Does eContent talk to the heart?

Alfredo M. Ronchi

EC MEDICI Framework - Politecnico di Milano

alfredo.ronchi@polimi.it

Summary

- eContent & Service languages

- Open Data

- Intellectual Property Rights management

Internet Content & Services

- *ICT is stimulating changes in the way most people earn their incomes; altering the balance between our roles as consumer and producers; changing the way we educate succeeding generation and train ourselves; changing the fruition of world's cultural heritage; transforming the delivery of health care; altering the way we govern ourselves; changing the way we form communities; altering the way we obtain and communicate information; contributing to bridge some cultural or physical gaps; and modifying pattern of activity among the elderly.*

- In recent times the digital domain, once strictly populated by professional users and computer scientists, open up to former digitally divided. Technology is evolving toward a mature "calm" phase, "users" are overlapping more and more "citizens" and they consider technology and eServices as an everyday commodity, to buy a ticket, to meet a medical doctor, to access weather forecast even to initiate "social" relation.

- How such a new wave influenced cultural diversity and languages? Till what extent eContent and Services are available in local languages? Is the Internet a melting pot creating a new lingua franca the "Engternet"? After different waves jeopardising cultural diversity such as the different aspects of globalisation including global markets and infrastructures the Internet and related services are a potential silver bullet to kill diversities. This aspect takes us to carefully consider the importance to preserve "diversity", especially in the digital age. Which is the real value of diversity?
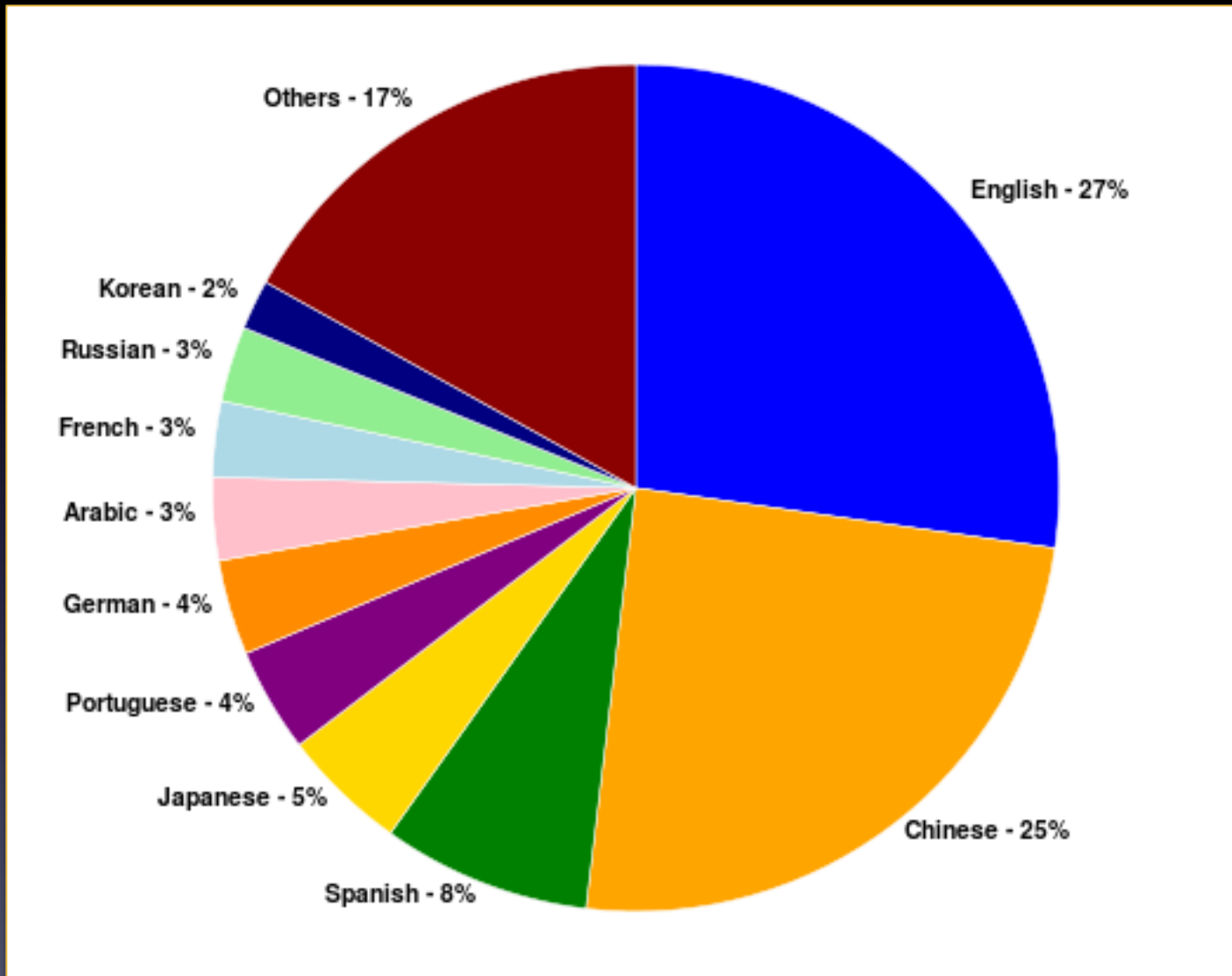
Communication

