# Opening Public Deliberations: Transparency, Privacy, Anonymisation

Eleonora Bassi[1], David Leoni[1], Stefano Leucci[1], Juan Pane[1], and Lorenzino Vaccari[2]

[1] University of Trento, Italy
{bassi,pane}@disi.unitn.it,
{david.leoni,stefano.leucci}@unitn.it
[2] Autonomous Province of Trento, Italy
lorenzino.vaccari@provincia.tn.it

**Abstract.** The open data movement is demanding publication of data withheld by public institutions. Wide access to government data improves transparency and also fosters economic growth. Still, careless publication of personal data can easily lead to privacy violations. Due to these concerns, the Italian law states that even public deliberations must be anonymised for long term publication. In the context of the Trentino Open Data Project (Italy), we first analyse privacy legislation and anonymisation techniques. Then, we propose a semantic open source stack based on entity and word sense disambiguation techniques for publishing anonymised deliberations edited with Norme in Rete software.

**Keywords:** open data, public sector information, digital administration, public deliberations, privacy, anonymisation, semantics, legal texts.

## 1 Introduction

Governments around the world are starting to recognize the value of the data kept in public institutions. The open data movement pushes for such data disclosure, as it allows broader public scrutiny and also boosts economies often choked by excessive bureaucracy. In this paper we analyze the problem of disclosing public deliberations as open data while preserving individual privacy. Which are the European and Italian legal frameworks in transparency and open data? Is it possible to use existing XML standards for legal documents? How can we assist the identification of personal data inside deliberations with semantic technologies? In the following, we try to answer these questions. We move from an overview of the European and Italian legal framework on open data (Section 2), in order to introduce our topic and its prominence after the new Italian rules on transparency (Section 3). In Sections 4 and 5 we analyse some technical and legal issues. In Section 6 we expose some anonymisation techniques and discuss their utility in our context. Finally, in Section 7 we propose a semantic open source stack to handle publication of anonymised deliberations in the Trentino Open Data Project.

P. Casanovas et al. (Eds.): AICOL IV/V 2013, LNAI 8929, pp. 41–53, 2014.

## 2 The European and Italian Legal Framework on Transparency and Open Data

In the past ten years, open data initiatives became every day more important for the digital information market: the main ambition is to enforce the innovation of public sector in order to enhance the transparency of public administrations activities and the participation of citizens. This goal is reached by publishing data previously withheld from public scrutiny, thus greatly improving governments accountability.

### 2.1 The European Legal Framework

Open data policies became a legislative program in Europe since the last 90: the European legislator adopted D-2003/98/EC [10], introducing rules that allow and encourage the reuse of public sector information (PSI), that is the information gathered and owned by public sector bodies (PSBs), in order to remove barriers such as discriminatory practices, monopoly markets and a lack of transparency [2],[15],[17],[19]. PSI is a very wide notion that often includes personal data. According to the PSI Directive, also personal data could be reused, but in a way that shall not affect the level of protection of the individuals according to D-95/46/EC (Privacy Directive). The difficulty to solve the problem of compatibility of these two directives (PSI Directive and Privacy Directive) striking a fair balance between all the fundamental rights and interests involved (transparency, freedom of information, right to privacy, access to public documents and reuse) made the case for the reuse of personal data a crucial point for the European legislator and Data Protection Authorities [7],[9],[16],[18],[24]. This matter affects the issues we are analyzing in this paper: how to assure data protection for the case of reuse of provisions and deliberations from PSBs (as pointed out by the Commission Decision of 12 December 2011 on the reuse of Commission documents(2011/833/EU)). The European legislator - both in the PSI Directive of 2003 and in its revision of 2013 (D-2013/37/EU) - preferred not to decide how to balance those different disciplines, and the consequent practical and technical measures for assuring a legitimate reuse of personal data - with the only exception of personal data from intelligent transportation system databases for which D-2010/40/UE (ITS Directive) prescribes that full anonymisation should be adopted.

### 2.2 The Italian Legal Framework

In 2006 the Italian legislator adopted the D. Lgs. n. 36/2006 that transposed the PSI directive: some local administrations implemented the European and national rules on PSI reuse, but updated them following the main core of European best practices on open data and the hints offered by the revision process of the PSI Directive. Finally, in the last year, Italy adopted a new framework of rules on transparency, accountability and the disclosure of data from public

administrations (D. Lgs. 33/2013). Although this new Decree is not directed primarily to the implementation of European rules on public sector information and open data, but to improve the functioning of public administrations, accountability and transparency, it requires the publication as open data of a large number of datasets and official documents, including deliberations. Thus, in the new Italian legal framework, the problem we analyze in this paper assumes an important role for enabling transparency and accountability through open data measures preserving privacy rights.

