

**European Language  
Resource Coordination**  
*Connecting Europe Facility*

## Deliverable Task 6

# ELRC Workshop Report for Italy



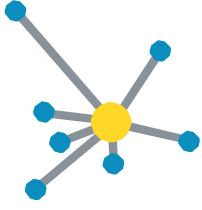
**Author(s):** Paola Baroni, Monica Monachini, Simonetta Montemagni, Claudia Soria (ILC-CNR)  
Serena Di Benedetto (DGT European Commission)

**Dissemination Level:** Public

**Version No.:** <V1>

**Date:** 2016-06-06



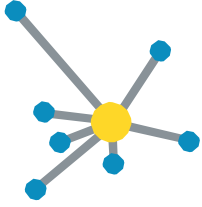


Roma, 15 marzo 2016 **European Language  
Resources Coordination**

---

## Contents

<b>1</b>	<b><u>Agenda</u></b>	<b>4</b>
<b>2</b>	<b><u>Workshop Attendance</u></b>	<b>5</b>
<b>3</b>	<b><u>Summary of Content of Sessions</u></b>	<b>6</b>
3.1	<b>Workshop Opening and Welcome by Local Representatives</b>	<b>6</b>
3.2	<b>Session S1: “Welcome by ELRC and Goal of the Workshop”</b>	<b>6</b>
3.3	<b>Session S2: “Europe and Multilingualism”</b>	<b>7</b>
3.4	<b>Session S3: “Languages and Language Technologies in Italy”</b>	<b>7</b>
3.5	<b>Session S4: “Welcome by On. Basso”</b>	<b>8</b>
3.6	<b>Session S6: “ How does machine translation work?”</b>	<b>8</b>
3.7	<b>Session S7: “How can Public Institutions benefit from the CEF.AT Platform?”</b>	<b>8</b>
3.8	<b>Session S8: “What Data is needed and Why?”</b>	<b>8</b>
3.9	<b>Session S9: “Legal framework for contributing data”</b>	<b>9</b>
3.10	<b>Session S11: “Data and Language Resources: Technical and Practical Aspects”</b>	<b>9</b>
3.11	<b>Session S12: “How can we participate?”</b>	<b>9</b>
3.12	<b>Session S13: “ Summary and next steps”</b>	<b>9</b>
<b>4</b>	<b><u>Synthesis of Workshop Discussions</u></b>	<b>10</b>
4.1	<b>Panel 1 - Session S5: “Multilingualism in Italian Public Services”</b>	<b>10</b>
4.2	<b>Panel 2 - Session S10: “ Data and Linguistic Resources for Public Administration in Italy”</b>	<b>11</b>
4.3	<b>Italian Brochure</b>	<b>13</b>



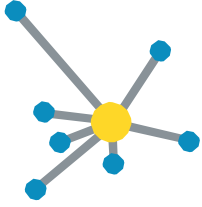
Roma, 15 marzo 2016 **European Language Resources Coordination**

---

### **Executive Summary**

This document reports on the ELRC Workshop in Italy, which took place on the 15<sup>th</sup> of March 2016 at the premises of the Italian Representation of the European Commission in Rome. It includes the agenda of the event (section 2) and briefly informs about the content of each individual, interactive and panel workshop session (sections 3 & 4). The event was attended by 55 participants spanning a wide range of ministries and public organisations. From the 55 persons present, 29 were potential data providers who communicated their availability to collaborate with ELRC. The dedicated event webpage can be found at <http://lr-coordination.eu/it/italy>.

The Workshop has been organized exactly in the same period when the Italian Parliament is discussing the multilingual requirements which need to be satisfied for the European Digital Single Market to become a reality. To testify this interest, a member of the Italian Parliament, On. Lorenzo Basso, was invited to welcome the audience of the ELRC Workshop and to illustrate the ongoing debate on the topic. It is also worth mentioning here the invitation of Simonetta Montemagni as a Computational Linguistics expert for a hearing on the topic of multilingualism and the Digital Single Market by the parliamentary boards “IX Trasporti” and “X Attività Produttive”: the hearing was held just after the Workshop, on March 23<sup>rd</sup>.



# Roma, 15 marzo 2016 **European Language Resources Coordination**

## 1 Agenda

**08:45 – 09:15** **Participants Registration**

**09:15 – 9:25** **Opening of the workshop**

- Emilio Dalmonte, Director of the European Commission Representation in Italy
- Riccardo Pozzo, Director of the Department of Social Science and Humanities, Cultural Heritage of the National Council of Research in Italy

**9:25 - 9:35** **Welcome by ELRC and Goals of the Workshop**

- Khalid Choukri, Secretary General of ELRA “European Language Resources Association” and CEO of ELDA on behalf of the “European Language Resource Consortium” (ELRC)
- Simonetta Montemagni, Istituto di Linguistica Computazionale “A. Zampolli” (CNR) and ELRC “Technological Italian National Anchor”

**9:35 – 10:05** **Europe and multilingualism**

- Serena Di Benedetto, EC DGT Field Officer, European Commission Representation in Italy, Rome

**10:05 – 10:35** **Languages and Language Technologies in Italy**

- Simonetta Montemagni

**10:35 – 11:05** **Coffee Break**

**11:05 – 11:15** **Welcome by**

- On. Lorenzo Basso, Rapporteur of the X Parliamentary Commission of the Chamber of Deputies on the EU COM(192) on the Digital Single Market

**11:15 – 12:15** **Panel: Multilingualism in Italian Public Services**

Moderator: Claudia Soria, Istituto di Linguistica Computazionale “A. Zampolli” (CNR)

Panelists:

- Claudia Foti, Ministry of Justice
- Annette Lenz Liebl, Office for Language Issues of the Autonomous Province of Bolzano
- Silvana Schiavi Fachin, International Centre on Multilingualism of the University of Udine
- Alessandra Terrosi, Prefecture of Florence
- Flavia Vecchione, Ministry of Interior

