of open data, which are increasingly available because of the high number of free web portals and datasets. Nevertheless, not so many datasets are entirely open. Many barriers exist and negatively affect the spread of open data, for instance: the inadequate and untimely update of data; the possibility that the data structure changes when an update is published; the difficulty to find unique and standard coding with a consequent expensive and not always feasible matching activity (Bacon and Goldacre, 2020).

The limited availability of completely open data affects particularly public health data all over the world without distinctions related to the country's level of economic and technological development (World Wide Web Foundation, 2018).

Some successful initiatives were presented to demonstrate the potential of open data. However, the healthcare data ecosystem is still mostly closed because of several complex reasons, for instance, the inappropriate level of updating and the need for consistency between the data considered and the accuracy and completeness. The lack in one of the critical data features may account for ineffective service.

In conclusion, government-held data need to be open by default and follow the principles set out in the G8 “Open Data Charter”. Besides, governments must maintain their commitments to open data to avoid backsliding. The improvement in the quantitative and qualitative terms of open data and the use of such data with an in-depth knowledge of their limits is essential to create new and useful public knowledge and to allow decision-makers to understand the territory and provide citizens with a powerful tool for participation and control.

REFERENCES

- Agenzia per l'Italia Digitale (2012). *I dati aperti del Ministero della Salute*. <https://www.dati.gov.it/content/dati-aperti-ministero-salute>. Last access: 10/02/2020.
- Agenzia per l'Italia Digitale (2017). *Italian Open Data License v2.0*. <http://www.dati.gov.it/iodl/2.0/> Last access: 10/02/2020.
- Anitec-Assinform (2018). *Il digitale in Italia 2018. Sanità*. <http://ildigitaleinitalia.it/il-digitale-in-italia-2018/il-mercato-digitale-italiano-2017-2020/consuntivi-e-previsioni-i-settori/sanita.kl> Last access: 10/02/2020.
- Bacon, S. and Goldacre, B. (2020). Barriers to Working With National Health Service England's Open Data. *Journal of Medical Internet Research*, 22(1), e15603.
- Banda, J. M. (2019). Fully connecting the Observational Health Data Science and Informatics (OHDSI) initiative with the world of linked open data. *Genomics & informatics*, 17(2), e13. DOI: 10.5808/GI.2019.17.2.e13
- Berners-Lee, T. (2009). *Linked Data*. <https://www.w3.org/DesignIssues/LinkedData.html> Last access: 10/02/2020.
- Bizer, C., T. Heath and T. Berners-Lee (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22.
- Bonacina, S. (2016). Linked Open Data in Health and Clinical Care A Review of the Literature. *European Journal of Biomedical Informatics*, 12(2), en2-en11.
- Bonacina, S. (2019). Linked Open Data in the Biomedical Information Area: A Keywords Analysis. *Studies in health technology and informatics*, 264, 1429-1430.
- Ceolin, D., Moreau, L., O'Hara, K., Schreiber, G., Sackley, A., Fokkink, W., van Hage, R. and Shadbolt, N. (2013). Reliability analyses of open government data.
- Direction interministérielle du numérique (2017). *Le site national des adresses*. <https://adresse.data.gouv.fr/> Last access: 10/02/2020.
- Droege, G., Karsch-Mizrachi, I., Barker, K., Coddington, J. and Seberg, O. (2019). Towards Linked Open Molecular Data: Recommendations for researchers, collections, infrastructures and publishers. *Biodiversity Information Science and Standards*.
- EpiCentro. *Osservatorio sui farmaci. La banca dati sui farmaci: VideoF@r*. <http://www.epicentro.iss.it/farmaci/videofar> Last access: 10/02/2020.
- Eurostat (2003). *Assessment of quality in statistics - Definition of Quality in Statistics*, Working Group, Luxembourg.
- Eurostat. *My Country in a bubble EU*. <https://ec.europa.eu/eurostat/cache/BubbleChart/?lg=en> Last access: 10/02/2020.
- Federfarma. <https://www.federfarma.it/> Last access: 10/02/2020.
- Formez PA (2011). *Linee guida per i siti web delle PA. Open Data. Come rendere aperti i dati delle pubbliche amministrazioni*. <https://www.istat.it/it/files/2013/01/vademecumopendata.pdf>.
- Gurstein, P. (2013). Social Equity in the Network Society: Implications for Communities. In N. Carmon, N., and S. S. Fainstein, editors, *Policy, Planning, and People: Promoting Justice in Urban Development*, 161-182.
- Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program. *Government Information Quarterly*, 33(1), 89-104.
- Kamdar, M. R., Fernández, J. D., Polleres, A., Tudorache, T. and Musen, M. A. (2019). Enabling Web-scale data integration in biomedicine through Linked Open Data. *NPJ digital medicine*, 2(1), 1-14.

