

Open Data in the Health Context

The Lombardy Region Experience

Luca Augello², Antonio Barone², Daniele Crespi², Michele Ercolanoni², Ferdinando Ferrari¹, Matteo Giacomo Jori³,
Luca Merlino¹, Simone Paolucci², Francesca Romana Rossi²

¹ Regione Lombardia, Italy

² Lombardia Informatica S.p.A., Milano, Italy

³ Università degli Studi di Milano, Italy

Abstract

The scope of this paper is to describe, within the complex national and regional legal context with respect to Open Data, the methodologies and opportunities for the use of such data in the health context in the Lombardy Region.

Keywords

Administrative data, Healthcare quality, Data re-use, Data protection

Correspondence to:

Antonio Barone

Lombardia Informatica S.p.A.

Address: Via Taramellii 26, 20124 Milano, Italy

E-mail: antonio.barone@lispia.it

EJBI 2016; 12(2):en12-en19

received: July 19, 2016

accepted: December 16, 2016

published: December 31, 2016

1 Introduction

In the digital age, data and information form a key resource for any social or commercial activity. A simple query for the nearest hospital or late-hour pharmacy requires access to data held and made available centrally, whose usage may allow the generation of innovative services of added-value.

Furthermore the availability of Open Data transforms the citizen from a passive subject into a person who can interact with foreknowledge and so be more greatly involved in the decisions that concern him/her. This represents a value that goes beyond mere transparency, giving the citizen the possibility of bringing his own contribution.

In a “democratic” system, the citizen should be able to know and share with others the policies and activities of the relative public administration. Transparency is not merely a question of access to data, but the possibility of sharing and re-using it.

2 Legal background

The origin of the concept of Open Data, and in particular, Open Government Data (Open Data in the public sector), that is making public sector data accessible and usable for whatever purpose, is usually associated with the legal system of North America. In fact, the first processes

of public sector data availability towards citizens were developed in 2009 in the United States. The US experience first influenced the UK government, and then, in a short time, the public administrations of other important countries, like Canada and New Zealand. The readiness of common law legal systems to appreciate the potential of open government data is directly derived from the relationship between citizens and the public administration in Anglo-Saxon countries, traditionally based on the generalized access to information (“any person has a right, enforceable in court, to obtain access to federal agency records”, as stated by the United States Freedom of Information Act of 1966 and in Directive 2003/98/EC on the re-use of public sector information, otherwise known as the PSI Directive. PSI Directive is an EU directive that encourages EU member states to make as much public sector information available for re-use as possible. For a more profound analysis of the directive and the principles on which it is based, see also [1–3].

Legal systems of continental Europe, on the other hand, traditionally consider access to public administration data as an exceptional event, justified only by specific circumstances, like for instance the non-observation of a citizen’s right, and the consequent need to access the documents in question. It is not therefore surprising that the first approach to the subject of usage of public sector information, proposed by the European Community directive 2003/98/CE, was undeniably of great importance,

but demonstrated a certain timidity in the definition of the objectives and instruments to enable such usage. Even though the directive was formulated with the specific intent of encouraging the availability of public data, it did not contain any obligation towards the member states to implement such availability, and let them require a specific request and fee payment for information access. Thus section 1, paragraph 2 of the Italian implementation of the European directive, Decree 36/2006, defined that public administrations could choose “to authorize the re-use of documents containing public data collected, produced and reproduced for institutional purposes”. This principle was partially balanced by section 1, paragraph 4 of the decree, according to which the goal of making information re-usable should be pursued according to methodologies that would guarantee the conditions of equality, adequacy and non-discrimination.

A rather more important step towards the implementation of a real open data system was made with the next directive, 2013/37/EU (which modified the aforementioned 2003/98/EC). This directive, while still allowing nations to locally limit access to some information, and guaranteeing in any case the laws of personal data protection, and intellectual and industrial property, defined an obligation towards the member states to consent re-use of public sector information. In conformance with the new policy defined by the European directive, Italian decree 33/2013, aka “the transparency decree”, introduced into the Italian legal system the obligation towards public administrations to consent the re-use of information that must be made freely accessible according to the current transparency laws, in order to guarantee a correct government functioning. The decree, recognising the citizen as having specific rights with respect to transparency, defined a “right to access administrative data” that includes the possibility of using such data for purposes over and above those for which the data was originally collected.