**12:15 – 12:45** **Machine translation: how does it work?**

- Marcello Federico, Fondazione Bruno Kessler, Trento

**12:45 – 13:15** **How can Public Institutions benefit from the CEAF.AT Platform?**

- Spyridon Pilos, EC, DG Translation, Head of Language Applications

**13:15 – 14:15** **Lunch Break**

**14:15 – 14.35** **What Data is needed and why?**

- Khalid Choukri

**14:35 – 15:05** **Legal framework for contributing data**

- Khalid Choukri
- Gabriele Ciasullo, Agency for Digital Italy

**15:05 – 16:00** **Panel: Data and Language Resources for the Public Administration in Italy**

Moderator: Monica Monachini, Istituto di Linguistica Computazionale “A. Zampolli” (CNR) and National Coordinator of CLARIN-IT

Panelists:

- Barbara Altomonte, Rete per l'italiano istituzionale (REI)
- Gabriele Ciasullo, Agency for Digital Italy
- Alessandro Lenci, University of Pisa
- Maria Teresa Paziienza, University of Roma “Tor Vergata”
- Marcello Soffritti, Institute for Specialised Communication and Multilingualism, EURAC Research, Bolzano

**16:05 – 16:30** **Coffee Break**

**16:30 – 16:50** **Data and language resources: Technical and practical aspects**

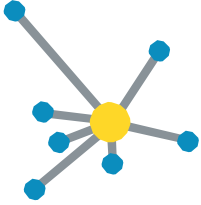
- Khalid Choukri

**16:50 – 17:20** **How can we participate?**

Interactive session moderated by Flavia Marzano, Associazione Stati Generali dell'Innovazione

**17:20 – 17:30** **Summary and next steps**

- Khalid Choukri and Simonetta Montemagni



*Roma, 15 marzo 2016* **European Language  
Resources Coordination**

---

## 2 Workshop Attendance

67 persons from 39 institutions, many of which were from public service translation departments, pre-registered the workshop.

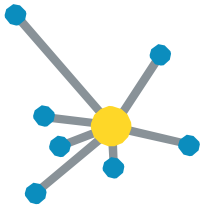
25 persons participated actively in the workshop either by presenting content or organizing the event or interpreting from and to English and Italian.

12 pre-registered persons from 12 institutions did not attend the workshop.

In total we had 55 persons from 32 institutions present during the workshop.

From these 55 persons, 29 persons from 17 institutions were potential data providers: six of them (namely, Barbara Altomonte, Gabriele Ciasullo, Claudia Foti, Annette Lenz Liebl, Alessandra Terrosi, Flavia Vecchione) participated actively as speakers.

Among the participants, 17 filled the feedback forms and 12 gave their contact details for being further involved with the ELRC activities.



## Roma, 15 marzo 2016 **European Language Resources Coordination**

### 3 Summary of Content of Sessions

#### 3.1 Workshop Opening and Welcome by Local Representatives

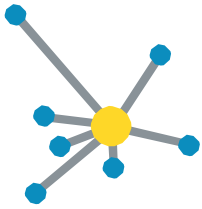
Claudia De Stefanis, Head of the Communication Sector of the EC Representation in Rome, welcomed the participants. She was particularly pleased to open the event, she said, because as a former assistant to DGT Director General she had followed closely the development of the MT@EC project. MT@EC main strength, she added, lays in the rich linguistic resources underpinning the system, a real treasure of high-quality human translation memories produced by DGT. This system is already supporting online services like IMI, SOLVIT and – very recently - the ODR platform, and is at the heart of the CEF platform. Thanks to DGT contribution to the internal consultation on the amended proposal for a Regulation on guidelines for trans-European telecommunication networks, she recalled, machine translation is among the “building blocks” of digital service infrastructures, and as such among the funding priorities in the CEF program. To support an increasing number of on-line public services the system needs to be further developed and adapted to the specific needs of these services as well as to the potential volumes of the new requests that will have to be addressed. This is why high-quality corpora are needed. She concluded by wishing all the participants a fruitful day of work and encouraging them to contribute to the project.

The welcome by Riccardo Pozzo, Head of the Department of Social Science and Humanities, Cultural Heritage of the National Council of Research of Italy (to which the “Istituto di Linguistica Computazionale A. Zampolli” belongs), was postponed immediately before the morning coffee break due to unexpected institutional duties. He welcomed the audience and pointed out the key role that Language Resources and technologies can play in Europe for a fully functional and multilingual Digital Single Market, with particular emphasis on their role for the integration of immigrants in the European society.

#### 3.2 Session S1: “Welcome by ELRC and Goal of the Workshop”

Simonetta Montemagni (ILC-CNR), the local ELRC Technology National Anchor Point, welcomed the participants to this important event aimed at fostering the development of a real multilingual Digital Single Market including Italy. The success of the event shows that the topic is of great interest and may help identify the needs and shared interests. In this respect, she reported an interesting coincidence: exactly in these days two Boards of the Chamber of Deputies of the Italian Parliament are discussing the Communication COM (2015) 192 from the European Commission on a Digital Single Market Strategy for Europe. She thanked the EC Representation in Italy for their valuable support in co-organizing and hosting the Workshop, Khalid Choukri, the ELRC representative for the region, and Spyridon Pilos, Head of Language Applications at the DG Translation, all speakers who accepted to contribute to the Workshop, and the participants who were numerous and from different Italian Public Administrations, Regions, research centers and companies.