## 3 Transparency, Public Availability and Disclosure of Public Deliberations

Deliberations are concrete and particular acts of public administrations necessary to the exercise of their activities. Publicity is a prerequisite for the validity of the act that allows the ability to know. According to the Italian law on local government (Art. 124, D. Lgs. 267/2000), all the deliberations of the municipalities and the provinces are published by publication on the city register, at the headquarters institution, for fifteen consecutive days, except for special provisions of law.

### 3.1 From Paper to Bits

The Italian Digital Administration Code (D. Lgs. 82/2005) provides that the electronic version produces legal effects of publicity in the cases and in the manners expressly provided by the law. The L. 69/2009 on simplification and competitiveness in public administration establishes the rule that from January 1st, 2012, the publication of acts and administrative measures which have the effect of legal publicity are read as acquitted with the publication of information on their web sites by government and public bodies.

### 3.2 Problems of Interpretation

The recent introduction of the Decree 33/2013 creates problems of interpretation. Art. 7 provides that data subject to mandatory disclosure are published in open format pursuant to the Italian Digital Administration Code. The problem arises with regard to the time criterion of publication: Art. 8, D. Lgs. 33/2013, provides that data subject to mandatory disclosure under the current regulations are published for a period of five years and in any case until the published acts produce their effects. We have here a conflict of interpretation between the D. Lgs. 267/2000 and the recent D. Lgs. 33/2013: fifteen days (Art. 124, D. Lgs. 267/2000) or five years (Art. 8, D. Lgs. 33/2013)? The question could be solved by the principle of succession of laws in time which prefers the idea of the D. Lgs. 33/2013 (five years). But the principle of specialty could be used to solve the problem: according to the special rule, that is an exception to the general one, the D. Lgs. 267/2000 would keep its effects. As a possible solution, the act must

be published for fifteen days (inclusive of all personal data contained within), and thereafter for the next five years it will be published in anonymous form, in accordance with art. 4, D. Lgs. 33/2013. Deliberations will remain available in their entirety to persons who advance an instance of access according to the requirements of L. 241/1990.

## 4 Opening Public Deliberations: Some Technical Remarks

The main problem in opening public deliberations concerns the structure of texts, which is not uniform across different Italian administrations. Several projects aim to solve this issue: the most complete and useful specifications for structuring legal texts are Norme in Rete and AkomaNtoso. The first is supported by the drafting environment xmLeges and the second by the application AT4AM, which is currently in use at the European Parliament. Since the Italian policy made by the Agenzia per l'Italia Digitale [11] recommends usage of NormeInRete mark-up schema, we decided to adopt xmLeges editor which best supports it.

### 4.1 NIR Project

NIR, developed by CNIPA (Italian National Center for Information Technology in the Public Administration) in conjunction with the Italian Ministry of Justice, ITTIG-CNR (Institute of Legal Theory and Techniques of the Italian National Research Council), University of Bologna and Italian Parliament, proposed the adoption of XML as a standard for representing legal documents using also additional meta information and a uniform cross referencing system (URN), providing documents with characteristics of interoperability and effectiveness of use. Another goal was to foster the building of legal texts access facilities for both citizens and legal experts. The standard for legal document description was created to increase degree of depth in text hierarchy description for different kind of legal documents by the definition of an XML-DTDs (NIR-DTDs), an example of which can be seen in Figure 1. The standard establishes constraints in the hierarchy of the formal elements of a legislative text (collections of articles), and a specification of the meta data which can be applied to a legislative document or to parts of it [6]. The advantages of XML format for legal documents are briefly summarized as follows: standardized definition of the structure of the document; automated assistance for the creation of legal texts; regulatory impact assessment on sorting; improved navigation within the legal texts; extensive research in the legislative databases; increased uniformity [5].

