Introductory message of Khalid CHOUKRI,
ELRA Secretary General
ELDA Chief Executive Officer

Welcome to LREC 2016, the 10th edition of one of the major events in language sciences and technologies and the most visible service of the European Language Resources Association (ELRA) to the community.
Let me first express, on behalf of the ELRA/ELDA and LREC team, our profound gratitude to His Excellency Mr. Borut Pahor, President of the Republic of Slovenia, for his Distinguished Patronage of LREC2016 and for honoring us with his presence.

ELRA & LREC

Since 1998 and the first LREC in Granada, it has been a privilege to speak to the language sciences and technologies communities every two years. We always feel gifted to have the chance to share ELRA's views, concerns, expectations, and plans for the future, with over a thousand experts gathered in a lovely and relaxed atmosphere. It is also an important occasion to report on our observations of the community activities, in particular on the recent trends, that we continuously monitor. It is very rewarding to do so, in a place to be remembered.
The ELRA Board has always done its best to make this event memorable, an event with a rich scientific content, offering great networking opportunities, occasions to initiate new projects, to form new friendship with colleagues in the same field, and a lot of emotions.
Let me fist share with you a deep feeling about our organization of LREC before expressing some remarks about our recent activities and those of our community.
When discussing LREC within the ELRA Board, the first major step to be taken relates to identifying the next location. The usual and typical requirements, concerning logistics, are the first and most critical debated criteria. The motivation and the support of a local community is also a decisive factor. Remember that there is no call for bids but rather spontaneous proposals from local teams. Since its inception, LREC has been organized by the same permanent team, with the support of local committees, representing our field. Last but not least, the attractiveness of the location is a crucial element, now part of the conference's DNA. Our motto has always been that working and meeting in very relaxing conditions can improve our efficiency and productivity and enhance our interactions. We never jeopardized the scientific part when selecting our locations.

ELRA monitoring the new HLT trends

Over the last 9 LRECs and this tenth one, we have had the privilege to witness the emergence of new research fields but also the deployment of impressive applications, that, we, as a community (and also as ELRA) have proudly set-up the ground for.
ELRA was established in 1995 and the first LREC took place in Granada in 1998. Last November 2015, we celebrated our 20th anniversary. It was very exciting to discuss how far we have really come in fulfilling some of the objectives that were essential to the development of our field: many of us remember how challenging it was to develop automatic speech recognition for very basic tasks (even to recognize the 10 digits and some command words, or to discriminate between all forms of Yes and No, in a few languages, the usual "big" ones), not to mention dialogue systems and speech understanding. How challenging and critical it was

to make tiny progresses in Machine Translation systems, mostly for a few "lucrative" languages and to develop methodologies to assess such performances.

Extending these tasks with more challenging achievements is today one of the core activities of many large research teams all over the world, and one can imagine the underlying techniques composing all these applications, from processing capacities to basic tools (image and acoustic processing, morphological, syntactic, sematic analysis) and curation of appropriate Language Resources. Everything that is affordable for many players, using open source packages, cloud computing facilities, etc. Breakthroughs are impressive and many applications are now deployed and used by the public at large.
However, the bottleneck remains, as in the past, the availability of open licensable Language Resources and their sharing within the community, in particular for the less-resourced languages.
Our community is probably one of the "Big Data" communities that benefited largely from the emergence of the web. The web came as a repository of treasures of data that pushed our "data-driven" paradigms. The web brought also new challenges and potential problems related to language processing, understanding, summarization, generation, translation, etc. The web boosted new needs for tackling data and text mining, sentiment analysis, opinion detection, etc. Twitter and other social media, for instance, have now become one of the main data sources while generating scientific problems the research communities have to address. These trends are now common ground to many research topics and activities, but they also highlight the serious gaps between languages and the many barriers that are hindering the progress in various fields and geographical areas.

ELRA and the legal issues

In addition to the problems related to discovering and interoperating data sets, a critical obstacle that has limited our investigations has to do with legal and ethical aspects.
These are important issues and ELRA always worked on raising the awareness of the various stakeholders (research and industrial communities as well as policy makers). I often stressed this in my speeches at LRECs: the legal and ethical issues are major topics that require more lobbying and petitioning from our community. The research progress, in particular when it is not-for-profit and carried out in academic spheres, should not be impeded by so many obstacles, most of which should have disappeared with the emergence of internet and the digital world. The most critical issue is the copyright, along with related laws and regulations, which prevents the re-use of data for research. The idea is not to restrict the rights of authors and other intellectual property of authors and creators, but, on the contrary, to ensure a legal fair use of copyrighted works by the research community to prevent misuses that are very common, as illustrated by the statement "the web as a corpus", which seems to imply that online content is by default freely reusable.