- *We all know that the world population today is bigger than the number of people that lived on the planet earth since the human race appeared, but incredibly today is easier to disseminate ideas and content through the planet reaching individuals. This is one of the effects of the global inter-communication in the digital era. Moreover global software tools are unleashing everyday creativity no regards about citizenship, language, gender or census.*

- On one hand the digital age is enabling better opportunities to exploit local cultures and knowledge due to minorities, on the other hand such a "global village" jeopardizes minorities and local cultures playing the role of standardization agent.

- A kind of English language, the one generated by spelling and grammar checkers, and translators is still placed in pole position but very close we find Chinese language quickly improving its ranking.
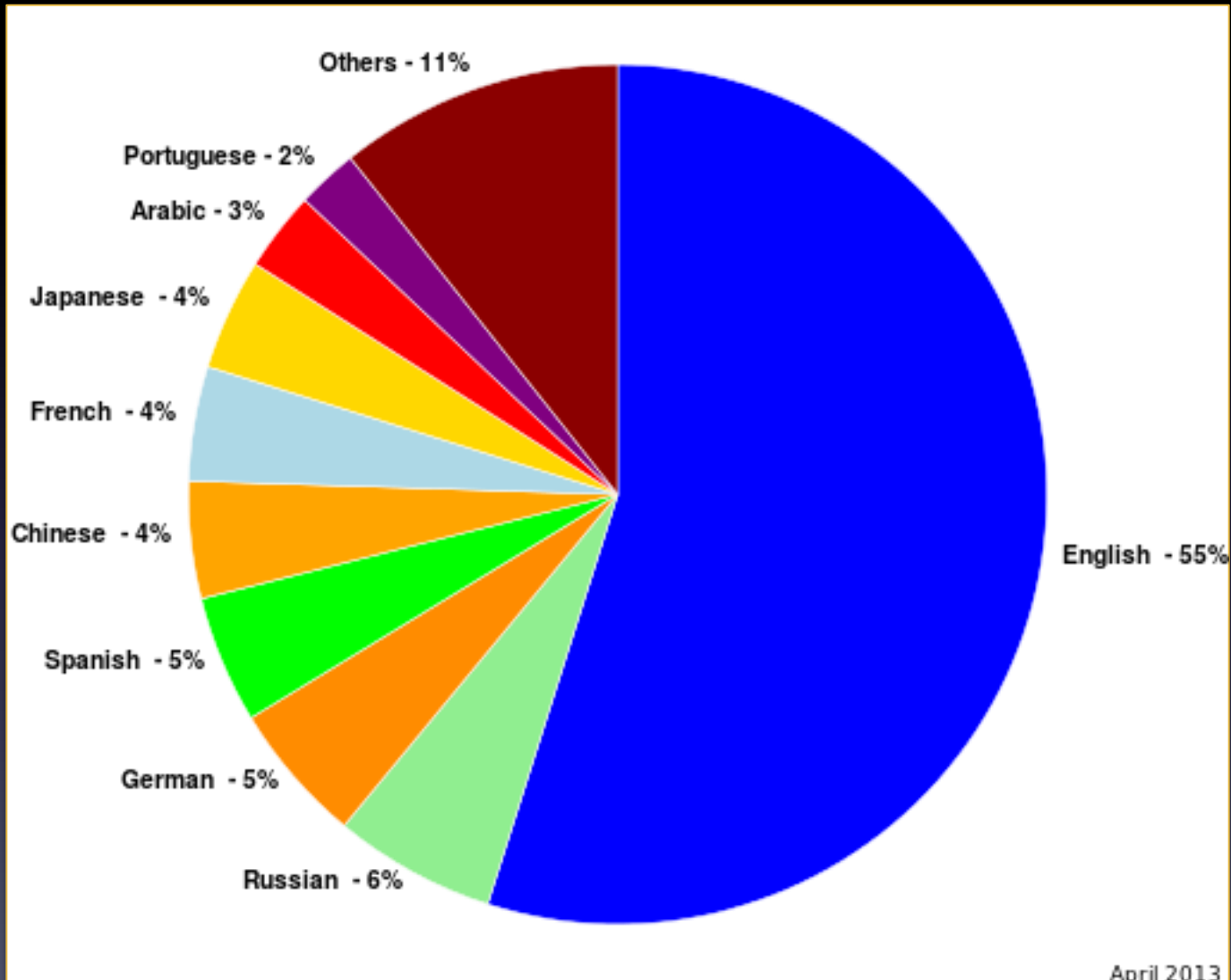
Citizens and Languages

