*Federico Morando

# Open Government Data

## Notes for a State of the Art 2016 and Focus on Italian Government Real Estate Data

DOI: 10.14609/Ti_1_16_1e

**Keywords: Open data, public sector information, copyright licenses, real estate, linked data, geodata.**

**Abstract** Public Administrations (PA) collect data. They do that to function and for accountability purposes. The digital revolution implies that the cost of making these data available for reuse is negligible, while it increases the opportunity cost of limiting their use to the purpose for which they were originally collected. The law encourages such reuse, and there is a growing number of technical standards and good practices making that easier and sustainable. In short, nowadays, the publication of open data is a good practice, but also a duty for PAs.

The paper at hand discusses the aforementioned concepts, with the purpose of providing the elements needed to understand the state of the art of open data (as an approach to government, a field of research and a movement). From the legal point of view, the focus in on the European and Italian jurisdictions. From the technical point of view, the "linked data" formalism is discussed in some details.

The last part of the paper analyses the case of open data concerning government real estate in Italy. Such example is relevant since it is at the border between the open data and transparency domains; moreover, the presence of geographical references (at minimum, the address) provides a key to cross these data with other datasets. An analysis of the current publication practices in this domain is also functional to showing the limitations of a publication duty which is not accompanied by the necessary degree of coordination and by detailed technical guidelines.

* President & CEO, Synapta SRL. Fellow, Nexa Center for Internet & Society - DAUIN, Politecnico di Torino

## INTRODUCTION

Public Administrations (PA) collect data, as any other complex organization. The do that to function and for accountability purposes. This is taking place since a long time, as shown by cuneiform writings. However, something changed in the nature of this huge amount of information in the last decades: we are increasingly talking of digital information. Such characteristic has disruptive consequences, since it reduces the cost of reusing such data, and simultaneously increases the opportunity cost (*i.e.*, the missed opportunities) of keeping such data "locked in a drawer". European and Italian laws encourages the making available of such data, and there is a growing number of technical standards and good practices to make data interoperable and therefore actually reusable.

In this context, the publication of "open" data became one of the pillars of Open Government, and one of the keywords of government in general.

The paper is structured in the following way. The first section defines the concept of "open data", and it is followed by a more detailed part about copyright licenses for opening data. The third section is devoted to technology, and it is followed by a focus on semantic interoperability. The paper is concluded by a case study on government real estate data, followed by some normative remarks.

## OPEN DATA

In the paper at hand, the definition of Open Data is based on the Open Definition project,[1] whose goal is to make precise "the meaning of 'open' in the terms 'open data' and 'open content'". The Definition can be summed up saying that "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)." Or, even more succinctly: "Open data and content can be freely used, modified, and shared by anyone for any purpose."

Openness can then be declined into practical accessibility, legal openness, technical openness.

Having open access to a work implies that the whole work is accessible online without additional charges with respect to a one time reproduction cost. Moreover, legal or technical constraints should not limit its re-use, apart from what is described below.

## OPEN DATA LICENCES

It is acknowledged (*e.g.* by Krötzsch and Speiser (2011), p. 356, with further references) and I already discussed elsewhere (Morando (2013b)) that the distribution of data also requires their licensing. In other words, Open Data should be published under Open Licenses. This is the case because of the current "copyright default", *i.e.* the set of rights that the current regime of copyright protection (in a broad sense, including "droit d'auteur", related rights, and the database *sui generis* right) automatically grants to authors.

This is what I name "copyright default" and it automatically implies that "all rights are reserved" for the maximum duration allowed by the law (typically, the life of the author plus 70 years). Moreover, no formalities are required to enjoy these rights, not even a statement that a certain work is

1 To put The Open Definition in context, it is worth mentionining that it was derived from the Open Source Definition, which in turn was largely based on the Debian Free Software Guidelines and the Debian Social Contract, drafteb by Bruce Perens and the community of Debian Developers.

protected. Similarly, the shorter (*i.e.* 15 years) protection granted to non-creative databases through the *sui generis* database right alone comes into existence automatically. Moreover, the shorter term of protection is automatically renewed by each new significant investment in updating the database and – considering that databases are nowadays frequently distributed as an online service – it may be practically difficult to have access to a 15 years old version of a data set, which in principle could be freely reusable (instead, this may apply to old CDs used to distribute databases in the pre-Internet age).

Therefore, without entering into further details, it is quite fair to say that data re-users should typically assume that everything (apart from a single datum or very trivial lists which are inexpensive to replicate) are somehow protected by intellectual property rights. (The most meaningful exception, worth mentioning for further analysis which cannot be performed here, are sole-source databases, where the creation of the data and the creation of the database cannot be disentangled: for further information see Hugenholtz (2004)).