Decree 33/2013 is not the only source of regulations that deal with open data access. The Italian Digital Administration Regulations (decree 82 of 7 March 2005), for example, defines important criteria for the correct definition of the data to be made available, and the instruments, both technical (data format) and legal (licensing), to use and transport the information.

The sharing of open data is considered to be part of a virtuous cycle, able to add value to the information assets of both private companies and public bodies, both in terms of research and development, and under the profile of market opportunity and the strengthening of new synergies, made possible by a more dynamic usage of databases. Therefore, in the context of open government data, with the objective of transparency, made possible by a greater monitoring by users of the actions of public administrations, advantages may be obtained that derive from the possibility of citizens to participate in the management of public resources, through information access and open data re-use.

However, the obligations of transparency and accessibility cannot be considered indiscriminate. The noble purposes of the doctrine of openness must necessarily be conditioned by other values guaranteed as constitutional rights, and in particular by the right to privacy and protection of personal data.

3 Open data and the protection of personal data in the health context

One of the most frequently recurring themes related to the implementation of open data is the possible conflict of such processes of data management and sharing with data protection laws. Without doubt, at least theoretically, the availability and re-use of information related to citizens could be in contrast with the rigorous limitations, defined by Decree 196/03 (Italian Personal Data Protection Regulations) and European Regulation EU 2016/679 (General data protection regulations), on the treatment of personal data. The above considerations are therefore particularly relevant if applied in the health sector, where databases generally contain sensitive data, in that it may reveal information relative to an individual’s state of health.

The importance of defining a correct compromise between openness of information assets and the protection of personal data is evident, both in the policies of the European Union and those of the European data protection authorities. At National level, the personal data privacy authority drew up: “Linee guida in materia di trattamento di dati personali contenuti anche in atti e documenti amministrativi effettuato da soggetti pubblici per finalità di pubblicazione e diffusione sul web” of 2 March 2011 (in G.U. 19 marzo 2011, n. 64; doc. web. n. 1793293), and then “Linee guida in materia di trattamento di dati personali, contenuti anche in atti e documenti amministrativi, effettuato per finalità di pubblicità e trasparenza sul web da soggetti pubblici e da altri enti obbligati” of 15 May 2014, (in G.U. 12 giugno 2014, n. 134; doc. web n. 3134436). The main principle of the current legislation is that, independently of the goals pursued (including therefore those of transparency that regulate public administrations), whenever the publication of data, information and documents involves a treatment of personal data, such goals must take into account the rights, the basic freedoms and the dignity of the subjects involved.

Both of the directives 2003/98/EC and 2013/37/EU on information re-use took into consideration the obligations of protection of personal data, but the “Transparency Decree” 33/2013, at national level, defined more clearly the application of the two directives. The general premise of section 1, paragraph 2, according to which the goals of transparency must be pursued, in all cases, guaranteeing the respect of laws concerning State secrets, official secrets, and personal data protection, finds its implementation in the section 4 of the decree, which, in