- Let's get a little bit into figures; accordingly with the latest International Telecommunication Union (ITU) survey 2014 on a world population round 7,1 billion we find a 61% of people not using at all the Internet and 39% of Internet active users where the gap between developing and developed countries is 31% to 77%. If we consider the subdivision by macro-regions of the world we find in 2013, again thanks to ITU surveys, Africa 16%, Americas 61%, Arab States 38%, Asia Pacific 32%, Commonwealth 52% and Europe 75%.

- More interesting are figures about Internet subscription subdivided by region subdivided by fixed or mobile connections. We find in 2013 and average value of 9.8% for fixed broadband line subscribers. In developed word this figure is 27.2% while in developing world is 6.1%. If we switch to wireless broadband the situation is quite different. The average value is 29.5% where 74.8% is due to developed world and 19.8% to developing world.

- The presence of different languages on the web may be summarized as W3Techs.com found in 2014, they ranked the first 36 languages but we can limit our insight to the first ten.
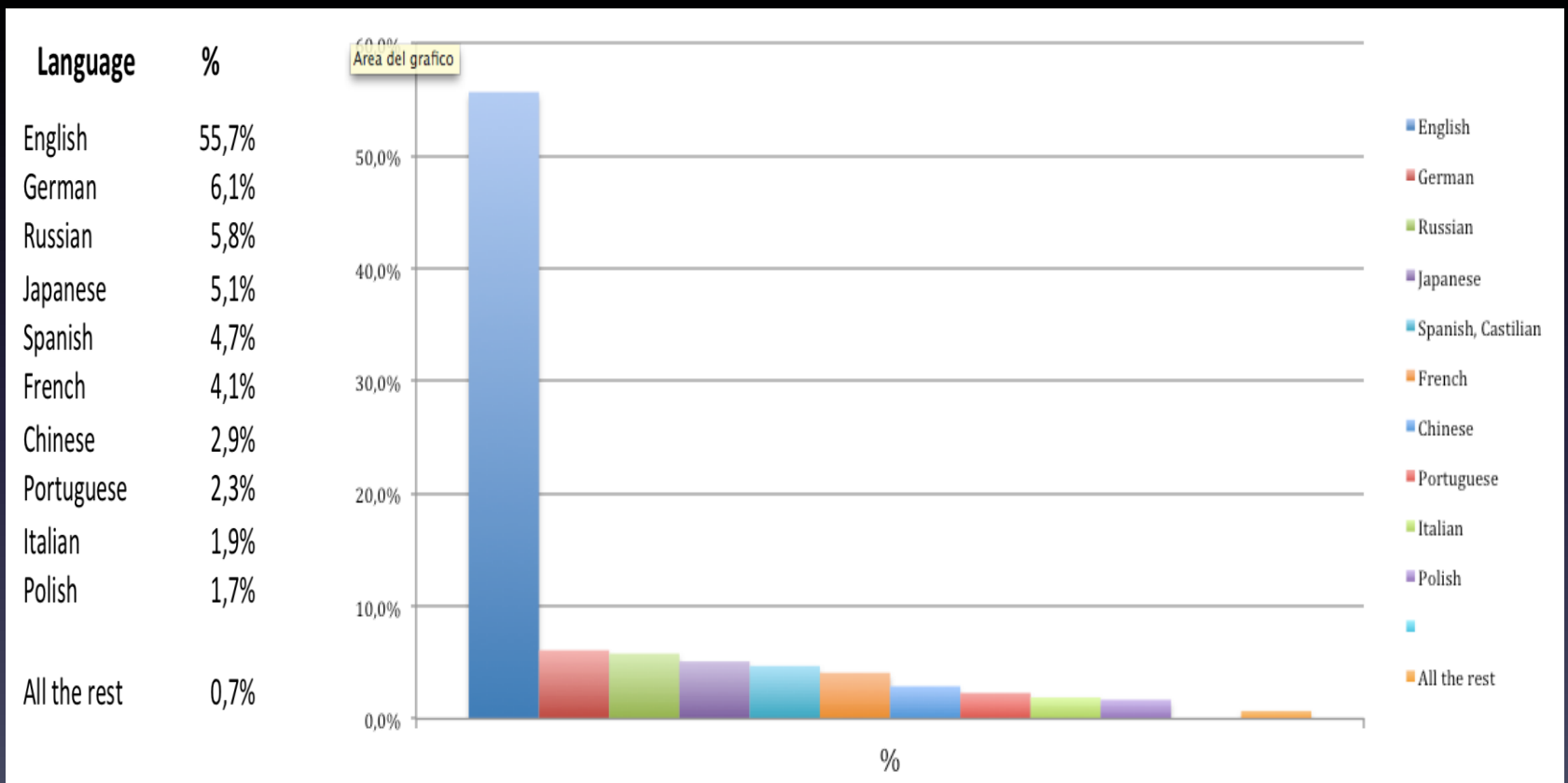
Internet Users By Language

Web Languages

Top 10 On Line Languages

Language	%
English	55,7%
German	6,1%
Russian	5,8%
Japanese	5,1%
Spanish	4,7%
French	4,1%
Chinese	2,9%
Portuguese	2,3%
Italian	1,9%
Polish	1,7%
All the rest	0,7%