Khalid Choukri (ELDA) started by thanking the EC Representation in Italy for their support to the event and the organizers from the Istituto di Linguistica Computazionale which bears the name of Antonio Zampolli, who played a key role in the foundation of ELRA. He also invited everybody to participate and to share their hopes and questions about the importance of Language Resources for Language Technologies. He continued by presenting the European context where language diversity is the basis of the European culture and identity. In this context, he introduced the automated translation services of the Connecting Europe Facility (CEF.AT), which focus on the translation needs of citizens and public bodies and are free of charge for public administrations in all Member



## Roma, 15 marzo 2016 **European Language Resources Coordination**

States, Iceland and Norway. To ensure that CEF.AT provides the best possible service, high-quality data is needed. He then illustrated the objectives of the workshop, which mainly stand in working with public services and administrations to find and share appropriate data and get the best results from the CEF.AT Machine Translation Technology for their needs and language. By sharing the right kind of data with CEF.AT, public administrations will receive better support for their own language (Italian in the case at hand) and multilingualism in Europe. Khalid Choukri concluded by emphasizing the fact that supporting our languages is supporting Europe and supporting Europe is supporting our languages.

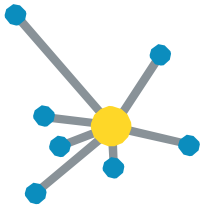
### 3.3 Session S2: “Europe and Multilingualism”

The presentation of Serena Di Benedetto, DGT Field Officer in Rome, provided a framework for the workshop, giving an overview of the legal basis of multilingualism in Europe and highlighting the political commitment of the EU towards the protection of the linguistic diversity. Multilingualism is also key for the Digital Single Market strategy, as both buyers and sellers show a marked preference for working/buying in their own language. In order to ensure quality and cope with the increasing workload, the EU translation services have adopted and developed several technological tools, including machine translation (MT@EC). This mature technology is at the heart of the effort to support pan-European public services through the CEF Automated Translation platform. To produce good-quality translations, the platform needs linguistic data from national public administrations.

### 3.4 Session S3: “Languages and Language Technologies in Italy”

First, Simonetta Montemagni presented multilingualism in Italy, which has one official language and twelve co-official ones (including local languages spoken by minority populations and regional border languages). In addition, in daily life, multilingualism in Italy is extended by tourism, trade, cooperation between regions of Europe, cross-border legal issues, international relations as well as by immigration. Languages involved are not restricted to European languages and also include – among others - Chinese and Arabic. For what concerns Italians’ attitude towards learning and using foreign languages, she pointed out that Italy is one of the countries where citizens are least likely to be able to speak a foreign language: according to the survey Eurobarometer 386 “Europeans and their languages” (2012), 62% of the Italian citizens do not speak any foreign language.

In the second part of the presentation, she introduced language technologies and the challenges connected to the automatic analysis of the Italian language and in particular of the language variety used by Italian public administrations. She then referred to the META-NET White Paper Series to describe the language technologies for Italian and introduced the research centers and companies working on language technologies spread across Italy (nowadays gathered within the newly-founded “Associazione Italiana di Linguistica Computazionale”, AILC). She concluded by emphasizing that multilingual communication between European public administrations requires the specialization of the machine translation system CEF.AT with respect to the domains and textual genres to be dealt with. Such a specialization can only be achieved by gathering textual and linguistic data available from Italian public administrations, which can significantly contribute to tailoring and improving the CEF.AT results for Italian.



## Roma, 15 marzo 2016 **European Language Resources Coordination**

### 3.5 Session S4: “Welcome by On. Basso”

To testify the interest of the Italian Parliament on the topic, Lorenzo Basso, Member of the Chamber of Deputies, welcomed the participants to the Workshop and illustrated the current discussion and activities of the Boards “IX Trasporti” and “X Attività Produttive” on the topic of the Digital Single Market with a specific focus on the need to address the challenge of multilingualism.

### 3.6 Session S6: “ How does machine translation work?”

Marcello Federico provided an overview on Machine Translation (MT), highlighting pros and cons of both human translation and machine translation. He claimed that integration of both systems is increasingly gaining attention in research and industry with a lot of work in progress. Nowadays, machine learning algorithms are part of machine translation systems. Statistical machine translation learns how to translate from existing translations (parallel corpora). The first phase is the training phase, in which a statistical model of translation is built, using a corpus of texts in both the source and target languages. The training phase is itself split into sentence alignment, word alignment and phrase extraction. The second phase is the translation phase, aimed at finding the best sequence of translations: the algorithm evaluates each sequence by calculating the probability of all phrase translations and their relative order, and of the fluency of the translation. This explains why data is very important for developing MT technology.

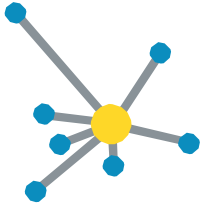
### 3.7 Session S7: “How can Public Institutions benefit from the CEF.AT Platform?”

Spyridon Pilos gave his presentation through Skype. He began his presentation by illustrating the typical interactions between administrations, citizens, and businesses within a Member State as well as between Member States or between Member States and the European Union and proposed the vision of accessing any service or business in any Member State using its own language. After illustrating the wider context, he elaborated on the reasons why MT was the only viable solution to tackle the challenge of multilingualism in Europe. He provided a detailed description of the MT@EC system, i.e. the supported languages, the base technologies, its user interface and other technical features, including input data format, results output and security in document transfer. He also presented the EU online services connected to MT@EC and invited the public sector representatives attending the workshop to access it. He concluded by emphasizing that the focus will be the coverage of more domains and increasingly in-domain text are needed to implement this move at sufficiently high quality.

### 3.8 Session S8: “What Data is needed and Why?”

Khalid Choukri detailed the data to be collected for MT: texts (translations, aligned translations, collections of comparable texts), glossaries, terminological databases, dictionaries and lists of words in one or more languages. ELRC is looking for multi-lingual data but mono-lingual data are also of interest to ELRC. He discussed how different representation formats (e.g. \*.tmx, \*.xliff, \*.txt, \*.doc, \*.docx, \*.odt, \*.ppt) are useful to varying degrees and stressed the importance of metadata (e.g. based on Dublin core). He showed several examples of how language resources are produced from data. Finally, he encouraged the participants to contribute to the project by providing or identifying the required textual content within their organizations.