- Kamdar, M. R., Zeginis, D., Hasnain, A., Decker, S. and Deus, H. F. (2014). ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research. *Journal of biomedical informatics*, 47, 112-130.
- KI/Seibo Software Studios (2009). *DRUGLE - A semantic search engine for pharmacology and medical drug information*. <http://drugle.se/> Last access: 10/02/2020.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
- Kundra, V. (2012). *Digital fuel of the 21st century: Innovation through open data and the network effect*. Joan Shorenstein Center on the Press, Politics and Public Policy.
- Mastroianni, F. (2018). *Nel mondo 2700 portali Open Data. La rimonta dell'Italia. Il Sole 24 ore*. <https://www.infodata.ilssole24ore.com/2018/09/20/nel-mondo-2700-portali-open-data-la-rimonta-dellitalia/>. Last access: 2020.
- Miller, E. (1998). An introduction to the resource description framework. *Bulletin of the American Society for Information Science and Technology*, 25(1), 15-19.
- Miller, P. (2010). Linked Data and Government. ePSI Topic Report nr. 7. Brussels, *The ePSI Platform*.
- Ministero della Salute (2011). *Open data*. <http://www.dati.salute.gov.it> Last access: 10/02/2020.
- Open Knowledge Foundation. *Open Definition 2.1*. <http://opendefinition.org/od/2.1/en/> Last access: 10/02/2020.
- Osservatori Digital Innovation (2018). *Gli open data non decollano in Italia*. https://www.osservatori.net/it_it/osservatori/comunicati-stampa/open-data-in-italia-non-decollano Last access: 10/02/2020.
- PROTECT Project Management Support Team (2017). *The Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium*. <http://www.imi-protect.eu/index.shtml> Last access: 10/02/2020.
- Raschetti, R. (1991). VIDEOFAR: an epidemiologic tool for drug utilization studies. *Annali dell'Istituto superiore di sanità*, 27(2), 191-3.
- Regione Emilia Romagna. *Open Data Emilia Romagna*. www.dati.emilia-romagna.it Last access: 10/02/2020.
- Regione Lazio. *Open data*. <https://dati.lazio.it> Last access: 10/02/2020.
- Regione Lombardia. *Open data*. www.dati.lombardia.it Last access: 10/02/2020.
- Regione Piemonte. *Dati Piemonte*. www.dati.piemonte.it Last access: 10/02/2020.
- The International Open Data Charter (2015). *Principles*. <http://opendatacharter.net> Last access: 10/02/2020.
- The GOV.UK team (2010). *Find open data*. www.data.gov.uk Last access: 10/02/2020.
- The U.S. General Services Administration. *The home of the U.S. Government's open data*. <https://www.data.gov/> Last access: 10/02/2020.
- Unione Europea (2018). *Portale Open Data dell'UE*. <https://data.europa.eu/euodp/it/> Last access: 10/02/2020.
- Verhulst, S., Noveck, B. S., Caplan, R., Brown, K. and Paz, C. (2014). The open data era in health and social care. *NYLS Legal Studies Research Paper*, (2563788).
- Wang, Y., Zhang, H., Wu, D. and Chen, J. (2019). Health data on data. gov: A research on status quo of open health data. *iConference 2019 Proceedings*.

-
- World Wide Web Foundation (2008). *The Open Data Barometer*. https://opendatabarometer.org/?_year=2017&indicator=ODB Last access: 10/02/2020.
- World Wide Web Foundation (2018). *Open Data Barometer - Leaders Edition*. Washington DC: World Wide Web Foundation.
- W3C SWEO Community Project (2017). *Linking Open Data*. <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData> Last access: 10/02/2020.
- Zehra, D., Jha, A., Khan, Y., Hasnain, A., d' Aquin, M. and Sahay, R. (2019). A Cancer Genomics Data Space within the Linked Open Data (LOD) Cloud.