the first ten content languages for websites as of 12 March 2014

Global Internet Languages

Language	%
English	27%
Chinese	25%
Spanish	8%
Japanese	5%
Portuguese	4%
German	4%
Arabic	3%
French	3%
Russian	3%
Korean	2%
All the rest	16%
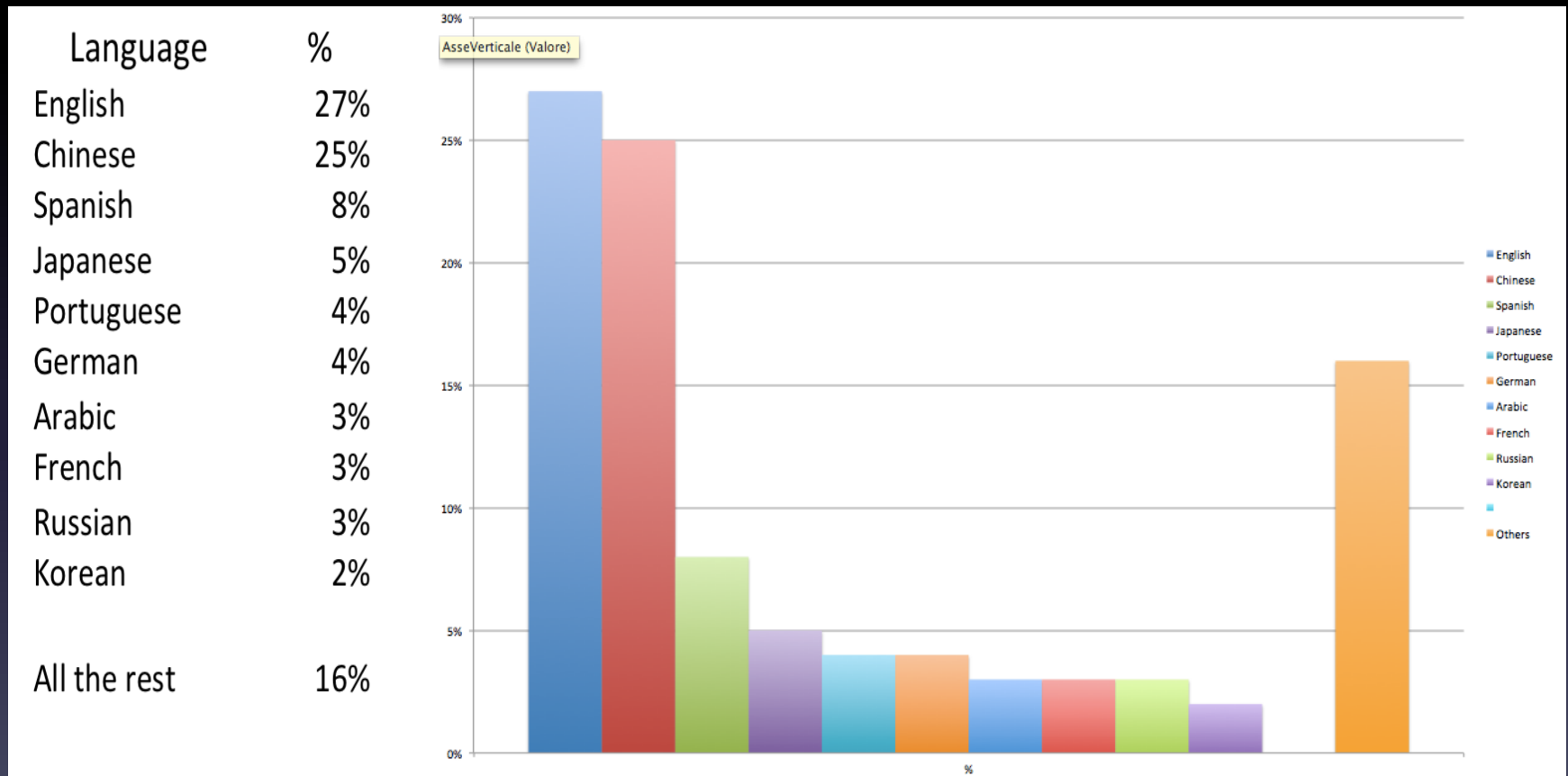
"Number of Internet Users by Language", *Internet World Stats*, Miniwatts Marketing Group, 31 May 2011, accessed 22 April 2012 - explanations on the methodologies used in the survey http://w3techs.com/technologies

New languages, new formats

- *New devices and communication standards are inspiring new languages built on abbreviations, phonetic equivalences, graphic signs and emoticons, will the 140 chars tweet became the new structure of verses?*

- *Smart phones and tablets are breaking time and space barriers including formerly divided people in the emerging cultural phenomenon. This is true both for young generation but even for elderly people that find tablets and smart phones more user friendly than "old" computers.*

- Digital technology is offering new ways to express creativity in different fields: music, images, videos, physical objects and more, enabling young generation to express their feeling and contribute to the creative industries.

mother tongue

"If you talk to a man in a language he understands, that goes to his head. If you talk to him in his language, that goes to his heart"   [Nelson Mandela]

From eContent to Open Data

- Content and services sometimes are built on top of existing data sets, more than ten years ago the European Commission created a framework to improve the added value reuse of public data sets.