### 4.2 XmLegesEditor

A tool was built in order to obtain an holistic approach to the drafting process: xmLegesEditor is a specific integrated legislative drafting environment developed at ITTIG/CNR for supporting the adoption of NIR XML standards. The effort

```
<intestazione><titoloDoc id="titolo">DISCIPLINARY ACTION 5/13</titoloDoc></intestazione>
<formulainiziale> Attendances <h:br/>
    Mayor        Bertrandi Paolo <h:br/>
    Councillors  Bianchi Maurizio  <h:br/>  Tomasi Alberto  <h:br/>
    <preambolo>
        <h:p> WHEREAS, disciplinary charges were served on City employee
            Mr Giovanni Pedrotti,  born in Rovereto on 18 March 1985, national identity number
            PDRGVN85P55L378O  and living in Mattei Street 73, Trento</h:p>
        <h:p> WHEREAS, the Councillor Bianchi Maurizio reminded the serious allegations
            which were proffered against Mr Giovanni Pedrotti in note 46/13  </h:p>
    </preambolo>  <h:br/> NOW, THEREFORE, BE IT RESOLVED <h:br/>
</formulainiziale>
<articolato>
    <articolo id="art1">  <num>Art. 1.</num> <corpo>
        Giovanni Pedrotti's employment with the City of Trento shall be immediately terminated.
    </corpo></comma></articolo>
</articolato>
```

**Fig. 1.** Deliberation excerpt with NIR XML markup. Text with personal data is underlined.

made with the development of xmLegesEditor has been to establish a trade-off between a user-friendly approach to text authoring hiding the underlying XML structure, and the maximum flexibility and extensibility in the exploitation of the high potentiality of content expression offered by XML documents [1]. Typically, WYSISYG word-processors are mainly oriented to texts' style markup rather than structural and semantic markup. XmLegesEditor proposes an original approach to this problem: the basic idea is that the user should be constrained by the editor to perform only valid operations on the document in such a way that, starting from a valid document, only valid documents can be produced [1]. A fundamental feature of xmLegesEditor is that it is a free resource, distributed with an open source license (GNU-GPL v3): the idea is to offer a shared highly customizable and extensible platform to develop specific functions and easily integrate existing or new designed tools as external modules.

## 5 Opening Public Deliberations: Some Privacy Remarks

Public deliberations contain in many cases personal data that requires to be protected due the disclosure. This is the typical case of balancing between transparency and privacy rights (see European Data Protection Supervisor (EDPS) [8,9] and [14,15],[18],[21]). The Italian DPA has stated several times about this problem for cases of publication of personal data in deliberations and administrative acts (Dec. 26/10/1998 [doc. web n. 30951], Dec. 2/9/1999 [doc. web n. 1092322], Dec. 23/2/2012 [doc. web n. 1876679], Dec. 7/10/2009 [doc. web n.

1669620]), prescribing the adoption of all technical measures for protecting privacy and to respect the principles of necessity, proportionality and minimization. Despite these DPA decisions were focused on privacy concerns related only to publication and not on reuse, they were complying with the recommendations of Art.29 Working Party and of EDPS on the reuse of PSI.

### 5.1 Call for Anonymisation

In the Opinion 7/2003 on the re-use of public sector information [24], Art. 29 Working Party insisted on the role that anonymisation can play in this sector and made the same recommendation in his Opinion 3/2013 on purpose limitation [26] and in the Opinion 6/2013 on open data and public sector information (PSI) reuse [27], stressing - in a stronger way - the necessity of anonymising personal data for the disclosure as open data, having in mind the connection between the scenario of reuse of personal open data and the potentiality of big data and data analytics (see Annex 2: Big data and open data). It is important to note that according to the WP29 anonymisation is not the only measure that a PSB must adopt in order to publish open data protecting privacy rights: the PSB should necessary conduct a robust and detailed privacy impact assessment identifying the risks and the measures adopted, following a case by case approach. However, although anonymisation is not considered a sufficient tool, in many cases it is strongly recommended or imposed as necessary. This position was followed in February 2013 by the Italian DPA in his Opinion on the draft of the Transparency Decree [doc web. n. 2243168]: anonymisation is required as necessary measure to assure the privacy of citizens for the publication (as open data) of public information for which the publication is not mandatory. The Transparency Decree adopted the solution proposed. We experienced firsthand the need for anonymisation by discovering with a simple Google search an ordinance where a mayor imposed a mandatory medical treatment to a citizen suffering from psychiatric illness. Name, birthdate and residence address of the citizen were all explicitly written resulting in a clear privacy breach. At the time of our search the ordinance wasn't present on the communality website anymore, yet we managed to found a copy inside Google cache.