## Roma, 15 marzo 2016 **European Language Resources Coordination**

### 3.9 Session S9: “Legal framework for contributing data”

Khalid Choukri presented the need for a clear and easy-to-follow regime for data re-use across the EU with the objective of developing innovative applications based on public data, for the the European DSM. In particular, he described how the Public Sector Information (PSI) rules addressed data re-use across the EU and legal and technical interoperability. He briefly presented the Public Sector Information (PSI) Directive and explained the structure of the data rights. For what concerns ELRC, Khalid Choukri explained that data will not be re-used as such but will rather be used to create language models, from which it is impossible to reconstruct the original data. The second part of the presentation was given by Gabriele Ciasullo, responsible for the area “Open Data” within the Agency for Digital Italy, who provided an overview of the national situation on sharing and re-use of public data based on the Italian Digital Administration Code (“Codice Amministrazione Digitale”).

### 3.10 Session S11: “Data and Language Resources: Technical and Practical Aspects”

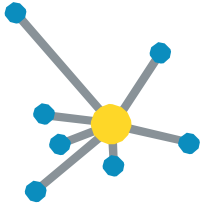
Khalid Choukri (ELDA) presented the detailed workflow for the collection, processing and sharing of Language Resources. Most of this process stages, e.g. the identification of the data sources and datasets, the basic metadata documentation, data cleaning and privacy and ethics management, are tasks in which the public sector providers will collaborate with ELRC. The presenter encouraged the audience to participate in these activities and work together with ELRC, and he showcased the mechanisms with which ELRC will fully support the providers throughout the whole process, i.e. the helpdesk and user forum, the ELRC repository and the ELRC website.

### 3.11 Session S12: “How can we participate?”

This interactive session was moderated by Flavia Marzano (Associazione Stati Generali dell’Innovazione). The discussion outlined the interest of the participants to get involved in the ELRC initiative and to contribute with different mono- and multi-lingual language resources available at the institutions they represented: e.g. aligned translations of different typologies of texts; monolingual corpora, possibly enriched with linguistic annotations; domain dictionaries and terminologies. In some cases, issues about the IPR and legal status have still to be clarified. Some of the participants engaged themselves to notify and inform the heads of their departments and the higher staff at their institutions about the topics discussed at the Workshop.

### 3.12 Session S13: “ Summary and next steps”

Khalid Choukri and Simonetta Montemagni summarised the day in presenting the next steps of the ELRC initiative for what concerns Italy. They explained that the goal of the ELRC workshop was to illustrate to the Italian public bodies that their data is valuable, and that everyone can benefit from data sharing given the foreseen consequences for what concerns the CEF.AT Platform. They also stressed the importance of maintaining this dialogue open, to proceed with the identification, for Italian, of the data sources and datasets, with the basic metadata documentation and with data cleaning: these are tasks in which the Italian public sector providers will collaborate with ELRC. They concluded by thanking everyone for attending, in particular the host, the interpretations team, the



## Roma, 15 marzo 2016 **European Language Resources Coordination**

speakers and the audience and invited all participants to fill the distributed feedback and ELRC participation forms.

### 4 Synthesis of Workshop Discussions

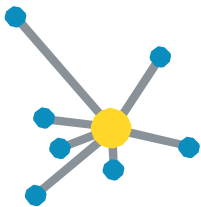
#### 4.1 Panel 1 - Session S5: “Multilingualism in Italian Public Services”

The panel was moderated by Claudia Soria. The panelists were representatives from the Ministry of Justice, the Ministry of Internal Affairs, the Office for Linguistic Issue of the Autonomous Province of Bolzano – Alto Adige, the Prefecture of Florence, the International Center on Plurilingualism of the University of Udine and ARLeF.

Claudia Foti, translator of the Ministry of Justice, explained how translations are performed internally by around 30 translators for English, French, German, and Spanish. Other languages are covered by resorting to external services. The type of documents to be dealt with mainly regards criminal matters: rogatory letters, verdicts, and correspondence related to international judicial cooperation. Most of the work concerns the Italian-Romanian, Italian-Albanian, and Italian-English language pairs. Foti reported how MT is very little used, and mostly regarded with suspicion. This lack of knowledge about the benefits and advantages of MT represents one of the biggest issues, together with data privacy, high variability of data formats, and copyright.

Flavia Vecchione gave an overview of the linguistic resources in use in the Ministry of Internal Affairs. The lack of a centralised linguistic service was one of the major problems reported; the translation work concerns around 350 translators, working in various offices and departments (e.g. Department of Public Security, police headquarters, Border Police, etc.). The Ministry of Internal Affairs works in 11 languages (English, French, German, Spanish, Portuguese, Russian, Slovenian, Arabic, Albanian, Chinese and Turkish); translations are outsourced only for less widely used languages and for non-classified information. Another important feature of the translation work being carried out is that translation can cover many different domains and is crucially bound to deep knowledge of foreign cultures: one of the main problems is the lack of equivalence between concepts, institutions, abbreviations and acronyms. No use is being done altogether of MT. The available multilingual resources are not systematically managed, and most terminologies and translation memories are not shared among colleagues. In the future, one can expect a rise in the demand for translation from and into Arabic and Chinese and other non-European languages, as well an increase in the volume of translation from and into English, French and Spanish, that have become the most important vehicular languages. MT could be profitably employed for the translation of modules, forms, standardised information and institutional web pages. Documents containing sensible data or classified information still pose substantial juridical problems.

