



ANNUAL ACTIVITY REPORT 2014

ELRA • 9, rue des Cordelières, 75013 Paris • France
Telephone: +33 1 43 13 33 33 • Fax: +33 1 43 13 33 30
info@elda.org • www.elra.info • www.elda.org

CONTENT

Content

Word by the President & Secretary General of ELRA	1
The ELRA Board	2
2014 at a glance	3
ELRA Members	4
ELRA Activities and Projects	6
LREC, the Language Resources and Evaluation Conference	14
Join ELRA	16

ELRA ANNUAL ACTIVITY REPORT 2014

Word by the President & Secretary General of ELRA

The Board has been drawing a new strategy for ELRA for the next 5 years in order to consolidate ELRA as a key player in the HLT community, focusing on anticipating the community's new expectations as well as strengthening its relationship with the European Commission. New trends are perceived through many initiatives popping up within the community, especially within the framework of the EC-funded R&D programs (FP7, CIP, Horizon2020, CEF, etc.). In line with this strategic orientation, several actions are being put into operation:

- The establishment of the **International Standard Language Resource Number (ISLRN)** (<http://www.islrn.org>), a unique identifier to be assigned to each LR allowing a unique identification of a resource, regardless from its physical storage, its availability or the associated license.
- The organisation of the **NLP12 meeting**, with the 12 major NLP associations and institutions from Language Resources and Technologies, Computational Linguistics, Spoken Language Processing and Digital Humanities aiming at boosting their co-operation and strengthening the bridges between various communities.
- The exploitation of the **META-SHARE** platform in order to become ELRA's next instrument of sharing LRs, including the development of new extensions such as a **legal Wizard** to help the right-holders share and/or distribute their resources under the right license.
- The celebration of the association's **18th anniversary** in November 2013, in Paris, through the organisation of the workshop *ELRA International Workshop on Sharing Language Resources: Landscape and Trends*.

“We continue to vision 2013-2015 with emphasis on a renovation of ELRA policy and activities. ELRA continues to discuss potential extension of its membership to new types of members in particular to attract individuals so it can serve them better.”