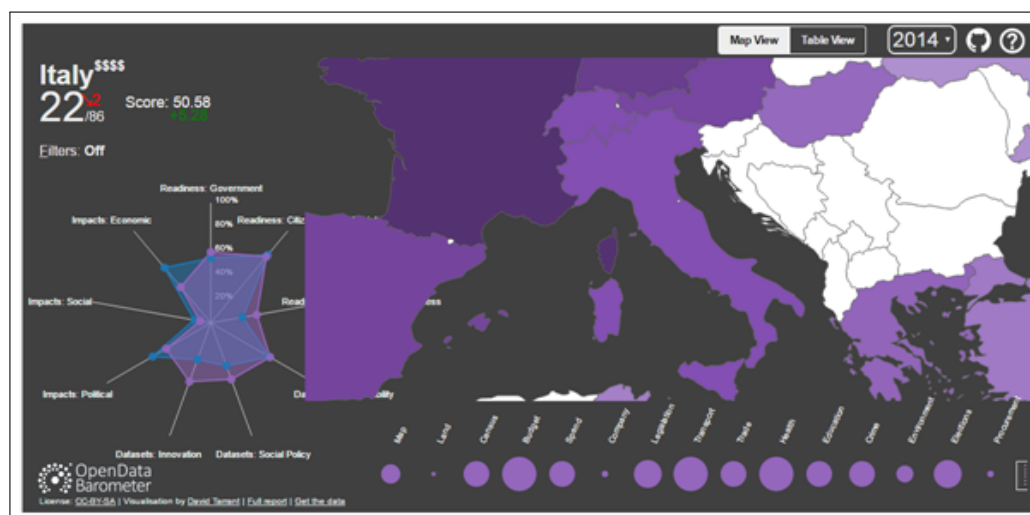


Figure 1: Open Data Barometer.

eight paragraphs defines the specifications of the “Limits to transparency”. Amongst these, paragraph 1 is of particular importance. It states that personal data, when subject to publication according to the decree, must always be treated, as far as distribution, indexing and reuse are concerned, with the guarantee of the respect of regulations of the protection of personal data [4]. Paragraph 3 defines that the publishing of data, information and documents for which no such obligation of publication exists, may occur only after anonymization of any personal data present. Paragraph 4 states that, in those cases in which laws or regulations require the publication of data or documents, public bodies must render unintelligible non-pertinent personal data or unnecessary sensitive or judicial data, with respect to the specific transparency goals of publication. And again, paragraph 5, while consenting access to information relative to the actions of anyone performing a public role, defines as not exposable, unless as foreseen by law, information concerning the nature of any infirmity or personal impediment of the subject or his/her family members that have caused an absence from the workplace, or any evaluation data or information concerning the subject’s job that could reveal sensitive data.

The cases disciplined by section 4 of Decree 33/2013 are not the only limitations on the publication and distribution of personal data by public bodies. Section 19, paragraph 3 of the Personal Data Protection Regulations, for example, declares the general principle according to which the communication of personal data by a public body to a private organisation or to economically autonomous public bodies, and the distribution of said data by public bodies, is allowed only when explicitly foreseen by a specific law (like for example Decree 33/2013) or regulation. Thus a public administration, before it can legally handle information assets, containing personal data, according to the criteria of open data, must first verify the existence of a law that explicitly declares that possibility.

On the other hand, the relationship between open data and the guarantee of privacy need not necessarily be one of conflict, and this is true also with reference to a delicate sector like health. Information treated as open data is often of a statistical nature, or in any case does not consent the identification of a specific citizen. The policies adopted by the Lombardy Region, and applied, for instance, to the datasets relative to the directory of pharmacies, or the analysis of hospital admissions, subdivided according to hospital, are obvious examples of how it is possible to implement processes of transparency and add value to information assets, without compromising privacy.

In this sense, data anonymization processes become particularly relevant, in that they make it possible to preserve the information value of a single datum, while removing from it any element that could allow that datum to be associated to a particular person [5]. It is useful to remember that, as defined by the Personal Data Protection Regulations, anonymous data is that which, originally or after processing, cannot be associated with an identified or identifiable subject, and to which Decree 196/03 does not apply. In the same way, point 26 of EU Regulation 2016/679 specifies that “The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable person, or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.”

The techniques of transformation of data into an anonymous form can be applied to so-called biosignals, whether these be signals really produced biologically, or the products of the interaction between the organism and an external agent (like, for instance, radiography, ultrasound, MRI’s, electrocardiograms, ...), facilitating thus the management of such data according to the criteria of open data. The nature of such data, which carries information that is intrinsically personal, requires par-

ticular attention in the process of evaluating the status of anonymity. While the results of generic blood analysis, simply disassociated with the person's identity can be reasonably considered to be anonymous, the same cannot be said in the case of bioimaging, which, in certain circumstances, could permit the identification of the person involved [6, 7].