Having summarized the reasons which make the use of open licenses necessary – or, at least, very appropriate – for the publication of open data, we may further discuss the characteristics that a license should possess in order to qualify as "open". According to the Open Definition, these are the following.

The license should be compatible with other open licenses. Its terms must also satisfy various conditions: it must allow free use of the licensed work, including its redistribution, modification (*i.e.*, the creation of derivatives), separation or compilation (*i.e.*, the conditions must apply to portions of the licensed material, and must permit the distribution together with other materials). Moreover, the license cannot be purpose-bound, *i.e.*, reuse must be possible for any (licit) usage in any field of endeavor. No charges or royalties can be imposed (on the first reuse, while commercial reuses of the material itself or its derivatives must always be possible).

The only acceptable conditions are attribution clauses prescribing the mention of authors, contributors, rights holders, sponsors, etc., and the requirement that modified versions are clearly described as such in order to guarantee the integrity of the data along the chain of re-uses. Moreover, share-alike clauses are allowed, *i.e.*, the license may require the use of the same and/or of a compatible license for the distributions of derivative materials. I remand to the full text of the Open Definition for further details.[2]

When a Public Administration actually tries to choose a license to open its data, it has to weight various elements, including the opportunity of adopting the most standard tools, following a recommendation included in the European Public Sector Information Directive (see recital 26 of Directive 2013/37/EU: "Any licences for the re-use of public sector information should in any event place as few restrictions on re-use as possible, for example limiting them to an indication of source. Open licences available online, which grant wider re-use rights without technological, financial or geographical limitations and relying on open data formats, should play an important role in this respect. Therefore, Member States should encourage the use of open licences that should eventually become common practice across the Union").

2 See http://opendefinition.org/od/2.1/en/ for the latest version of the Definition at the time of writing these lines.

A first option could be to use Free and Open Source Software licenses. However, this is not a very widespread approach, since FLOSS licenses are very specialized tools and using them for things which are not pieces of software is typically suboptimal.

Another option could consist in using one of the licenses from the Creative Commons (CC) suite, which are general purpose licensing tools. These licenses offer to right-holders a menu of elements/modules from which they can pick their favorite combination and including: "Attribution" (BY); "Non-Commercial" (NC); "No Derivative Works" (ND), meaning that only verbatim copies could be produced; and "Share Alike" (SA), meaning that the author requires the creators of derivative works to adopt the same license used by him/her (the so-called "viral" or "copyleft" effect).[3] The (meaningful) combinations of the previous elements generate six different licenses, two of which can be defined as "open licenses" according to the Open Definition: CC BY and CC BY-SA. On top of these standard licenses, CC also offers a right waiver or dedication to the public domain (with a fall-back clause to a very permissive license in jurisdictions where some rights cannot be waived): Creative Commons Zero (CC0).