Nicoletta Calzolari, President






Khalid Choukri, Secretary General



ELRA ANNUAL ACTIVITY REPORT 2014

The ELRA Board

BOARD OFFICERS

NICOLETTA CALZOLARI PRESIDENT	KHALID CHOUKRI SECRETARY GENERAL	JOSEPH MARIANI HONORARY PRESIDENT
		
ILC CNR, Italy	ELDA, France	LIMSI & IMMI, France

NICK CAMPBELL VICE-PRESIDENT	ISABEL TRANCOSO VICE-PRESIDENT	MARIA GAVRILIDOU SECRETARY	ANDREJS VASILJEVS TREASURER
			
TCD, Ireland	INESC, Portugal	ILSP, Greece	Tilde, Latvia

BOARD MEMBERS

THIERRY DECLERCK	MARKO GROBELNIK	HENK VAN DEN HEUVEL	FREDERIQUE SEGOND
			
DFKI, Germany	JSI, Slovenia	CLST, The Netherlands	Objet Direct, France

ELRA ANNUAL ACTIVITY REPORT 2014

2014 at a glance

Jan	<ul style="list-style-type: none">The LREC 2014 Programme Committee meeting takes place in Reykjavik (Iceland).
Feb	<ul style="list-style-type: none">ELRA and ELDA move to new offices. The new address is: 9, rue des Cordelières, ParisFinal Submission and Online registration to LREC 2014 opens.
Mar	<ul style="list-style-type: none">As a KHRESMOI project partner, ELDA attends the CeBIT2014 in Germany (March 10-14) and the European Data Forum in Greece (March 19-20).
Apr	<ul style="list-style-type: none">ELRA launches the ISLRN Portal. Both LDC and ELRA allocate ISLRN numbers.
May	<ul style="list-style-type: none">LREC 2014 takes place in Reykjavik (May 26-31), under the patronage of UNESCO and with the support of Madame Vigdís Finnbogadóttir, former President of Iceland.
Jun	<ul style="list-style-type: none">Khalid Choukri chairs the ISO/TC/SC35 Big Data and Human Interfaces Meeting in Paris (June 5-6, 2014)
Jul	<ul style="list-style-type: none">ELRA sponsors TALN 2014 in July 2014ISO/SC35 Plenary Meeting in Hang Zhou, China (July 7-10)
Sep	<ul style="list-style-type: none">ELRA sponsors the 15th edition of INTERSPEECH in Singapour (September 14-18) and runs a booth during the conference.
Oct	<ul style="list-style-type: none">ELDA participates in the KHRESMOI Project Final Review in Geneva, Switzerland (October 16).ELRA General Assembly is held in Paris (October 21).ELRA participates in the UNESCO International Expert Meeting on Improving Access to Multilingual Cyberspace, Paris (October 28-29).
Nov	<ul style="list-style-type: none">First visit to LREC 2016 expected site: Portoroz in Slovenia (November 10-11)ISO/IEC/JTC1 Plenary Meeting on November 17-20, 2014 in Abu Dhabi (United Arab Emirates)ELRA participates in the ATALA Workshop on Ethics and NLP (November 22)ELRA participates in the International conference on ITC for Amazigh in Rabat, Morocco (November 24-25)ELRA attends the meeting on languages of France at DGLFLF in Paris (November 24)
Dec	<ul style="list-style-type: none">ELRA announces the next LREC edition: LREC 2016 will be held in Portoroz in Slovenia, on May 23-28, 2016.JRC, the EC's Joint Research Centre, an important LR player, is the first to adopt the ISLRN initiative.Cracker, a new CSA project under H2020 is accepted for funding by European Union.First review of MLi in Luxemburg

ELRA ANNUAL ACTIVITY REPORT 2014

ELRA Members

ELRA membership scheme is **institution-based**. Any organisation, public or private, European or non-European, can join. However, full membership, with voting rights, is available only to organisations legally established in Europe, as per the article 5 of the Statutes.

2014 MEMBERSHIP AT A GLANCE

Who are our members?

48 total number of members

13 new members joined in 2012

Members having renewed their membership **35**

6 members joined in 1995 and have renewed membership continuously ever since

European members **81%**

56%

European non profit organisations



Where are our members?



ELRA ANNUAL ACTIVITY REPORT 2014

SERVICES TO MEMBERS

- **Price discounts** on the resources, up to 70% reduction on some items, including the resources produced by ELRA.
- **Discounted fees** for members registering to the biennial conference LREC.
- **Legal and contractual assistance** are also provided to the members of the association
- **Regular information** through the monthly Members' News bulletin and the updates brought to the www.elra.info web site.
- Free online access to **Language Resources and Evaluation (LRE) Journal**, published by Springer, is provided to all ELRA members.

Legal Support HelpDesk

Using, producing, sharing or distributing Language Resources can trigger legal questions related to Intellectual Property Rights (IPR) management which are not always easy to sort out. For nearly 20 years now, ELRA has established close co-operation with legal experts to clear such IPR issues, to design licensing schemas and draft licenses, but also to provide assistance on any contractual or legal matter that may arise during the Language Resources life cycle, including acquisition, production, sharing, or distribution phases.

With this Helpdesk service fully dedicated to IPR issues, we are extending our legal support to the whole Human Language Community. This Helpdesk provides services similar to those offered by the META-SHARE network.

Please contact us by sending a message to elra-ipr-helpdesk@elra.info.

Support hlt-related events

A number of events have been sponsored by ELRA in order to support and endorse their activities within our field. In 2014, these are:

- European Data Forum 2014
- TALN 2014
- INTERSPEECH 2014
- CLiC-it 2014

In addition, ELRA continues to support the Linguist List.

Our members are welcome to contact us if they seek such support for events (conferences, workshops) that focus on HLT and preferably on LRs and Evaluation.

LRE Friends

The Forum of Language Resources and Evaluation Friends (LRE-F) was established at the last LREC in Istanbul. The LRE Friends are individuals interested in Language Resources and Evaluation. They can be scientists, students or professors, involved in HLT research activities in universities, small and medium companies or international groups, decision-makers or project managers in large public institutions, etc.