Alessandra Terrosi, from the Prefecture of Florence, showed the immigration service P.A.eS.I. It was created as part of the project P.A.eS.I. (Public Administration and Foreign Immigrants, <http://www2.immigrazione.regione.toscana.it/>), in order to provide an electronic access to information and services involving public and foreign governments. The PAeS.I project, included in the Regional Action Plan eToscana, aims to create, in an integrated and coordinated way, a circularity and a sharing of information among the public administration, professional associations, and professionals for the streamlining and transparency of administrative procedures and building a correct relationship between the public and private systems.



## Roma, 15 marzo 2016 **European Language Resources Coordination**

Annette Lenz-Liebl, who represented the Office for Linguistic Issue of the Autonomous Province of Bolzano – Alto Adige, introduced the main duties and responsibilities of the Office. The Office operates in a trilingual region, and in addition to its main translation services, it provides linguistic assistance and consultancy, under the form of revision services or as consultancy and linguistic training for public officers. The terminology developed during the translation or revision services is collected in a shared database. The Office is particularly interested in evaluating new opportunities deriving from assisted translation, with the aim of facilitating the work of translators, reducing costs and avoiding duplication of efforts.

Finally, Silvana Schiavi Fachin (International Center on Plurilingualism of the University of Udine and ARLeF) presented a survey of the multilingual policies implemented by the websites of public administrations of Friuli-Venezia-Giulia region, noting the lack of multilingualism of the three minority languages that are spoken in the region (Slovene, German and Friulian). In order to improve the bilingualism Italian-Friulian, the Region had distributed to all public services an automatic translation software Italian-Friulian, Jude.

### 4.2 Panel 2 - Session S10: “ Data and Linguistic Resources for Public Administration in Italy”

The panel, moderated by Monica Monachini, was aimed at discussing the topic of data and linguistic resources for the Public Administration in Italy. The topic was dealt with from a three-fold perspective, namely of:

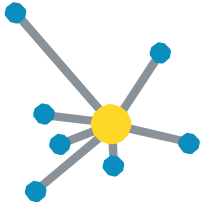
1. who already designed and developed linguistic resources, either corpora or lexica, for the Public Administration domain in Italy;
2. who built the Italian national catalogue of Public Administration data;
3. who provides support in the preparation of Italian Public Administration documents.

The panel was thus articulated in three different types of interventions, each addressing a different set of questions.

The first set of speakers was asked to address the following issues: 1) type of developed resource(s) in projects focusing on Public Administration data (e.g. lexicons, terminologies, corpora); 2) main features of developed resources (Mono- vs multi-lingual, raw vs. enriched with annotations, cross-lingual alignment); 3) representation format; 4) development methodology (manually, semi-automatically, automatically); 5) availability and distribution policy (with particular attention to whether they will be made available to ELRC).

Alessandro Lenci (University of Pisa) illustrated the experience gained and the resources developed in the project SEMPLICE (“Semantic Instruments for Public Administrators and Citizens”, funded by the Tuscany Region) for semantic indexing of Public Administration documents: namely, the PAWaC corpus (Public Administration Web as a Corpus), a monolingual corpus of 25.218.385 tokens with different types of administrative documents from municipalities automatically lemmatized and PoS-tagged; a Public Administration ontology including 2700 terms which were manually revised and structured; a corpus of 100 documents with manually revised Named Entity annotation.

Maria Teresa Pazienza (University of Roma Tor Vergata) presented the activities carried out within the framework of successful collaborations between the University of Roma Tor Vergata and Public Administrations. She illustrated the NLP-based methodology they developed for the design and construction of resources such as: VocBench (in collaboration with FAO), a system for the development of multilingual thesauri; AGROVOC, a multilingual controlled vocabulary for Food and



## Roma, 15 marzo 2016 **European Language Resources Coordination**

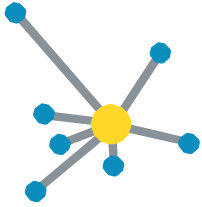
Agriculture, available in LOD. Both of them are already being used by national and international public bodies.

Marcello Soffritti (Institute for Specialized Communication and Multilingualism, EURAC Bolzano and Department of Interpreting and Translation of the University of Bologna) presented different resources for automatic translation and assisted translation, all of which can be made available to ELRC. These resources are multilingual and range from corpora to terminologies and translation memories: they were built either manually or semi-automatically to meet the specific needs of different collaborations/projects. For these resources to be re-used and possibly combined strong actions are required, ranging from technical validation to format conversion, up to the availability of a technical infrastructure, in order to promote a radically new planning of the workflows which produce bilingual or multilingual versions of lexicons and corpora.

Gabriele Ciasullo contributed to the panel by illustrating the activities of the Agency for Digital Italy (AgID) for what concerns databases and open data of national interest. The questions addressed were concerned with used representation formats, distribution of data, collection, storing and maintenance, availability and current usages. He discussed the National Catalogue of PA which aims at facilitating the dissemination, knowledge and sharing of data between public administrations and their reuse according to the principles of open data. He also described the Repertorio Nazionale dei Dati Territoriali ([rndt.gov.it](http://rndt.gov.it)) and INSPIRE, the national Infrastructure for Spatial Information in Europe which provides metadata-based services for searching and displaying the content of spatial data sets. He also presented the portal [dati.gov.it](http://dati.gov.it). He concluded by claiming that open data is the driving force of the economy.

The panel ended with a contribution by Barbara Altomonte, from REI, Rete per l'Italiano Istituzionale ([http://ec.europa.eu/translation/italian/rei/index\\_it.htm](http://ec.europa.eu/translation/italian/rei/index_it.htm)). The questions were concerned with the typology of linguistic resources built within or simply available through REI and who is the expected target for these resources. She stressed that REI wants to provide ELRC not just with a set of data and resources, but also with a shared vision: translated text has an ethical value, and quality is the means to pursue this value. Citizens should be able to retain the right to use their mother tongue and receive basic services in their own language. Translated text should be clear, easy and understandable: this is also a pre-requisite for automatic processing. There is data already available in two languages that public administrations could make available to ELRC. The most crucial documents for citizens and migrants are the civil registry authority services.