4 Open data in Italy. Distribution and usage in the health context

In the general context of the participation of the citizen in the events that concern him/her, the concept of empowerment, obtained through the acquisition of knowledge and the distribution of information, becomes a key element.

The Presidency of the Council of Ministers, together with the Ministry of Economic Development and the Italian Digital Agency, has presented a national plan – “Strategy for Digital Growth 2014-2020”, in which the “open data” platform goes in that direction.

Strictly speaking, not all Public Administration data can be considered open data. It is necessary to verify certain conditions; that is, data is open if it is:

- Available in raw, non-aggregated format;
- Available according to the terms of a licence that permits its re-use, even for commercial purposes (Italian Open Data License v2.0 or the international Creative Commons 4.0);
- Accessible through information and communication technologies in open format (i.e. a documented and technology-neutral public format);
- Available at no cost, or at most at low cost to take into account reproduction and distribution.

The Open Data Barometer of the World Wide Web Foundation provides statistics on the level of adoption of open data policies throughout the world. Italy comes out both positively, considering that it is positioned in the high part of the chart (22nd of 77), and negatively, considering the distance from the leading countries. Figure 1 shows a screenshot of the Open Data Barometer statistics.

The portal of the national Public Administration, <http://www.dati.gov.it/>, managed by AGiD, contains 10,348 datasets coming from 76 administrations. Of these only 256 are health-related.

There are just 30 datasets of the Health Ministry portal <http://www.dati.salute.gov.it/>. Table 1 shows the most popular downloads from this site.

Table 1: Downloads from <http://www.dati.salute.gov.it/>.

Dataset	Downloads
Pharmacies	54,301
Parapharmacies	26,132
Medical devices	47,172
Pharmaceutical Distributors	21,473

The potential of open data in the Health context is however in the forefront. During the event “HACK4DIGITALGOV: PA in your hands” of 2015, of the 3 ideas receiving awards, the first 2 regarded health themes, and used open data of a sanitary nature.

In particular, the first classified was the app “First Aid 2.0”, which had the objective of communicating the overcrowding status of Emergency Rooms according to White, Green and Yellow codes.

In second place was “Health Advisor”, with the scope of letting users express evaluations of the national health service according to their own experience.

5 Open data in the Lombardy region. Health data available

The Lombardy Open Data portal (www.dati.lombardia.it) forms part of the process defined by the Lombardy Digital Agenda (approved by the Regional Government Resolution IX/2585 of 30 November 2011), and in particular is one of its priority initiatives: the valorisation of the public information assets.

The objective is to make available to everyone the Region's information assets, such that large amounts of information of public interest (e.g. health data, demographic data, maps, economic, environment and climate data, ...) can be used by anyone to create new services, perform studies, create applications.

This is the added value: starting from the data that has been made available, it is possible to create innovative services for the Region, favouring not only transparency, but also participation and the collaboration between institutions and the private sector.

Four years later, with its 1500 published datasets relative to 20 different contexts, the Lombardy Open Data portal, based on the platform Socrata, is one of the most active in Italy, and is characterised by the quality of the data, both in terms of completeness and speed of update.

The usage of the Open Data is increasing constantly, as seen in the Table 2.

Table 2: Increase in page visits.

2014	2015
1,566,000 pages visited	3,341,000 pages visited (+113%)
2,814,000 lines accessed	5,336,000 lines accessed (+89%)
49,431 downloads	97,777 downloads (+102%)

The graph in Figure 2 shows visualisations and downloads (data through May 2016).

The Region has made the platform available also to other local authorities of Lombardy. It currently collects data of around 15 bodies, and 3 of these (the city of Bergamo, the city of Monza, and the Province of Monza-Brianza) have activated dedicated micro-sites.

The platform offers several advanced functions: citizens, developers and researchers can consult data, build

maps and graphs and save them on the platform in order to enrich the catalogue. Furthermore, Lombardy Region has promoted various initiatives with the scope of encouraging the use of the data, stimulating digital creativity, in particular for young people, with the participation of schools and universities.