### 5.2 Anonymisation Level

Some doubts arise on what kind of anonymisation the European DPAs and Member States legislators are referred to [7]. In the Opinions mentioned before, Art. 29 Working Party refers to a strict concept of anonymisation, that is required to avoid the constraints imposed by privacy legislation, while, in other cases, the argumentation is open to different levels of anonymisation and different technical possibilities, in relation to the probability of re-identification, to its costs and to the context of processing ([9],[14,15],[25]).

### 5.3 What to Anonymise

Deliberations may contain personal information under the form of names, addresses, birthdates, sex. In Fig. 1 we may see an example of some word that must (and must not, like Council members) be anonymised. Additional personal information can be found in documents referenced by the deliberation, such as *note nr 46/13* in the example. Since these documents might contain identifying information about physical persons named in the deliberation, if they are publicly available in non-anonymised form, references to them must be cancelled out. Also, referenced documents such as addendums can be in any format, including images. Trying to aid anonymisation of images by automatic means is much more difficult than dealing with plain text.

## 6 Anonymisation Techniques for Open Data

During last years several clamorous cases of privacy breaches occurred after the publication of supposedly anonymised datasets [4],[13],[23]. In 1997, Sweeney showed it is possible in the US to find the identity of a person by just knowing his age, sex, ZIP code with 5 digits and crossing this data with voting records, which are public in the US [20].

### 6.1 Reference Guide

Since UK is spearheading open data movement in Europe, its citizens are increasingly worried about their personal data being published on the internet. To address their concerns, UK government released a valuable Code of practice for anonymisation [22]. It targets a broad audience, explaining in simple terms risks and methods related to anonymisation. Anonymising deliberations falls into the so-called case of qualitative data anonymisation, where identifying information such as names and addresses is either cancelled out before publication, or generalized by applying a method called banding. An example might be substituting the address *Mattei Street, 73, Trento 38122* with a generic *Trento, 38XXX*. Banding preserves more information and it is valuable for researchers in social sciences when studying anonymised transcripts of interviews with people. Another option could be to use a technique called pseudo-anonymisation to associate a unique key to each anonymised person and substitute names in the text with that key. This would allow to recognize that the same person is mentioned in different deliberations without disclosing the actual identity of that person. To validate the effectiveness of the anonymisation ICOs Code of practice recommends performing the so-called *motivated intruder* test before publishing anonymised data. The test prescribes to play the role of an individual who wants to identify people in the anonymised dataset *if* motivated for some reason (i.e. sell data, blackmail people, stalking, etc). The intruder is supposed to try crosslinking anonymised data to existing sources by only using legal means, like searching the internet, enquiring people, looking at public records and so on.

The Working Party Art.29 in the Opinion 06/2013 [27] cites the motivated intruder test but seems skeptical about its effectiveness: among other things, it stresses how not all possible motivations can always be foreseen. The Opinion recommends so-called re-identification tests, where attempts to re-identification are done regardless of the possible supposed gains. Recently the Working Party Art.29 also published a detailed guide on anonymisation techniques [28] cast in the EU legal framework, where it offers a much welcomed quantitative approach to the problem of anonymisation. Reviewed techniques range from the simplest $k$-anonymisation by generalization to the most advanced randomization method of differential privacy. The report concludes there is still no silver bullet, and a case by case analysis must be performed prior the publication of any dataset.

### 6.2 Solution for Deliberations

Pseudo-anonymisation can be discarded right away because deliberations are published both in original and later in anonymised form, allowing to easily associate a person name to its key. While Working Party report on anonymisation [28] is clearly of importance when publishing statistical datasets, unfortunately is less relevant in our case. Statistical data is usually provided under the form of a table where it is relatively easy to understand how persons could be grouped to protect their anonymity. On the other hand, personal information in deliberations is scattered all over the text, and people mentioned in them are mostly unrelated. References to other documents containing additional information about persons in the deliberation also offer lots of clues for cross-linking attacks. In order to make a quantitative assessment of the amount of disclosed information, it would be necessary to mark and collect all such data (in the case of disciplinary action of Figure 1 it could be the office where the employee was working, his position, etc). Over time, this would give a clear historical picture of what has been released and allow more precise choice of anonymisation to perform in new documents. Although interesting, conducting such an analysis at present seems too onerous for a public administration. For these reasons, our current choice is to adopt the approach of cancelling out identifiers (such as names, social security numbers) and main quasi-identifiers (such as gender, birth-dates and postal codes).