- The recently emerged keyword "Open Data" represents one of the nowadays' challenges. Institutions and companies are investing time and resources in order to turn such a concept into reality.

Open Data

Open data refers to the idea that certain data should be freely available for use and re-use.

Dealing with Open Data we must take into consideration, among the others, two main aspects:

- the public body can legally dispose of the processed data using them freely and eventually re-firing them as it may consider useful?

- How it can be wise to behave in managing their rights?

Dealing with Open Data

Classify mandatory datasets not yet publicy available

Amount of errors affecting publicy available datasets

Institutions Evaluation index (errors, conformity)

Preferences in publication formats

European Regulations

- "The Directive on the re-use of public sector information provides a common legal framework for a European market for government-held data (public sector information). It is built around two key pillars of the internal market: transparency and fair competition." (PSI)

- Directive 2003/98/CE – 17 Nov 2003 - GUCE n L 345 del 31 Dec 2003 also known as PSI Directive is the Reference document in the field of Public Bodies data re-use in the European Union

- This Directive aims to facilitate the "creation of products and services in information content, based on public sector documents, extended to the whole community, to promote the effective use across national borders, the use of public sector documents due to private companies, in order to obtain value-added products and services and to limit distortions of competition on the market .

- It focuses on the economic aspects of re-use of information rather than on the access of citizens to information. It encourages the Member States to make as much information available for re-use as possible. It addresses material held by public sector bodies in the Member States, at national, regional and local levels, such as ministries, state agencies, municipalities, as well as organisations funded for the most part by or under the control of public authorities (e.g. meteorological institutes). The Directive covers written texts, databases, audio files and film fragments; it does not apply to the educational, scientific, broadcasting and cultural sectors.

- Direttiva 2013/37/UE - 26 giugno 2013 revision of the Directive has been adopted by the Union legislator. Member States now have 2 years to transpose the provisions of the revised Directive into national law. In July 2014, the Commission published guidelines to help the Member States transpose the revised rules and to indicate best practice in several fields of importance for the re-use of public sector information.

- Directives put obligations only on Member States. Therefore the Commission has adopted a separate decision to allow re-use of its own documents - going beyond the rules of the PSI Directive.

Summary of the Directive

- Charges for re-use have to be limited at a ceiling calculated on the basis of actual costs. Public sector bodies need to calculate charges per re-user in a way so that the total income from charging does not exceed the costs incurred to produce and disseminate the information, together with a reasonable return on investment.

- Public sector bodies are encouraged to apply lower charges or to apply no charges at all. On request, public sector bodies must indicate the method used to calculate charges.

- Conditions for re-use shall be non-discriminatory for comparable categories of re-use.

- Prohibition of cross-subsidies: If public sector bodies re-use their own documents to offer added-value information services in competition with other re-users, equal charges and other conditions must apply to all of them.

- Prohibition of exclusive arrangements: Public sector bodies may not enter into exclusive arrangements with individual re-users, excluding others. Such exclusive rights may only be authorised in exceptional circumstances if they are necessary to provide services in the public interest.

- Charges and other conditions for re-use have to be pre-established and published. If a request for re-use is refused, the grounds for refusal and the means of redress need to be explained.

- Requests for re-use shall be processed within a specific timeframe (20 days for standard cases).

- Licences should not unnecessarily restrict possibilities for re-use or be used to restrict competition.

- Member States are encouraged to use standard licences in digital format.

DIGITAL AGENDA FOR EUROPE
Implementation of the directive

13 Member States have adopted specific PSI re-use measures:

- Belgium (Federal, Flanders, Walloon, French, German, Brussels)
- Cyprus
- Germany
- Greece
- Hungary
- Ireland
- Italy
- Luxembourg
- Malta
- Romania
- Spain
- Sweden
- United Kingdom

3 member stated used previous laws

- Austria
- Denmark
- Slovenia

9 memer states adapted their regulations

- Bulgaria
- Croatia
- Czech Republic
- Estonia
- Finland
- France
- Latvia
- Lithuania
- Netherlands
- Poland
- Portugal
- Slovak Republic

Ownership, Intellectual Property, Privacy, etc

All the Public Bodies are mainly concerned about data ownership, intellectual property, privacy.