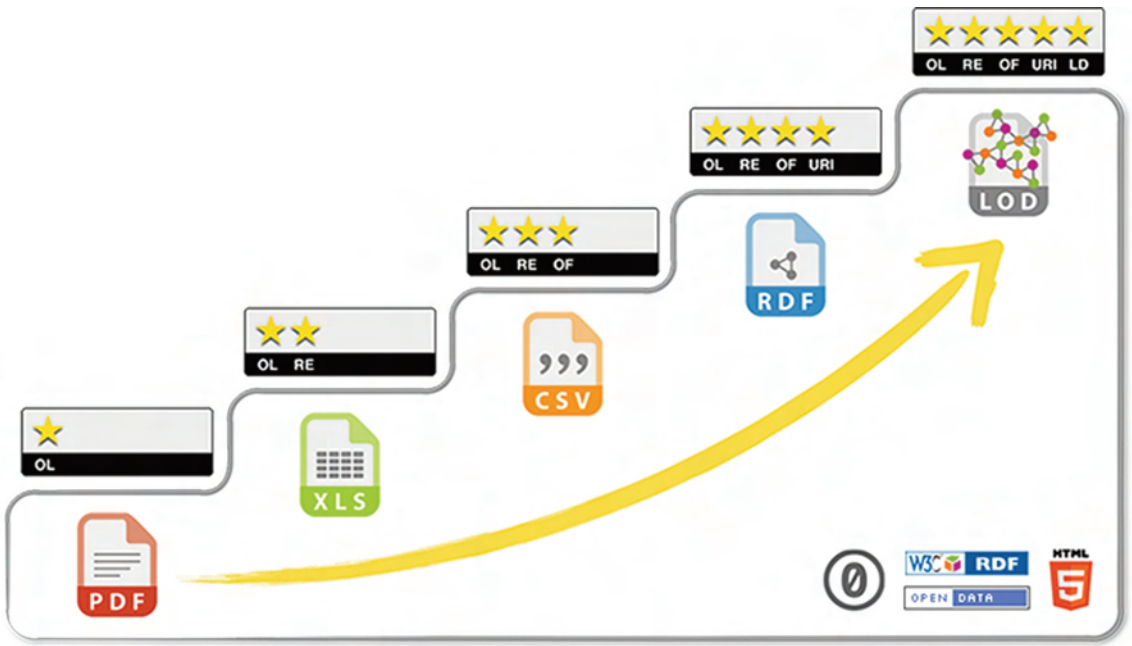
Despite the availability of standard public licenses, such as the ones from Creative Commons (and others from Open Data Commons), several national governments decided to draft their own licenses for the release of (open) Public Sector Information. One of the first countries to do so (also because of the choices of CC concerning the *sui generis* database right that I discussed elsewhere, see Morando (2013b)) was the United Kingdom, with its "Click Use" license and its current non-transactional evolution, the Open Government License (OGL). The OGL is essentially equivalent to a CC Attribution license, but it includes some specific provisions concerning "Crown copyright" and other clauses addressing standard public sector worries, such as forbidding the use of the released information in such a way that suggests any official status. The OGL approach adopted in the UK was almost immediately and is still followed all over the world (and in Europe in particular). For instance, France adopted its own License Ouverte, while Italy produced the Italian Open Data License (IODL), which was released in various versions, starting from a non-commercial beta version to arrive (going through a 1.0 copyleft version) to the current 2.0 version, which is a more permissive attribution license.

3 You may find more practical information about the CC licenses at http://www.creativecommons.org. For a more theoretical and impartial commentary about CC licenses, see (the first part of) Elkin-Koren (2006).

## TECHNOLOGY: FORMATS AND FORMALISMS

The Open Definition has two main technical requirements: machine readability and open format. The first amounts to requiring that the material is provided in a form which may be easily processed by a computer, so that the individual elements of the work – *e.g.* a single cell in a table – can be easily accessed and modified.[4] The latter requires that the serialization format of the data is fully documented and its implementation is not subject to patents or other royalties. Moreover – with a requirement which is more demanding than most definitions of open format – at least one open source implementation of the specification must already be available and must allow for the full processing of the data.[5]

Tim Berners-Lee suggested a 5-star deployment scheme for open data, indicating linked open data (see below) as the reference publication approach.[6]




*Source: http://5stardata.info/en/*

4 See Art. 2, par. c-bis, of D.Lgs., January 24, 2006, n. 36, (Italian implementation of Directive Direttiva 2003/98/EC), that defines "machine readable format: a file format structured so that software applications may easily locate, recognize and extract specific data, including individual statements of facts and their internal structure".

5 See Art. 68, par. 3, letter a), of D.Lgs. March 7, 2005, n. 82, according to which open data format means "a data format made public, exhaustively documented and neutral with respect to the technological means necessary to make use of the data", which may be interpreted as equivalent to the Open Definition requirement, or less demanding.

6 This approach was originally described in Berners-Lee T. (2006).

According to this model, data holders (and, in particular, public sector information holders) may publish their data under the following approaches:

1) basic open data: publishing data on the Web, in whatever format, *e.g.*, PDF files (including raster ones), under an open license;
2) structured open data in proprietary format: as in 1), but published as machine readable data, *e.g.*, Excel spreadsheet;
3) structured open data in open format: as in 2), but published in an open format, *e.g.*, CSV, XML or JSON;
4) RDF data: as in 3), but using URIs (see below) to denote entities, so that third-parties can easily make reference to the published data;
5) Linked Open Data: as in 4), but also linking the published data to other data RDF data to provide context.

As it appears from the list, the aforementioned approaches are in a growing scale of technical sophistication and re-usability, in the sense that it becomes easier and easier to build pieces of software re-using the data, but also their production is a growing scale of complexity and cost.

Therefore, according to Time Berners-Lee, but also following the recommendations of the European Commission or of the Agency for Digital Italy,[7] the ideal way of publishing open data is the Linked Open Data paradigm.