## 7 Use Case: Open Data Initiative of Trentino

The Autonomous Province of Trento (PAT) has promoted territorial development based on competitiveness and innovation through specific innovation programs and laws. In particular, the Provincial Development Plan (PSP  Piano di Sviluppo Provinciale) aims to adopt the information society as the fundamental resource for its territorial development. This vision was confirmed by the Provincial law 16/2012 which foresees the adoption of the Open Source software and of the Open Data paradigm. Then, the PAT approved the provincial guidelines about the Open Data formats, metadata and licences (Del. 2858/2012).

Deliberations (from the Province and from other municipalities too, including the City of Trento) are public information that the local government open data initiative is planning to open following new transparency rules. As we have seen in the previous sections, opening these deliberations matters for privacy protection. Moreover, as deliberations are only a part of the data to be published in the open data catalog *dati.trentino.it* from the Province, the issue of ensuring privacy of information requires a more comprehensive solution than only anonymising data in the deliberations. However, the same techniques that we apply to tackle the problems in the deliberations, can be used to deal with the anonymisation on other types of text.

### 7.1 The Semantic Stack

In order to support anonymisation, the open data initiative of Trentino includes a semantic stack that encompasses tools to parse and understand content of the datasets. Considering the semantification of text, the semantic stack will include Natural Language Processing (NLP) [12] techniques to parse sentences, and also Word Sense Disambiguation (WSD) [3] and Named Entity Recognition (NER) and Disambiguation (NED) [12] techniques, among others. By relying on NLP tools, we can parse all the sentence to its components, such as subjects, predicates, verbs, tagging each word with its parts of speech (verb, noun, adjective, adverb, ...), then using WSD, we can disambiguate the meaning of each word in the text, which would allow us to recognize synonyms in the text, such as car and automobile, and differentiate homonyms, such as bank (of the river) and bank (the financial institution). Once we know the meaning of each word, this will simplify the task of identifying name references in the text, task that is performed using NER, and later to disambiguate the exact entity that is being referred, using NED. In the case of the deliberations we want to be able to automatically recognize person name references in the text, but it is also very important to know who is being referred by the name, because not all the names need to be anonymised, for example, public names, such as the President of the Council, or the signatories of the Deliberations do not need to be anonymised.

### 7.2 A Possible Solution: xmLeges Extension

We propose an extension to the xmLeges software and a possible *ex ante* procedure for the anonymisation of the deliberations problem, rather than an *ex post* solution. We suggest a workflow that includes the editing step of the deliberations, which would allow the authors to identify, with some automatic support, the parts that need privacy protection. The workflow is outlined in Figure 2 as follows:

1. the deliberations are edited inside xmLeges software, which will automatically suggest the common XML structure;
2. during the editing process, all the text is parsed and disambiguated using NLP, WSD, NER and NED tools, allowing the editor to automatically find the name references in the text;

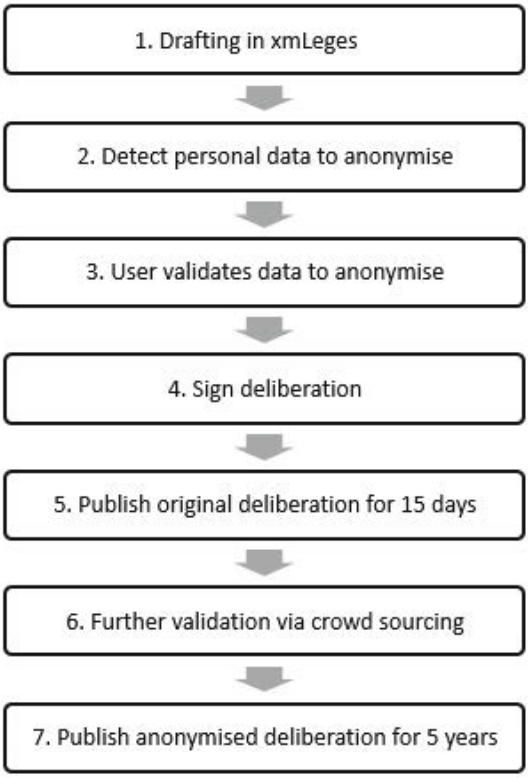


**Fig. 2.** *Ex ante* approach for anonymisation

3. the xmLeges allows the user to manually check all the text that was automatically marked to be potentially anonymised, to accept or reject these elements. The user should also be able to manually mark the names, addresses, and other text that s/he thinks needs to be anonymised;
4. the user sends the finished deliberation for signing and receives it back when this is done;
5. the deliberation is published for 15 days as is;
6. during these 15 days, the user can still further mark the parts of the deliberation that need to be anonymised. At this step, one can design a crowdsourcing-like approach for marking (tagging) the text that needs anonymisation;
7. the anonymised deliberations are published for 5 years, allowing the readers to ask for the original deliberations using the proper channels to obtain them.