A few countries have adopted the doctrine of fair use, giving to their research community a very "competitive" advantage. This is the case of the USA (where the Federal Government works are exempt from copyright protection). An important move from the European Commission (EC) is its challenging objective to establish a Single Digital Market across the European Union Member States. The language barriers are only one of the obstacles that the European Commission will face in establishing the SDM. To boost the innovation based on the data held by the public sector bodies, the EC has issued an important regulation, the Public

Sector Information (PSI)[1] Directive. It requires all public bodies to release the data they produce to the public so the data can be used for innovative applications. Linguistic data is part of the deal and we hope to see more resources with cleared IPR in our catalogues and in our repositories soon, for use by all. The USA and European Union are simple examples of a large international movement that hopefully will be beneficial to our community.
However, as legal aspects are always subject to interpretation, the "Territoriality and Extraterritoriality in Intellectual Property Law"[2] does not help to clear these issues. It is crucial that an international harmonization addresses this (maybe a serious amendment of existing conventions). Our community should contribute to the initiatives going on in various countries about the copyright and other related rights in the information society[3], to boost data sharing and re-use.
ELRA strongly supports the Open Data movement and advocates for making public sector information more accessible and re-usable, without any license or through very permissive and open licenses.
Another topic debated at the 20th anniversary of ELRA, is the ethical issues that have to be considered in our field, either when dealing with Data Management (for instance the crowdsourcing approach to data production) or when replicating experiments and citing publications of other colleagues.
I am very happy to see that the dedicated workshop on "Legal Issues" (http://www.elra.info/en/dissemination/elra-events/legal-issues-workshop-lrec2016/) is taking place at LREC, once more, with a large number of registered participants. Another workshop on ethical issues, is scheduled within this LREC (ETHics In Corpus Collection, Annotation and Application, http://emotion-research.net/sigs/ethics-sig/ethi-ca2).

Replicating experiments and Data Citation

This crucial topic of replicating experiments covers a large spectrum of behaviors and was reviewed by the experts group, gathered to discuss the activities of our field at the twentieth anniversary of ELRA. The topic refers to "Reproducibility of the Research Results and Resources Citation in Science and Technology of Language". The topic has emerged as a hot topic for discussion within many fields of research activities. With António Branco and Nicoletta Calzolari, we made a proposal for a discussion workshop to debate this topic at LREC. I am very grateful to António Branco who agreed to take the lead on this sensitive topic (4REAL Workshop: http://4real.di.fc.ul.pt/). Like in many scientific fields, several dimensions are important and require specific considerations. Of course, maintaining research integrity is essential and requires that replication of published results is possible, and even guaranteed. Such replication can only be done through sharing of resources and approaches. The same requirements are needed to compare results across approaches which, in addition, require that one clearly identifies the resources used in the benchmarking. The identification of the resources, and thus, their citations, are correlated and more attention is required than what

[1] https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information
[2] Alexander Peukert, "Territoriality and extra-territoriality in intellectual property law", in Günther Handl, Joachim Zekoll & Peer Zumbansen (eds.), Beyond Territoriality: Transnational Legal Authority in an Age of Globalization, Queen Mary Studies in International Law, Brill Academic Publishing, Leiden/Boston, 2012, 189-228.
[3] Study on the application of Directive 2001/29/EC on copyright and related rights in the information society (the "Infosoc Directive"), Jean-Paul Triaille, Séverine Dusollier, Sari Depreeuw, Jean-Paul TRIAILLE (ed.), ISBN: 978-92-79-29918-6, DOI: 10.2780/90141, © European Union, 2013.

we did so far. The use of the ISLRN (International Standard Language Resource Number, http://www.islrn.org/) has been introduced and is being supported by the major data centers. We hope that monitoring this process will show how impactful it is on our field. Citation mechanisms, applied to Language Resources, affect the so called "impact factor" and the underlying research indexes, and have to be taken care of by the community.