## TOWARD ACTUAL SEMANTIC INTEROPERABILITY

Linked data are data structured following the Resource Description Framework[8] (RDF) formalism: knowledge is expressed in the form of subject-predicate-object "triples", and each entity is represented by a unique identifier, which is also a URI[9] that answers to HTTP calls. Such technology was designed and is strongly pushed by W3C as a standard[10] to favor interoperability amongst heterogeneous data sources on the Web. Here, interoperability is not guaranteed by a predefined semantic; instead, it is favored by the use of a conceptually simple formalism, by the existence of identifiers which are globally unique by construction (the URIs), and by making it easy to reuse existing vocabularies and taxonomies, possibly by annotating data *ex post* and even if they are published by completely independent and unrelated organizations.

7 See AgID (2012).

8 For an introduction, see https://www.w3.org/RDF/.

9 Oversimplifying, the concept of URI is similar to the more familiar HTTP URL, substantially synonymous with "web address", since a URL is used to retrieve content on the Web. A URI, however, is a more abstract concept and is first of all a unique resource (*i.e.*, entity, concept) identifier, which is then "resolved", also pointing to a URL, which provides useful information about the same resource (a Web page to describe it to a human being, or a set of RDF triples describing it to a computer program in a machine processable way; indeed, different "agents" can receive different URLs calling the same URI, thanks to a process known as "content negotiation"). To learn more, you can start from the Wikipedia page https://en.wikipedia.org/wiki/Uniform_Resource_Identifier.

10 A list of the technical specifications is available at https://www.w3.org/standards/techs/rdf. For an informal introduction to the key concepts, see Berners-Lee T. (2006).

It is also worth mentioning that such an approach, that was born to enable the Web of Data,[11] is actually very useful within complex organizations, where there are several data sources, which are not subject to a strong governance in the first place. Therefore, public administrations may represent an especially appropriate use case, not only for their publishing of open data, but also for interoperability and data exchange purposes within a single big or medium administration or amongst administrations.

Indeed, when data are structured according to the linked data formalism, linking different datasets becomes trivial, as it happens to extracting new knowledge relying on information published by third parties. A case worth mentioning is the one where the integration of existing databases goes through the enrichment of internal data using external sources, *e.g.*, geo-referencing some information using an external database, to then project the same data on an existing internal cartography.

Linked Data also represent a powerful data curation tool, thanks to the opportunity of discovering new matching keys from existing databases, but also thanks to the fact that making the links between different databases explicit allows you to analyze and manage inconsistencies and conflicts.

To achieve an increasing degree of semantic interoperability, it is necessary that, whenever possible, the RFD formalism is coupled with the use of standard vocabularies, such as Dublin Core or Friend of a Friend, or Schema.org, or knowledge representation systems as SKOS, and taxonomies such as Eurovoc, as well as standard metadating models at the dataset level, as DCAT. All this in order to allow interoperability between datasets produced by different and diverse authorities.

The RDF formalism and the SPARQL query language are general purpose tools intended to represent and access any kind of data. That said, and considering the relevance of geospatial data and their role as a reference layer connecting other sources of information, it may be worth mentioning that a specific sub-sets of standards exist to represent geographic information as Linked Data.

In particular, GeoSPARQL[12] is a standard from the Open Geospatial Consortium (OGC) for representing and querying geospatial linked data. Data may be saved in an RDF graph as, *e.g.* Geography Markup Language (GML) or well-known text (WKT) literals, and various kinds of reasoning are available to query them, *e.g.*, using Simple Features, RCC8, or DE-9IM/Egenhofer topological relationships /representations.

Since, as I mentioned above, the geospatial level is one of the most useful layers to cross amongst them public datasets from different sources, the paper at hand is concluded by an example of (partial and imperfect, but already significant) opening of government data about real estate.

**11** Even private forms contribute – of course in their own interest – to gradually implement a Web that contains more and more structured data; for example, Google and Facebook support this approach through Schema.org and the Open Graph Protocol, in order to facilitate the ingestion of data from the Web into their platforms. Linked data are also at the center of platforms such as IBM's Watson or similar projects in the cognitive computing domain.

**12** Specification available at http://www.opengeospatial.org/standards/geosparql.

## THE CASE OF GOVERNMENT REAL ESTATE DATA

The case of open government data relating to the real estate of PAs is noteworthy, since the opening of such data is required by the legislation on administrative transparency, but the same data are also susceptible to various forms of re-use for different purposes (*e.g.* business intelligence). Moreover, the presence of geospatial references in the data (as a minimum, the street address; often, structured cadastral references) provides a useful key for the integration with other the datasets.