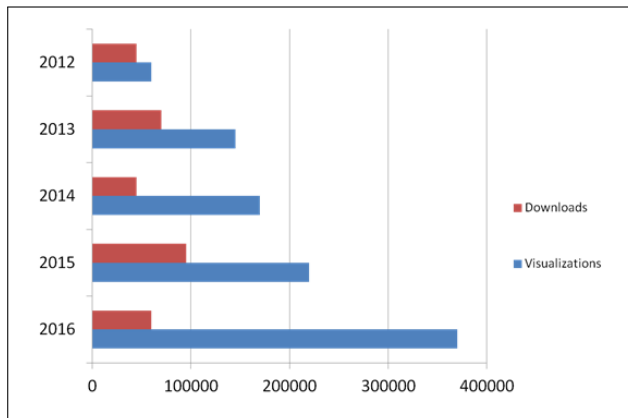


Figure 2: Visualizations (blue) and downloads (red) related to the Open Data portal in the last five years.

In the health context, 25 datasets are available (making a total of 61 taking into account the entire Welfare context), and a similar number of different views of the data, with various filters or applied to maps, which have been created by users of the portal and shared with all visitors.

Since their first publication through to May 2016, health data has been viewed 55K times, with more than 11K downloads; the same figures for the first 5 months of 2016 being 7640 and 2900 respectively.

Table 3 shows, for the most frequent keywords, the number of datasets available.

Table 3: Number of available datasets for the most frequent keywords.

Admissions	13
DRG	8
Repetitions	7
Pharmacies	6
Performance	6
Day-hospital	6
Pharmacy	6
Hospitals	6

As well as simple directory datasets (lists of hospitals, lists of pharmacies, ...), Lombardy offers performance data using the principle indicators of health episode outcome (repeated admission and return to operating room being the most important).

In fact, these queries, along with those relative to hospital productivity data, are amongst the first places of accessed data (cumulative data updated in May 2016) as shown in Figure 3.

Using the published datasets, users have made views based on filters or applied to maps. For example, the

maps of all the pharmacies of the region, shown in Figure 4, or just those of the city of Milan, shown in Figure 5, based on the dataset that contains the complete index of the pharmacies in the region.

Further, in compliance with the open philosophy, it is also possible to share and discuss the results so obtained with a simple, user-friendly interface shown in Figure 6.

6 Usefulness of data opening in the Lombardy health context

The usage of open data in the context of clinical studies is only at the beginning.

Clinical studies require a level of detail in data that cannot be found in data organised for statistical purposes, and so it is necessary to make available data with a sufficient level of granularity. To do so requires a significant effort in the processing of data to make it anonymous (for example, the application of hash functions that, while still guaranteeing the correct association between records associated with the same individual, do not permit his/her identification).

Many and complicated are the processes within the Health context. These generate a vast amount of information, which are archived, not only within hospitals, but also within databases of the national public administration. Many different types of organisation or groups may benefit from the use of health data, even though it is not possible, a priori, to foresee where this added value will be created. In spite of this situation, the Lombardy Region has started in its path towards the opening of its health and administrative information assets to external bodies associated with research organisations and university studies.

The Region, within its institutional functions of the healthcare of its citizens, needs to perform activities of evaluation and monitoring of the effectiveness of the health treatments provided, appropriateness, quality of assistance, user satisfaction, and health risk factors.

The treatment of data performed by Lombardy Region has the objective of evaluating and comparing (between groups of people or hospitals) the appropriateness, effectiveness and efficiency of the healthcare provided, also taking into account specific illnesses or health problems, exposure to risk factors, the reconstruction of diagnostic, therapeutic and care pathways, and the analysis and comparison of healthcare outcomes.

This treatment of data has particularly important goals of public interest, in terms of planning, monitoring, evaluation and appropriateness of healthcare, as foreseen by section 85, paragraph 1b of Decree 196/2003.

One example of open data usage is the portal “Dove e come mi curo” – “Where and how to get well”. This portal (see Figure 7) uses public data to help patients choose the most adequate hospital for some specific medical treatment.