ELRA & Data Management Plan

Another major topic that was briefly discussed at the ELRA 20th anniversary is Data Management Plans and the underlying Language Resource Sustainability factors (DMP&S). Within its activities and through a number of projects, ELRA has always advocated for a comprehensive data management strategy that would ensure efficient management of the production, repurposing or repackaging processes, a clear and adequate validation process, sharing and distribution mechanisms, and a sustainability plan[4], (http://www.flarenet.eu/sites/default/files/D2.2a.pdf). It is now common practice by most of the funding agencies to request a Data Management Plan to be part of the proposals they receive. We hope that such plans are seriously designed and not only as administrative sections of the proposals. ELRA will continue to work on this challenging task for the emerging resource types and contexts of use and will continue to offer its support (including through its helpdesk on technical and legal matters) to the proposers and project managers. ELRA is working on a very specific tool to help Language Resource managers produce their own DMP. ELRA is working on a specific wizard, based on its background and input from other projects (the MIT Libraries' Research Data Management Plan[5], and the initiative held by the Inter-university Consortium for Political and Social Research[6]).
To assess how this "sustainability" dimension is taken into consideration, ELRA has established a set of resources that have been monitored, by our Language Resource experts, since 2010. We started the list with resources that were not part of the catalogues of data centers. One would be surprised to see **that over 28% have simply disappeared from the radars**, with many others requiring a thorough search before one can find them again. And it is always complicated to identify them accurately and verify they correspond to the ones on our list. Again this emphasizes the need for repositories that adhere to a clear code of conduct. The simple "web" (and URLs) will not constitute a reliable and persistent repository.

LREC 2016, some features

As usual, LREC 2016 features a large number of workshops. We are proud to continue to support the Sign language workshop that is building bridges between several research communities and setting partnerships for research on so many modalities (http://www.sign-lang.uni-hamburg.de/lrec2016/workshops.html). We are happy to see that the community is very active in paying attention to the less-resourced and under-resourced languages (http://www.ilc.cnr.it/ccurl2016/index.htm), a topic that is now taken care of by a dedicated ELRA committee, the LRL committee. Many other specific topics are covered in this edition's

[4] Khalid Choukri and Victoria Arranz, "An Analytical Model of Language Resource Sustainability", Proceedings of LREC2012.
[5] https://libraries.mit.edu/data-management/plan/write/
[6] http://www.icpsr.umich.edu/

workshops: Arabic and Indian languages, social media, emotions, affect analysis, multimodal resources, and several workshops dealing with MT and MT-related aspects.
LREC 2016 is also featuring a panel discussion with some of the major funding agencies. We hope to draw some conclusions about their past activities but also to discuss roadmaps for the next decade.
Last but not least, the importance of our technologies was emphasized when a strong earthquake hit Nepal in April 2015. Many teams offered their technologies to help the rescue groups, in particular the multilingual applications. Some teams designed and quickly developed new applications for this. ELRA donated Nepalese resources for this purpose. We are thankful to our partners who agreed to waive all fees on these resources and very grateful to those who developed the applications that may have helped a little bit.

Acknowledgments:
Finally, I would like to express my deep thanks to our partners and supporters, who throughout the years make LREC so successful. I would like to thank our Bronze sponsors: EML (European Media Laboratory GmbH) and Intel, our supporter: Viseo, and our media sponsor: MultiLingual Computing, Inc.

I also would like to thank the HLT Village participants, we hope that such gathering offers the projects an opportunity to foster their dissemination and hopefully to discuss exploitation plans with the participants.

I would like to thank the Local Advisory Committee. Its composition of the most distinguished personalities of Slovenia denotes the importance of language and language technologies for the country. We do hope that it is a strong sign for the long-term commitment of Slovenian officials.

I would like to thank the LREC Local Committee, chaired by Dr Simon Krek and the LREC Local Organizing Committee, chaired by Marko Grobelnik, in particular Špela Sitar *and* Monika Kropej *for providing support to the organization of this LREC Edition in Slovenia.*

Finally I would like to warmly thank the joint team of the two institutions that devoted so much effort over months and often behind curtains to make this one week memorable: ILC-CNR in Pisa and my own team, ELDA, in Paris. These are the two LREC coordinators and pillars: Sara Goggi and Hélène Mazo, and the team: Roberto Bartolini, Irene De Felice, Meritxell Fernández-Barrera, Riccardo Del Gratta, Francesca Frontini, Lin Liu, Valérie Mapelli, Monica Monachini, Vincenzo Parrinelli, Vladimir Popescu, , Caroline Rannaud, Irene Russo, Priscille Schneller and Alexandre Sicard.

Now LREC 2016 is yours; we hope that each of you will achieve valuable results and accomplishments. We, ELRA and ILC-CNR staff, are at your disposal to help you get the best out of it.

Once again, welcome to Portorož and Slovenia, welcome to LREC 2016