These questions take us to directly refer to aspects ranging between:

- the origin of the data set (responsible of the project , data providers , harvesting procedure, ..)
- the procedure activated in order to collect them at the time (who did the harvesting, information type, etc)
- the intellectual rights ownership and transfer (who is the actual owner, which rights have been transfered, etc)
- protection of sensitive data and related citizens' privacy issues
- statistic confidentiality (process to make data anonymous)

-

PSI Guidelines

- Adopted by the College as Commission Communication in form of a Notice

- • Published in the EU Official Journal C Series in all 23 language versions (2014/C 240/01)

- • Non-binding document

- • Gives guidance and best practice examples

- • Useful for the Member States during the transposition phase and afterwards

Licensing

- Licences are not compulsory – if no need to license, a short notice is preferable.

- • The use of open standard licences, such as Creative Commons, is advisable.

- • CC0 is recommended. If it cannot be used, CC-BY or similar attribution-only should be adopted.

- • When drafting national OGLs, the MS are advised to draw inspiration from the guidelines.

- Provisions on scope as broad as possible, with indicative list of examples

- • Definitions – layman's language, concise

- • Exemptions – explicitly indicated

- • Restrictions limited to attribution requirement

- • Other possible licensing terms: versioning, compatibility, liability disclaimer, non-compliance

- • Privacy: reference to EDPS and Art 29 WP, with a possibility to use 'smart notices'.

PSI Guidelines - Datasets

Inspired by international initiatives such as G8 Open Data Charter and the results of the public consultation

- 5 thematic categories of data for priority release:

    o Geospatial data

    o Environment and Earth observation o Transport

    o Statistics

    o Companies data

PSI Guidelines – Datasets

MS can adapt priorities after consulting stakeholders and keeping in mind expected impact (innovation/efficiency/transparency)

A number of recommendations on technical aspects:

- o Stable online location at highest level
- o Original data, complete as far as possible
- o Machine readable and open formats
- o Standard vocabularies, APIs, data-dumps, etc...

PSI Guidelines – Charging

- Reproduction of one more copy and making it available to a re-user

- Marginal cost charging related to data 'distribution' (NOT production/sales/marketing)

- Physical distribution and special requests may generate costs that need to be recovered

- However, online distribution should aim at 'o charge'

- In a cost-recovery model, income generated to be subtracted from total costs (see: EFTA ruling)

- • Quantifiable sets of data to be taken as reference

- • Cultural institutions can add 'IPR clearance' costs and take into account private sector charges without need to apply o/t/v criteria.

- • Suggested limit on RRI at 5% above fixed interest rate

- • Transparency with regard to revenue generated

Privacy Issues

Release and re-use of public bodies datasets may impact citizens' privacy.

- Public domain data (datasets)

- Sensitive data

Personal information represent a wide range of data they include any data concerning any identifiable individual and in some countries applies even to companies if their data may involve individuals.

Typical sensitive data are:Name, Surname, private address, phone, VAT and Social Security numbers, email, car registration plate and even photo and voice recording.

Personal data in addition means physical, physiologic, psychic, economic, social and cultural identity.

Off-limits data are the ones pertaining the intimate sphere of the individual: racial or ethnic origins, religious or philosophical beliefs, political issues, enrolment in political parties, associations plus health conditions, sexual behaviours and more

My data belongs to me

Final remarks

- eContent & Services in local languages

- Open data, availability, ownership, IPR

- Privacy and personal information

S2: Content, Creation, Communication, Copyrights

Does eContent talk to the heart?

Alfredo M. Ronchi

EC MEDICI Framework - Politecnico di Milano

alfredo.ronchi@polimi.it