The sole paragraph of art. 30 of Legislative Decree. N. 33/2013 ("Transparency Decree"), titled "Publication requirements relating to real estate and assets management," provides[13] that "public administrations publish identifying information of the properties owned as well as lease or rent fees paid or received." This provision should be implemented in the "Transparent Administration" section of the website of each PA.[14]

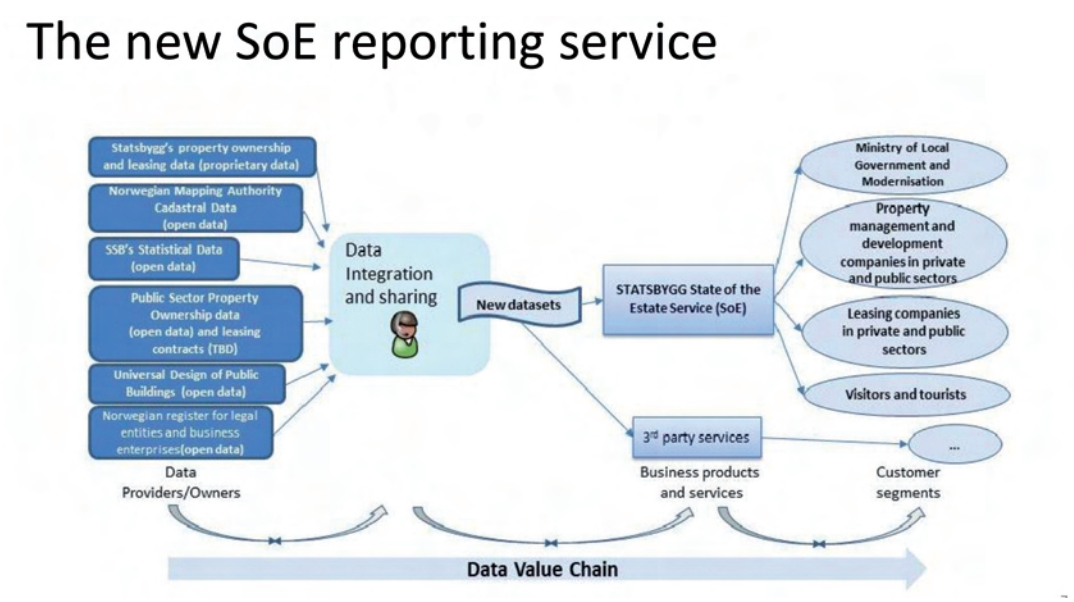
Despite the apparently self-explanatory nature of this provision, the analysis of its implementation from the perspective of a re-user for research or commercial purposes offers excellent opportunities to illustrate the potential of publishing open data, but also the limitations of a publication performed without the necessary degree of coordination or detailed technical guidelines.

From the point of view of the potential value of the publication, the benefits in terms of transparency of the activities of the public administration are self-evident, but one should add to these the informative and managerial advantages that may be achieved correlating this information with other already available open or proprietary data. In this regard, it seems appropriate and effective refer to a concrete example relating to a business case of the proDataMarket European project.[15] This is the case of Statsbygg, a government owned company which is responsible for managing and developing publicly owned real estate in Norway. This case showed that crossing the datasets from the sources shown in the figure below led to a significant increase in data quality and other advantages in terms of downstream innovation.

13 The draft decree reorganizing the rules governing access to Public Administration data, known as "Madia Decree" (in implementation of art. 7 of the Act no. 124 of 2015 for the "review and simplification of the provisions on prevention of corruption, openness and transparency") includes, at the time of writing these lines, an amendment to Art. 30, which extends the obligation to publish to properties in any way "detained", and not just to those "owned" by the administrations. That said, no other material changes to this publication duty are expected. The real news, however, is in article 9-*bis* and it concerns the publishing mechanism of the information about the Estate of PAs, which will to occur through a national database called "PA Estate", held by the MEF-DT: see the Concluding Remarks of this article for further details.

14 By "public administrations" the Transparency Decree means (according to Art. 11, c. 1) all government departments, including institutes and schools of all levels and educational institutions, companies and autonomous state administrations, the regions, provinces, municipalities, mountain communities and their consortia and associations, academic institutions, autonomous public housing institutes, chambers of commerce, industry, handicrafts and agriculture and their associations, all public non-economic national, regional and local authorities, government departments, enterprises and institutions of the national health service, some other agencies (ARAN and the agencies under Decree. n. 300/1999), as well as some more specific cases (*e.g.*, CONI). Without deepening this analysis of the scope of the Decree, in all cases of which the author is aware, also independent administrative authorities behaved as if the Decree applied to them. Finally, it is worth mentioning that the aforementioned reform proposal (*supra*, note 13) also includes a subjective modification to the scope of the law, so that it is explicitly extended (to the extent compatible) to the following categories of entities: (a) public economic entities, including port authorities and professional orders; (b) companies in public control, apart from listed ones; (c) associations, foundations and private sector bodies mainly funded by or whose administrative organs are appointed by public administrations. Some obligations would also extend to companies owned by the government or to associations, foundations and private law entities exercising administrative functions, etc.