Nome	Popolarità	Tipo
1. Dati regionali ricoveri per DRG Sanità h1, salute, sanità, annuale, drg, prestazioni, ... Analisi delle prestazioni di ricovero in degenza ordinaria: dati regionali per DRG.	1,743 visualizzazioni	
2. Farmacie Sanità farmacia, farmacie, h1 Anagrafica delle Farmacie di Regione Lombardia.	2,451 visualizzazioni	
3. Strutture di ricovero e cura Sanità ospedale, h1 Elenco delle strutture di ricovero e cura di Regione Lombardia.	1,832 visualizzazioni	
4. Performance degli Ospedali Sanità performance, ospedali, h1 Analisi dei dati di performance degli ospedali. L'analisi propone dati a partire dall'anno 2010. Si precisa di	1,560 visualizzazioni	
5. Letti per struttura sanitaria di ricovero Sanità posti letto, h1 Report annuale del numero di posti letto medio per struttura di ricovero con disaggregazione per discipli	2,038 visualizzazioni	
6. Prestazioni Ambulatoriali Sanità prestazioni ambulatoriali, h1, sanità Prestazioni Ambulatoriali	1,389 visualizzazioni	
7. Bacino di utenza delle strutture per ospedale Sanità h1, salute, sanità, annuale, drg, prestazioni, ... Prestazioni Ricoveri - Bacino di utenza delle strutture per ospedale.	849 visualizzazioni	

Note: The values in the Popularity column indicate the total views of the datasets at the time of consultation in May 2016.

Figure 3: Screenshot showing the most popular datasets.

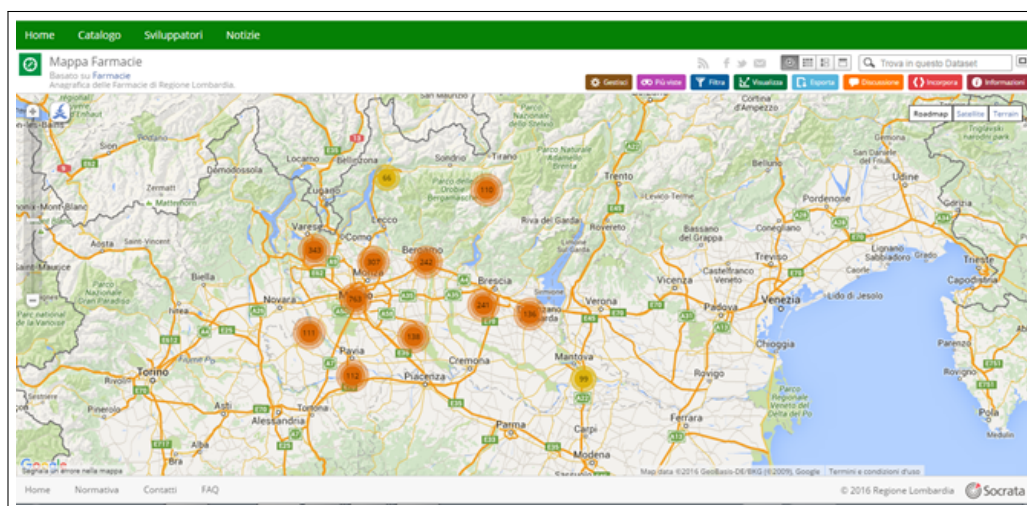


Figure 4: The map of pharmacies in the Region, grouped by Province.