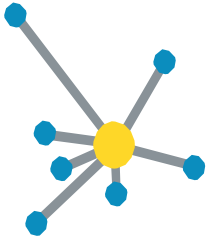
Roma, 15 marzo 2016 **European Language  
Resources Coordination**

4.3 Italian Brochure

# **VERSO UN ACCESSO AI SERVIZI PUBBLICI SENZA BARRIERE LINGUISTICHE**

Workshop ELRC Italia  
Roma, 15 marzo 2016





## I veri confini nell'Europa digitale

La creazione di un Mercato Unico Digitale (o “Digital Single Market”) europeo rappresenta un obiettivo chiave della Commissione europea per il 2020.

Se da un lato il Mercato Unico Digitale è fondamentalmente multilingue, dall'altro finora la maggior parte delle lingue europee non ha potuto beneficiare di un adeguato supporto da parte delle tecnologie del linguaggio che rendesse possibile un lavoro, una vita e un commercio realmente senza frontiere.

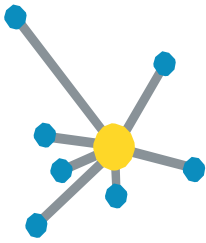
In effetti, la ricchezza costituita dalla diversità e varietà linguistica dell'Europa si trasforma in un ostacolo se il libero scambio di informazioni viene intralciato, se l'accesso imparziale ai servizi pubblici, alle opportunità commerciali, di lavoro e di sostegno vengono impediti.

Nel nascente Mercato Unico Digitale, le barriere linguistiche creano confini invisibili: mentre il 99% delle imprese europee è costituito da piccole e medie imprese (PMI), solo il 7% di esse riesce a vendere al di là dei confini, a consumatori che parlano una lingua diversa.

Analogamente, il 90% dei clienti europei preferisce navigare all'interno di siti web nella propria lingua.

Un supporto adeguato sul versante delle tecnologie linguistiche garantirebbe il multilinguismo portando vantaggi significativi sia per l'economia sia per la società: si stima che un Mercato Unico Digitale potrebbe effettivamente creare fino a 340 miliardi di Euro di crescita aggiuntiva, centinaia di migliaia di nuovi posti di lavoro e una società realmente basata sulla conoscenza.





## Benefici derivanti dalle nuove opportunità

La Commissione europea ha compiuto un passo importante a supporto di un Mercato Unico Digitale veramente multilingue, creando i presupposti per servizi pubblici dedicati ai cittadini e alle imprese europee in grado di operare liberamente senza barriere linguistiche.

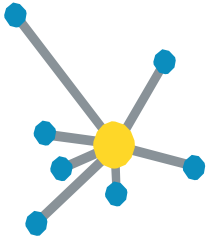
La piattaforma di Traduzione Automatica sviluppata all'interno del programma europeo "Connecting Europe Facility" (CEF.AT) faciliterà la comunicazione multilingue e lo scambio di documenti e altri contenuti linguistici in Europa tra amministrazioni pubbliche nazionali, e tra queste e i cittadini europei e le imprese.

La piattaforma CEF.AT opererà in vari scenari della pubblica amministrazione in aree come, ad esempio, la protezione dei consumatori, la salute, gli appalti pubblici, la previdenza sociale e la cultura.

Svolgerà un ruolo centrale nel far crescere servizi pubblici online, come Europeana, il portale degli Open Data e la piattaforma online di risoluzione delle controversie.

Con il supporto del servizio CEF.AT, le pubbliche amministrazioni in Europa potranno fare un importante passo avanti che permetterà loro di operare riducendo o addirittura abbattendo le barriere linguistiche.





## La tecnologia alla base di servizi pubblici senza frontiere

La tecnologia alla base della piattaforma CEF.AT è rappresentata da un sistema di traduzione automatica di tipo statistico, MT@CE, che impara come tradurre un testo a partire da traduzioni esistenti.

All'interno di un approccio di tipo statistico alla traduzione automatica, dati come raccolte di testi mono- e bi-lingui, dizionari, terminologie, ontologie e repertori di nomi propri svolgono un ruolo centrale in quanto forniscono l'evidenza sulla base della quale il sistema di traduzione automatica "impara" a tradurre un testo da una lingua sorgente a una lingua target.

MT@CE nel suo stato attuale è disponibile gratuitamente per tutti i servizi pubblici europei, nell'ambito del programma CEF.

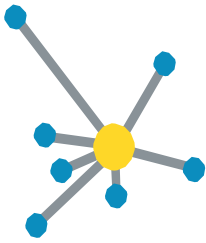
Ad oggi, il sistema è addestrato principalmente sulle traduzioni della legislazione della UE e sui testi prodotti dalle istituzioni dell'UE: questo fatto rende il sistema non sempre in grado di soddisfare appieno le esigenze e le richieste che emergono quotidianamente da parte dei servizi pubblici nazionali europei.

Per migliorare il servizio di traduzione automatica offerto dalla Commissione europea, è necessario specializzarlo in relazione alle esigenze e peculiarità delle amministrazioni pubbliche nazionali ed europee: tale obiettivo può essere perseguito arricchendo l'evidenza linguistica a disposizione del sistema di traduzione automatica con dati testuali e linguistici (mono- e bi-lingui) raccolti dalle istituzioni nazionali come ministeri, governo, amministrazioni pubbliche e organizzazioni non governative.