15 See Shi *et al.* (2015) and the website of the proDataMarket project (Grant number: 644497), in particular: http://blog.prodatamarket.eu/2015/08/prodatamarket-business-cases-at-ruleml2015-industry-track/.




*Source: Slide #7 of the presentation "Industry@RuleML2015: Norwegian State of Estate – A Reporting Service for the State-owned Properties in Norway"*[16]

Coming to what can be improved, with respect to the case of government real estate data publication in Italy, there is no shortage of examples. Among these, with regard to the analytical description of the "identifying information" of the property, the analysis of the practices of major Italian municipalities shows that the rule is not always applied, or it is sometimes interpreted discordantly.

In the following sections, I will describe the results of an analysis carried out in the period between July 2014 and January 2015,[17] which considered all provincial capitals and other municipalities with a population greater than or equal to 100,000, for a total of 119 municipalities.[18] In the analysis, the duty to publish concerning "Real estate ownership and management" was divided into three parts: (a) Real estate, (b) Active rents, (c) Passive rents.

In summary, about ¾ of the administrations published their data, but less than 20% of them did that in a technically open mode (machine readable and in open formats).

16 Available at http://www.slideshare.net/ruleml2012/industryruleml2015-norwegian-state-of-estate-a-reporting-service-for-the-stateowned-properties-in-norway (under a CC Attribution License).

17 The survey has been performed and/or verified by Roberta Bruno, in the period betweeb October 2014 and January 2015, with the technical support of Alessio Melandri.

18 The available data at the moment of performing the analysis were the ones from the monthly ISTAT Demographic survey (Bilancio demografico mensile), release of June 30, 2014, as reported by https://it.wikipedia.org/wiki/Comuni_d%27Italia_per_popolazione.

Although the differences are minimal, the data on real estate in the strict sense have been published marginally better than those related to rents, and active rents slightly better than passive ones:

| | Usable data | Absent data | Unusable data | Data with 3+ stars |
|---|---|---|---|---|
| **Real estate** | 75% | 18% | 8% | 19% |
| **Active rents** | 73% | 20% | 7% | 18% |
| **Passive rents** | 71% | 26% | 3% | 17% |

The first column indicates the percentage of usable (*i.e.*, readable and understandable) information from the point of view of a human being. The second and third columns divide the non-usable information between absent (there are no or empty files) and formally present but irrelevant or illegible data. The "data with 3+ Stars" column indicates the percentage of datasets that can be classified as having 3 or more stars according to the "five star Open Data" model (or, more specifically, available in CSV, ODS, XML format, or as an HTML table). All percentages are calculated on the 119 cases analyzed.

What impedes the most the reuse of data, however, it is not their format, but the absence of a standard record layout (*i.e.*, of standard headings for columns): for example, some kind of "identifying information" of the property is normally provided, but only in 40% of cases there is a full address, and the presence of cadastral data is limited to 55% of the cases, and it falls to about 10% with respect to the description of rents (and, if this could be obviated in some cases crossing the dataset about active rents with the one about real estate ownership, data concerning passive rents are simply unavailable).

| AVAILABILITY OF INFORMATION CONCERNING: | | | | | | |
|---|---|---|---|---|---|---|
| | Full address | Cadastral data | Floor area | Rent amount | Rent time period | Signing date of contract |
| **Real estate** | 43% | 55% | 27% | n.a. | n.a | n.a |
| **Active rents** | 34% | 13% | 21% | 76% | 41% | 19% |
| **Passive rents** | 37% | 8% | 24% | 71% | 37% | 24% |

Please refer to the technical report by Canova *et al.* (2015) for more information about the aforementioned research. For the purposes of this paper, I also conducted a qualitative update of the analysis during the month of May 2016, focusing on the most populous Italian cities. This analysis confirmed the 2015 study results: for example, in some cases, as in Milan, cadastral data are absent, as any other information to uniquely identify the portion of property owned or its dimensions; in other cases, as in Turin, the dataset is richer, but the only available format is PDF, making it difficult to perform any further elaboration.