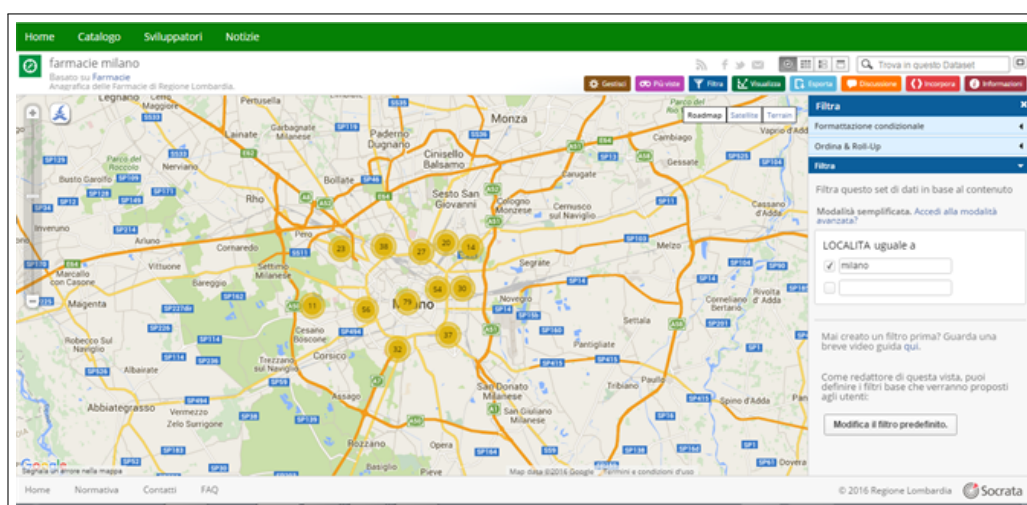


Figure 5: The map of pharmacies in the city of Milan.

Another example is the use of the georeferential coordinates of hospitals, together with Emergency Room real-time usage data, in order to allow users to locate the

nearest hospital together with an estimate of its state of overcrowding. The app, SALUTILE PS, implemented in Lombardy, brings together these two datasets.

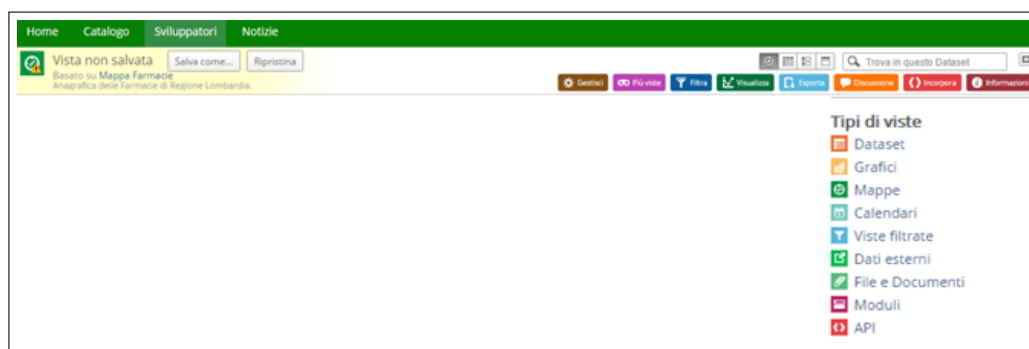


Figure 6: Interface for sharing and discussing the results obtained through the queries.



Figure 7: The portal (www.doveecomemicuro.it).

In order to achieve these goals, Lombardy Region foresees the participation of public and private Universities, public and private Research Hospitals, located within Lombardy and possessing specific requirements of competence in the contexts of epidemiology and research, and with high expertise in those areas¹.

These institutions, in order to collaborate with the Lombardy Region, must submit an application, which will be properly assessed and will allow their inclusion within a specific register.

Once the areas of interest related to a study to be performed have been defined, Lombardy Region chooses the body which, according to its profile in the register, is the most suitable to support it in the activities.

The involvement of these bodies is useful and of fundamental importance considering their profiles of great competence relative to the goals of evaluation of the effectiveness, appropriateness and quality of health-related activity.

The data that is made available concerns in particular:

- contagious and common diseases
- vaccinations
- early diagnosis programmes
- general practitioners
- specialist and rehabilitative outpatient care
- home care
- overseas care
- mental health
- dependencies
- hospital care
- emergency care
- residential and semi-residential care
- midwifery certificates and pregnancy outcomes
- pharmaceutical care and pharmacovigilance

¹DGR N° X / 4893 on 07/03/2016 – “Disciplina delle collaborazioni di enti esterni con regione Lombardia nell’ambito delle attività di programmazione, gestione, controllo e valutazione

dell’assistenza sanitaria, previste dall’art. 85, comma 1, lettera b) del d.lgs.196/2003 e del conseguente accesso ai dati del Data Warehouse Regionale”

- physical and sporting activities
- integrated care
- thermal care
- accident and health risks related to the workplace
- road accidents
- disabilities and handicap
- exemption from healthcare cost payment
- customer satisfaction research
- mortality data
- prosthesis

all without any elements directly identifying the patients involved (name, surname, tax code, healthcare code) such that each individual is assigned a specific unique identifier that does not consent the direct identification of the person during the treatment of data.