In order to accomplish the above steps, we plan to extend the open source xmLeges with the semantic technologies available in the Trentino Open Data Project. The semantic xmLeges (S-xmLeges) will also be made available as an open source project.

In some cases it will be difficult to adopt the *ex ante* approach requiring the usage of the S-xmLeges tool for the creation and edition of the deliberations, as this would require training and switching editing tools that people in public administrations are already familiar with. When this is the case, we can adapt the *ex ante* approach to convert it into an *ex post* approach by allowing people to create the deliberations as they want, publish them for the required 15 days as is, and then, we adapt the steps described above as outlined in Figure 3:

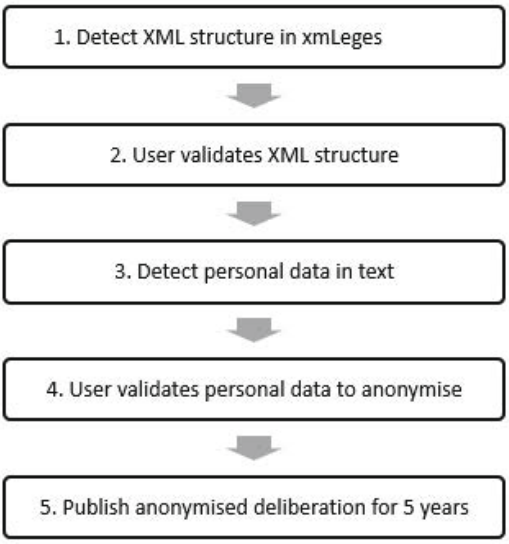


**Fig. 3.** *Ex post* approach for anonymisation

1. the deliberations are loaded into xmLeges software, which will parse them into the common XML structure;
2. user validates the XML structure, making sure the suggested tagging is appropriate
3. all the text is parsed and disambiguated using NLP, WSD, NER and NED tools;
4. the xmLeges allows the user (or via crowdsourcing-like approaches) to validate all the text that was automatically marked;
5. the anonymised deliberations are published for 5 years.

### 7.3 Possible Issues

The main technical issues that we foresee with this approach is the ability to fully automatically recognize rather technical terms that are part of the lexicon

in the legal domain. This can be dealt with a vocabulary that can be built based on existing legal dictionaries, and creating crowdsourcing tasks whenever new terms are not found in the dictionary, asking the crowd of experts in the legal domain to define these terms. Also, given that the state of the art WSD, NER and NED tools are not perfect, a human would need to double check some of the annotations created by these tools, when the confidence in the disambiguation or recognitions is below a threshold.

## 8 Conclusions

In this paper we proposed how to manage legal text according to Italian transparency laws and open data principles, balanced with privacy rights. We suggested an *ex ante* solution that enhances the Norme in Rete software with a semantic open source stack for publishing anonymised deliberations, combined with an *ex post* solution. The proposed S-XmLeges extension will be tested in the Trentino Open Data project.

**Acknowledgements.** This work has been partly supported by the Trentino Open Data project (see `http://dati.trentino.it/`).