Attraverso l'azione "European Language Resources Coordination" (ELRC) all'interno del programma europeo "Connecting Europe Facility", la Commissione europea sta intraprendendo uno sforzo di raccolta di dati testuali e linguistici senza precedenti come primo passo nella direzione della specializzazione del CEF.AT in relazione alle esigenze e peculiarità dei servizi pubblici degli Stati membri dell'UE, dell'Islanda e della Norvegia, creando così i presupposti per l'erogazione di servizi multilingui per i cittadini, le amministrazioni e le imprese in Europa.

In tal modo, l'azione ELRC non solo aiuterà a colmare il divario tra i servizi di traduzione automatica attualmente offerti dalla Commissione europea e le reali esigenze quotidiane dei servizi pubblici nazionali di tutta Europa, ma contribuirà anche in modo significativo alla sopravvivenza delle lingue nazionali europee.





## Workshop ELRC Italia

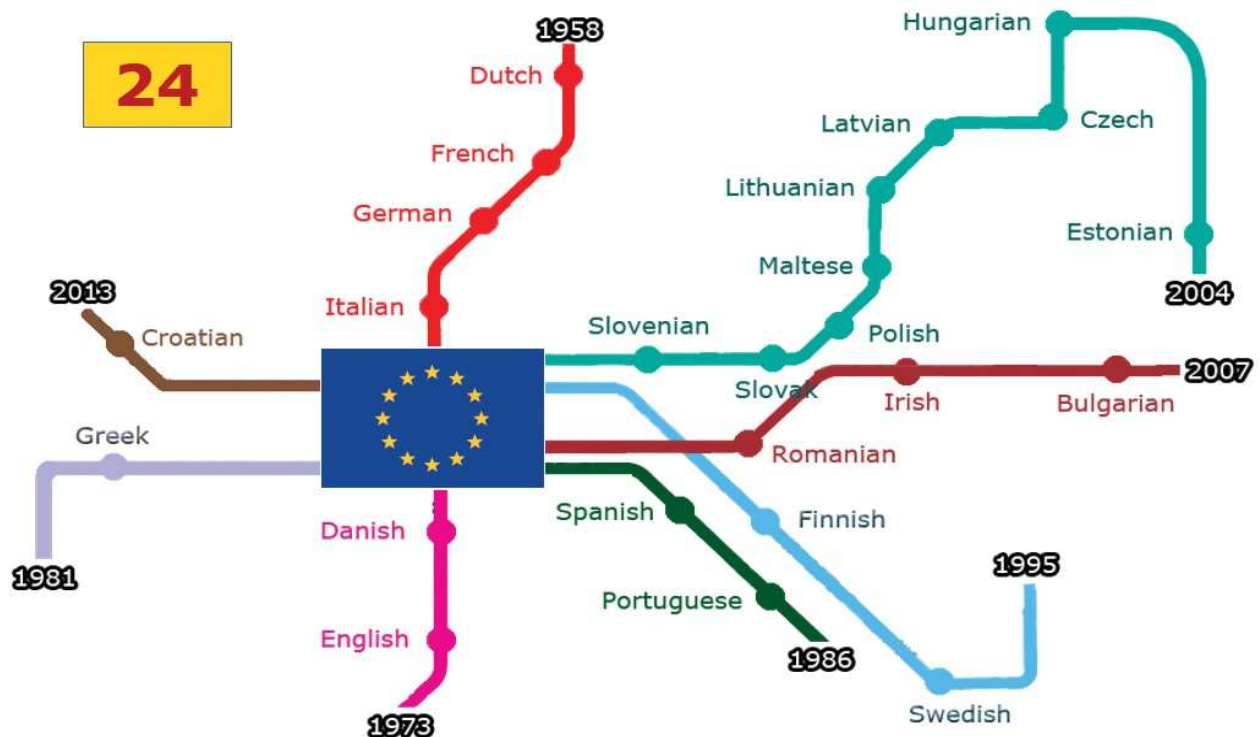
A nome della Commissione europea, il consorzio ELRC sta organizzando una serie di seminari nazionali in ciascuno dei 30 paesi partecipanti per aumentare la consapevolezza circa l'importanza dei dati testuali e linguistici disponibili presso le pubbliche amministrazioni: si tratta di un patrimonio testuale che può contribuire in modo significativo alla specializzazione dei risultati di sistemi di traduzione automatica e di cui non vi è sempre consapevolezza circa il suo valore.

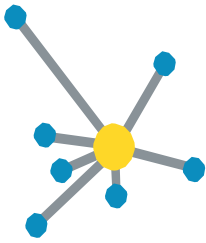
Al fine di coinvolgere tutte le parti interessate nel rendere la piattaforma CEF.AT a misura delle loro esigenze, ELRC invita tutti i rappresentanti delle amministrazioni pubbliche in Europa a partecipare all'evento in corso di organizzazione nel proprio paese.

In Italia, l'organizzazione dell'evento è stata affidata all'Istituto di Linguistica Computazionale "Antonio Zampolli" del Consiglio Nazionale delle Ricerche, che opera nel settore della Linguistica Computazionale dal 1967.

Il **Workshop ELRC Italia** (<http://lr-coordination.eu/it/italy>) si svolge il 15 marzo 2016 a Roma presso lo Spazio Europa della Rappresentanza in Italia della Commissione europea (Palazzo dei Campanari, Via IV Novembre 149).