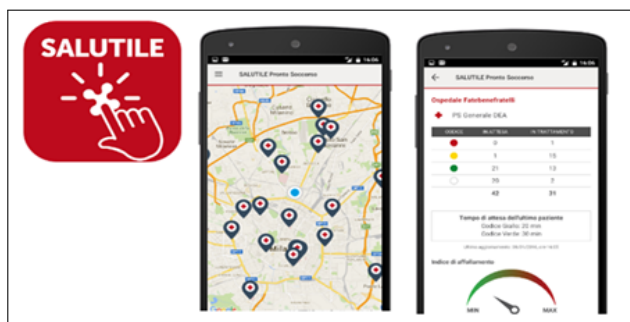


Figure 8: Emergency Room locator.

Furthermore, additional measures of data generalisation are defined and adopted in order to guarantee the non-identifiability of patients, such as using an age-range instead of the date of birth, or a province or geographical area instead of the town of birth.

The information will be used, for processing by the external bodies involved, within a specific IT platform, realised with the necessary measures of security that does not allow data to be exported from it.

The results obtained from the processing of data by the external bodies are owned exclusively by Lombardy Region, which reserves the right to consent the use of the results obtained in aggregated format (for publications, statistical comparisons with other information coming from other regions or countries) according to specific formal agreements, defined within the context of a convention stipulated before the start of the activity.

7 Conclusions

Most health-related datasets available in open data format are indexes of sanitary structures and services. As has already been said, Lombardy is one of the few regions to offer data on the evaluation of its health structures, and has in its objectives the increase in the range of datasets, which however can only happen with the necessary attention to the concerns of privacy already discussed.

The open data philosophy brings us to make ever more datasets available, and the fact of their availability stimulates the creativity of new uses in new systems and applications previously unimaginable. It must however be noted that over and above the increase in the number and quality of available datasets, it is necessary to stimulate the market that is not as yet aware of the resources that it could use; the culture of open data must be encouraged to grow within public bodies and private enterprises.

One of the more interesting areas, with recently a high growth rate, is in the integration of wearable devices (IoT – Internet of Things), equipped with a vast range of sensors able to measure the elementary physical parameters of an individual. Experiments in this field are already underway and regard, for example, remote monitoring of vital signs for patients in homecare.

References

- [1] G. DE MINICO, Gli open data: una politica costituzionalmente necessaria?, in Forum di quaderni costituzionali rassegna, 12.6.2014.
- [2] B. PONTI, La trasparenza amministrativa dopo il d.lgs.14 marzo 2013, Maggioli editore, Santarcangelo di Romagna, 2013, p. 115.
- [3] F. MERLONI, La trasparenza amministrativa, Giuffrè, Milano, 2008, p. 364.
- [4] C. ROMANO, Open data e riutilizzo nel decreto trasparenza: propulsore per la democrazia e lo sviluppo o sfida ulteriore per i diritti fondamentali?, in L. CALIFANO, C. COLAPIETRO (a cura di), Le nuove frontiere della trasparenza nella dimensione costituzionale, Editoriale Scientifica, Napoli, 2014, pp. 263-277.
- [5] ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion 05/2014 on Anonymization Techniques, 0829/14/EN WP216, adopted 10 April 2014.
- [6] G. PALIWAL, A.W. KIWELEKAR, A comparison of mobile patient monitoring systems, in G. HUANG, X. LIU, J. HE, F. KLAWONN, G. YAO, Health Information Science, Second International Conference, Springer, New York, 2013, 205-206.
- [7] W.J. SONG, S.H. SON, M. CHOI; M. KANG, Privacy and security control architecture for ubiquitous RFID healthcare system in wireless sensor networks, in Digest of Technical Papers International Conference on Consumer Electronics, 2006, 239-240.