## References

1. Agnoloni, T., Francesconi, E., Spinosa, P.: xmLegesEditor, an OpenSource visual XML editor for supporting Legal National Standards. In: Proc. of V Legislative XML Workshop, pp. 239–252. European Press Academic Publishing (2007)
2. Aichholzer, G., Burkert, H.: Public Sector Information in the Digital Age: Between Markets, Public Management and Citizens' Rights. Edward Elgar Publishing, Incorporated (2004), `http://books.google.it/books?id=a0AbDHMb5rAC`
3. Andrews, P., Pane, J.: Sense induction in folksonomies: a review. Artificial Intelligence Review, 1–28 (2013), `http://dx.doi.org/10.1007/s10462-012-9382-7`
4. Barbaro, M., Zeller, T.: A Face Is Exposed for AOL Searcher No. 4417749. The New York Times (August 2006), `http://www.nytimes.com/2006/08/09/technology/09aol.html`
5. Biagioli, C.: Modelli funzionali delle leggi. Verso testi legislativi autoesplicativi. Series in legal information and communication technologies, EPAP (2009), `http://books.google.it/books?id=A6RqQgAACAAJ`
6. Biagioli, C., Francesconi, E., Spinosa, P.L., Taddei, M.: A legal drafting environment based on formal and semantic XML standards. In: ICAIL, pp. 244–245 (2005)
7. Dos Santos, C., Bassi, E., De Terwangne, C., Fernandez Salmeron, M., Tepina, P.: Policy Recommendation on Privacy and Personal Data Protection as Regards Re-Use of Public Sector Information (PSI). Masaryk University Journal of Law and Technology (MUJLT) 6(3) (2012)
8. EDPS: Public access to documents containing personal data after the Bavarian Lager ruling of 24 March 2011 (2011)
9. EDPS: Opinion on the open-data package (April 18, 2012) (2012)

10. European Parliament: Directive 2003/98/EC, PSI Directive (2003)
11. Giovannini, M.P., Palmirani, M., Francesconi, E.: Linee guida per la marcatura dei documenti normativi secondo gli standard NormeInRete, Agenzia per l'Italia digitale. Series in Legal Information and Communication Technologies, vol. 9. Agenzia per l'Italia digitale, vol. 9 (November 2012)
12. Margonar, S., Giunchiglia, F., Pane, J.: A Large Scale Name Matching and Search Framework. Technical report, DISI - University of Trento (March 2013), http://eprints.biblio.unitn.it/4161/1/DISI-13-026.pdf
13. Narayanan, A., Shmatikov, V.: Robust De-anonymization of Large Sparse Datasets. In: Proceedings of the 2008 IEEE Symposium on Security and Privacy, SP 2008, pp. 111–125. IEEE Computer Society, Washington, DC (2008), http://dx.doi.org/10.1109/SP.2008.33
14. Nissenbaum, H.: Privacy as Contextual Integrity. Washington Law Review 79(1) (2004)
15. O'Hara, K.: Transparent government, not transparent citizens: a report on privacy and transparency for the Cabinet Office. Technical report, Cabinet Office (September 2011), http://eprints.soton.ac.uk/272769/
16. Pagallo, U., Bassi, E.: Open Data Protection: Challenges, Perspectives and Tools for the Re-use of PSI. In: Hildebrandt, M., O'Hara, K., Waidner, M. (eds.) Digital Enlightenment Yearbook 2013, pp. 179–189. IOS Press (2013)
17. Ponti, B.: La trasparenza amministrativa dopo il d. lgs, 33 (marzo 14, 2013). Analisi normativa, aspetti organizzativi ed indicazioni operative. Series in legal information and communication technologies, Maggioli (2014)
18. Raab, C.: Privacy Issues as Limits to Access. In: Aichholzer, G., Burkert, H. (eds.) Public sector information in the digital age: between markets, public management and citizens' rights, pp. 23–46. Edward Elgar (2004)
19. Ricolfi, M., Sappa, C. (eds.): Extracting Value From Public Sector Information: Legal Framework and Regional Policies. Quaderni del Dipartimento di Giurisprudenza dell'Universitá di Torino, ESI (2013)
20. Sweeney, L.: Computational Disclosure Control - A Primer on Data Privacy Protection. Tech. rep. MIT (2001)
21. Turilli, M., Floridi, L.: The ethics of information transparency. Ethics and Inf. Technol. 11(2), 105–112 (2009), http://dx.doi.org/10.1007/s10676-009-9187-9
22. UK Information Commissioner's Office: Anonymisation: managing data protection risk code of practice (2012)
23. Wikimedia: What are readers looking for? Wikipedia search data now available. Wikimedia Blog (September 2012), http://blog.wikimedia.org/2012/09/19/what-are-readers-looking-for-wikipedia-search-data-now-available/
24. WP29: Opinion 7/2003 on the re-use of public sector information and the protection of personal data - Striking the balance (wp83) (July 2003)
25. WP29: Opinion 4/2007 on the concept of personal data (wp136) (April 2007)
26. WP29: Opinion 03/2013 on purpose limitation (wp203) (April 2013a)
27. WP29: Opinion 06/2013 on open data and public sector information ('PSI') reuse (wp207) (June 2013b)
28. WP29: Opinion 05/2014 on Anonymisation Techniques (wp216) (April 2014)