Il workshop vede coinvolti relatori della Commissione europea, rappresentanti delle amministrazioni pubbliche nazionali e degli enti governativi, per arrivare a fornitori di servizi linguistici e di tecnologie del linguaggio, in un confronto mirato a comprendere le esigenze delle amministrazioni nazionali del settore pubblico in materia di traduzione automatica, ad effettuare una ricognizione e una raccolta di dati relativi alla lingua italiana e a discutere di problemi tecnici e giuridici connessi con il loro uso



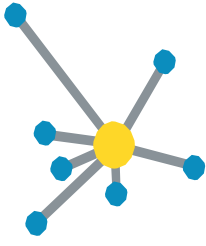


## Programma

---

08:45– 09:15	<b>Registrazione dei partecipanti</b>
09:15 – 09:25	<b>Apertura dei lavori e benvenuto</b> <ul style="list-style-type: none"><li>• Emilio Dalmonte, Direttore della Rappresentanza in Italia della Commissione europea</li><li>• Riccardo Pozzo, Direttore del Dipartimento Scienze Umane e Sociali e Patrimonio Culturale del Consiglio Nazionale delle Ricerche</li></ul>
09:25 – 9:35	<b>Benvenuto di ELRC</b> <ul style="list-style-type: none"><li>• Khalid Choukri, Segretario Generale di ELRA “European Language Resources Association” a nome dello “European Language Resource Consortium” (ELRC)</li><li>• Simonetta Montemagni, Istituto di Linguistica Computazionale “A. Zampolli” (CNR) e ELRC “Technological Italian National Anchor”</li></ul>
9:35 – 10:05	<b>L’Europa e il Multilinguismo</b> <ul style="list-style-type: none"><li>• Serena Di Benedetto, Antenna della Direzione generale della Traduzione, Rappresentanza in Italia della Commissione europea</li></ul>
10:05 – 10:35	<b>Lingue e Tecnologie Linguistiche in Italia</b> <ul style="list-style-type: none"><li>• Simonetta Montemagni</li></ul>
10:35 – 11:05	<b>Pausa Caffè</b>
11:05 – 11:15	<b>Intervento</b> <ul style="list-style-type: none"><li>• On. Lorenzo Basso, Relatore X Commissione della Camera dei Deputati sulla COM(192) UE sul Digital Single Market</li></ul>
11:15 – 12:15	<b>Tavola Rotonda: Servizi Pubblici Multilingui in Italia</b> <p><i>Moderatore</i></p> <ul style="list-style-type: none"><li>• Claudia Soria, Istituto di Linguistica Computazionale “A. Zampolli” (CNR)</li></ul> <p><i>Partecipanti</i></p> <ul style="list-style-type: none"><li>• Claudia Foti, Ministero della Giustizia</li><li>• Annette Lenz Liebl, Ufficio Questioni Linguistiche della Provincia Autonoma di Bolzano</li><li>• Silvana Schiavi Fachin, Centro Internazionale sul Plurilinguismo dell’Università di Udine</li><li>• Alessandra Terrosi, Prefettura di Firenze</li><li>• Flavia Vecchione, Ministero dell’Interno</li></ul>
12:15 – 12:45	<b>Traduzione Automatica: Come funziona?</b> <ul style="list-style-type: none"><li>• Marcello Federico, Fondazione Bruno Kessler, Trento</li></ul>

---

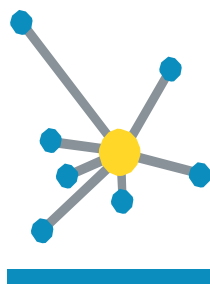


## Programma

---

12:45 – 13:15	<b>Quali vantaggi per le Istituzioni Pubbliche dalla Piattaforma di traduzione automatica CEF.AT?</b> <ul style="list-style-type: none"><li>• Spyridon Pilos, Direzione generale della Traduzione, Commissione europea</li></ul>
13:15 – 14:15	<b><i>Pausa Pranzo</i></b>
14:15 – 14:35	<b>Quali dati? Perché?</b> <ul style="list-style-type: none"><li>• Khalid Choukri</li></ul>
14:35 – 15:05	<b>Aspetti legali della condivisione dei dati</b> <ul style="list-style-type: none"><li>• Khalid Choukri</li><li>• Gabriele Ciasullo, Agenzia per l'Italia Digitale</li></ul>
15:05 – 16:05	<b>Tavola Rotonda: Dati e Risorse Linguistiche per la Pubblica Amministrazione in Italia</b> <p><i>Moderatore</i></p> <ul style="list-style-type: none"><li>• Monica Monachini: Istituto di Linguistica Computazionale “A. Zampolli” (CNR) e Coordinatore Nazionale di CLARIN-IT</li></ul> <p><i>Partecipanti</i></p> <ul style="list-style-type: none"><li>• Barbara Altomonte, Rete per l'italiano istituzionale (REI)</li><li>• Gabriele Ciasullo, Agenzia per l'Italia Digitale</li><li>• Alessandro Lenci, Università di Pisa</li><li>• Maria Teresa Pazienza, Università di Roma “Tor Vergata”</li><li>• Marcello Soffritti, Istituto di Comunicazione Specialistica e Plurilinguismo dell'EURAC Research, Bolzano</li></ul>
16:05 – 16:30	<b><i>Pausa Caffè</i></b>
16:30 – 16:50	<b>Dati e Risorse Linguistiche: Aspetti Tecnici e Pratici</b> <ul style="list-style-type: none"><li>• Khalid Choukri</li></ul>
16:50 – 17:20	<b>Come partecipare?</b> <p>Sessione aperta ai contributi del pubblico</p> <p><i>Moderatore</i></p> <ul style="list-style-type: none"><li>• Flavia Marzano, Associazione Stati Generali dell'Innovazione</li></ul>
17:20 – 17:30	<b>Conclusioni e Piano di lavoro</b> <ul style="list-style-type: none"><li>• Khalid Choukri</li><li>• Simonetta Montemagni</li></ul>

---



## Organizzazione



Web: [www.lr-coordination.eu](http://www.lr-coordination.eu) | e-Mail: [info@lr-coordination.eu](mailto:info@lr-coordination.eu)

Coordinamento italiano



**Istituto di Linguistica Computazionale "A. Zampolli" – CNR**

Web: [lr-coordination.eu/it/Italy](http://lr-coordination.eu/it/Italy) | e-Mail: [elrc.italy@ilc.cnr.it](mailto:elrc.italy@ilc.cnr.it)