### Standardization already in place

The poor quality and lack of any uniformity in open government data about Italian government real estate is a paradox, since the Budget Law 2010 already provided the following. Government departments and other public national or local public administrations, their consortia and associations, as well as the agencies, etc., "that use or hold, for whatever reason, properties owned by the state or other public properties, transmit to the Ministry of Economy and Finance - Treasury Department the list of these real estate assets for the purposes of preparing the general balance sheet of the State assessed at market values. By July 31 of each following year, the administrations [...] communicate any changes. [...] The method of communication and transmission of the information described above is established by a decision of the Director of the State Property Agency."

I am not aware of how many administrations are implementing these legal obligations, but the aforementioned decision of the Director of the State Property Agency is public and well-detailed and it is a component of the project called "PA Estate". These guidelines explain that both land and building data should be reported, and the types of information to be collected (always at the level of single cadastral identifiers), which concern the typological classification of the property, the location, the size, the kind of current use, the presence of any cultural or environmental restrictions, the period of construction, as well as the balance sheet value (or estimated market value, and the year of its estimate, if available). In addition, the procedure provides for the possibility of uploading data massively, with import and export in open format (CSV).[19]

In other words, there is already a unique national database – even if it is not openly accessible – that contains much of the information that public administrations are due to publish, although in a less organized way, on their institutional websites, with the addition of some information about rental income and expenses. It is also interesting to note that the portal OpenDemanio, an open government initiative about real estate information from the State Property Agency, is currently making available the same kind of information mentioned above, but limited to the real estate assets directly managed by the Agency itself. Therefore, in practice, a pilot project for the opening of the "PA Estate" database has already been implemented. Quantitatively, these assets represent a small percentage of the information supplied by Italian administrations: for example, with respect to the territory of the municipality of Turin, there are less than 400 assets inside the files which may be downloaded at http://dati.agenziademanio.it/#/opendata (*i.e.*, only the properties managed by the Agency), compared to about 6,000 properties in respect of which the City of Turin shall forward information to the Treasury Department. The initiative, however, is of particular importance, both because it offers a significant amount of certified information, and because it represents an authoritative "precedent" of open data publication in this domain. Indeed, since its homepage, the portal OpenDemanio makes explicit reference to the Open Definition, uses standard open licenses (Creative Commons Attribution 3.0 Italy) and publishes data in an open and machine readable format (CSV). If this example was followed by all PAs for their own real estate data published pursuant to the Transparency Decree, the quality of open government real estate data in Italy would experiment a dramatic increase.

19 For further practical details I remand to the Web page "FAQ - Rilevazione del Patrimonio della PA" (http://www.dt.tesoro.it/en/faq/faq_patrimonio.html).

## FINAL REMARKS

The paper at hand had the objective of collecting the main concepts related to open data in a concise text addressed to the community of researchers and practitioners that play – in the academia and in and around Public Administrations – a key role in the modernization of the analysis and management of territorial information.

Regarding the case of government real estate data, one of the purposes of this short paper is to call attention to the existence of a subset of such data that is already available as open data, promoting their reuse and, indirectly, increasing incentives to publish them and generating feedback about their publication. In fact, in my experience, public administrations tend to invest in data quality, especially when there are re-users that pretend such quality, which may be a reasonable approach, considering the severe budgetary constraints that they face, and the sheer number of administrative requirements burdening them.

It also seems appropriate to conclude by mentioning the fact that many of the problems related to the quality of government real estate data described in this article should gradually be overcome by virtue of the reorganization of the law about access to the public administration data, known as "Madia Decree". Indeed, the currently available draft provides, in Art. 9-*bis* (Publication of databases), that public administrations must fulfill some of their publication duties through the communication of the data, information or documents to another administration managing one of several national databases. In the case of real estate, it would be the "PA Estate database", held by the Department of Treasury of the Ministry of Economy and Finance.[20] The publication in a single database would ensure format uniformity and hopefully also semantic standardization. Subsequently, any single administration could fulfill its publication duties by publishing a simple hyperlink on its website.

In the long run, the existence of a national database establishing a standard for the publication of government real estate data is a good news for those interested in the reuse of data on the assets of PAs for research or business reasons. In the short term, however, what I described in the paper at hand about the quality of currently available open data on government real estate is a warning for those who will have the task of curating the data flows feeding the "PA Estate" database. In the absence of strict guidelines and/or IT-driven processes, as well as proper governance, there could be a risk of transferring within the unified database the current data Babel which characterizes open government real estate data. Fortunately, the detailed guidelines already developed by the Treasury Department in relation to the PA Estate project, as well as the good practice represented by the OpenDemanio project bode well for the rapid publication of all this information as high quality open data.

While waiting for that to happen, right now and on a unilateral basis, each PA could proceed to publish in its own Transparent Administration website section the CSV file that it periodically uploads on the State Property Agency portal, enriched by information about rental income and costs.

20 The database was set up by the aforementioned Art. 2, par. 222, of Law no. 191 of 2009 and Art. 17, co. 3-4, Law no. 90 of 2014, ratified with amendments by Law no. 114 of 2014.

## References

AgID (2012), *Linee guida per l'interoperabilità semantica attraverso i Linked Open Data*, versione 2.0, pubblicata il 12 November 2012, available at http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf.

Belisario, E., Cogo, G., Epifani, S., & Forghieri, C. (2011). *Come si fa Open Data? Istruzioni per luso per Entie Amministrazioni Pubbliche*.

Berners-Lee T. (2006), *Linked Data*, W3C Design Issues series, availabe at https://www.w3.org/DesignIssues/LinkedData.html.

Biallo G. (2013) *Dati geografici aperti: istruzioni per l'uso -Italia OpenGeoData*, Open Knowledge Foundation Italia, Open data handbook, Retrived at http://opendatahandbook.org/, May 2013

Canova L., Iemma R., Morando F. (2015), *#WikiTrasparenza – Chi apre davvero i dati sulle risorse pubbliche*. Nexa Working Paper n. 2015-1, disponibile su http://nexa.polito.it/working-paper/2015-1.

Elkin-Koren N. (2006), *Exploring Creative Commons: A Skeptical View of a Worthy Pursuit*, in Hugenholtz P. B. and Guibault L. (eds.), *The Future of the Public Domain*, Kluwer Law International, 2006.

Hugenholtz P. B. (2004), *Abuse of Database Right – Sole-source information banks under the EU Database Directive*, presentato alla conferenza "Antitrust, Patent and Copyright", École des Mines/UC Berkeley, Paris, January 15-16, 2004. Pubblicato in Lévêque F. and Shelanski H. (eds.), *Antitrust, patents and copyright: EU and US perspectives*, Cheltenham: Edward Elgar 2005, p. 203-219. Disponibile su http://www.ivir.nl/publicaties/download/953.

Krötzsch M. e Speiser S. (2011), "ShareAlike Your Data: Self-referential Usage Policies for the Semantic Web." ISWC , Ed. Aroyo S. L., Welty C., Alani H., Taylor J., Bernstein A., Kagal L., Noy N. e Blomqvist E. Springer Berlin / Heidelberg ISWC 2011: 354-369.

Morando F. (2013a), *Gli obblighi di pubblicazione concernenti l'uso delle risorse* (artt. 26-27, 29-30). In: La trasparenza amministrativa dopo il d.lgs. 14 marzo 2013, n. 33 / Ponti B. Maggioli spa, pp. 239-257. ISBN 8838781028.

Morando F. (2013b), *Legal interoperability: making Open Government Data compatible with businesses and communities*. In: JLIS.IT, vol. 4 n. 1, pp. 441-452. ISSN 2038-1026.

Morando F., *et al.* (2013), "Guida alla pubblicazione dei dati delle pubbliche amministrazioni", Edizioni ETS srl, Pisa, pp. 9-83. ISBN 9788846735782.

Rizzo G., Morando F., De Martin J.C. (2011), *Open Data: la piattaforma di dati aperti per il Linked Data*. In: *Informatica e diritto* n. 20, pp. 493-511. ISSN 0390-0975.

Shi L., Pettersen B. E., Østhassel I, Nikolov N, Khorramhonarnama A., Berre A. J., Roman D., (2015) *Norwegian State of Estate: A Reporting Service for the State-Owned Properties in Norway*, in *Rule Technologies: Foundations, Tools, and Applications: 9th International Symposium,* RuleML 2015, Berlin, Germany, August 2-5, 2015, Proceedings, pp 456-464, Springer.

## Webography

The Open Definition project, setting out principles that define "openness" in relation to data and content. http://opendefinition.org/ (last visit: 01/05/2016)

Open Government License v. 3. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ (last visit: 1/05/2016)

Italian Open Data License v2.0. http://www.dati.gov.it/iodl/2.0/ (last visit: 1/05/2016)

Tim Berners-Lee 5-star open data deployment by examples. http://5stardata.info/en/ (last visit: 01/05/2016)

The OGC GeoSPARQL standard for representing and querying geospatial linked data. http://www.opengeospatial.org/standards/geosparql (last visit: 01/05/2016)