ELRA ANNUAL ACTIVITY REPORT 2014

ELRA Activities and Projects

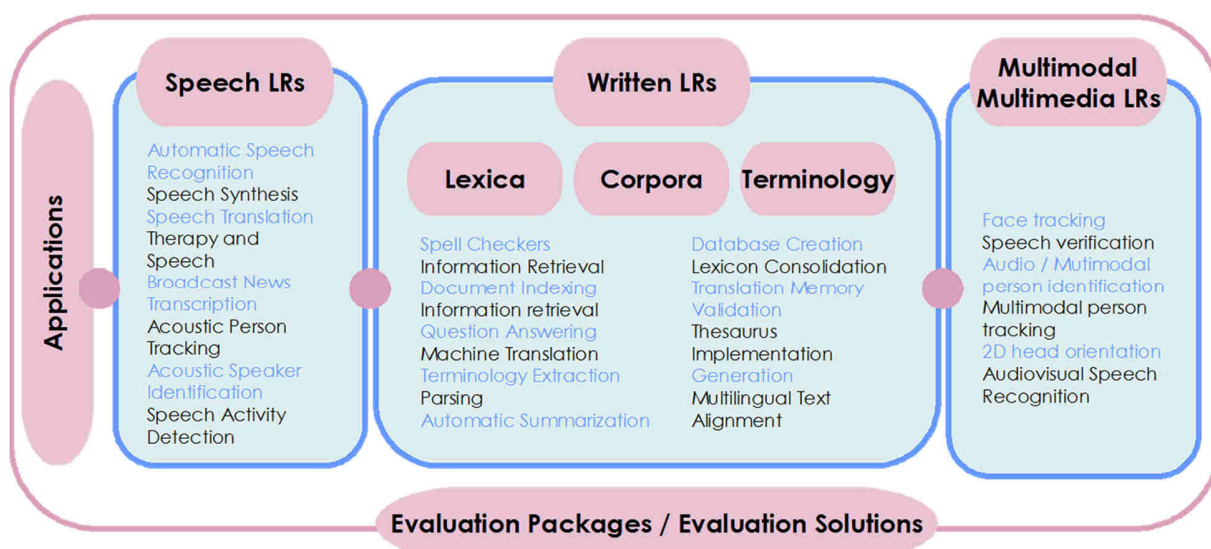
In addition to the promotion of language resources for the Human Language Technology (HLT) sector through market surveys, publication of a regular newsletter and organisation of LREC conference and other workshops, the activities of ELRA are organised around the following services:

- Identification, collection and distribution of language resources
- Production of language resources
- Validation of language resources
- Evaluation of systems, products, tools, etc., related to language resources
- Standardisation

The promotion of the language resources production also includes our support of both the infrastructure for evaluation campaigns and the development of a scientific field of language resources and evaluation, e.g. via the LREC conference.

LANGUAGE RESOURCES CATALOGUE

The Language Resources collected by ELRA are made available to the public through the ELRA Catalogue that can be accessed online at <http://catalog.elra.info>*. The distribution of resources, which cover a wide variety of languages and belong to different modalities, is shown below:

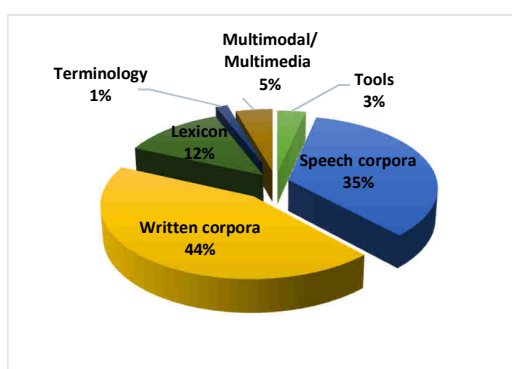


* Now also part of the META-SHARE network at www.meta-share.net

ELRA ANNUAL ACTIVITY REPORT 2014

The term **Language Resources** refers to sets of language data and descriptions in machine readable form, used in many types of areas, components, systems, applications, including the creation and evaluation of natural language, speech and multimodal algorithms and systems, the software localisation and language services, the language-enabled information and communication services or the knowledge management; e-commerce, e-publishing, e-learning, e-government.

Universal Catalogue



The Universal Catalogue is an important identification feature. Information regarding Language Resources (LRs) identified all over the world are gathered in this publicly available repository. The LRs are generally located by the ELRA team, but external feedback from our members, collaborators, and web-site visitors is also included. Our aim is to provide researchers and developers information about existing LRs and spare them the effort of searching or rebuilding similar resources. The Universal Catalogue is a collaborative feature: through a simple form, anyone interested can

add some basic information, point to an existing language resource or enrich the current description of a LR already present in the Universal Catalogue.

The Universal Catalogue is accessible through the following web site: <http://www.universal.elra.info>.

LRE Map

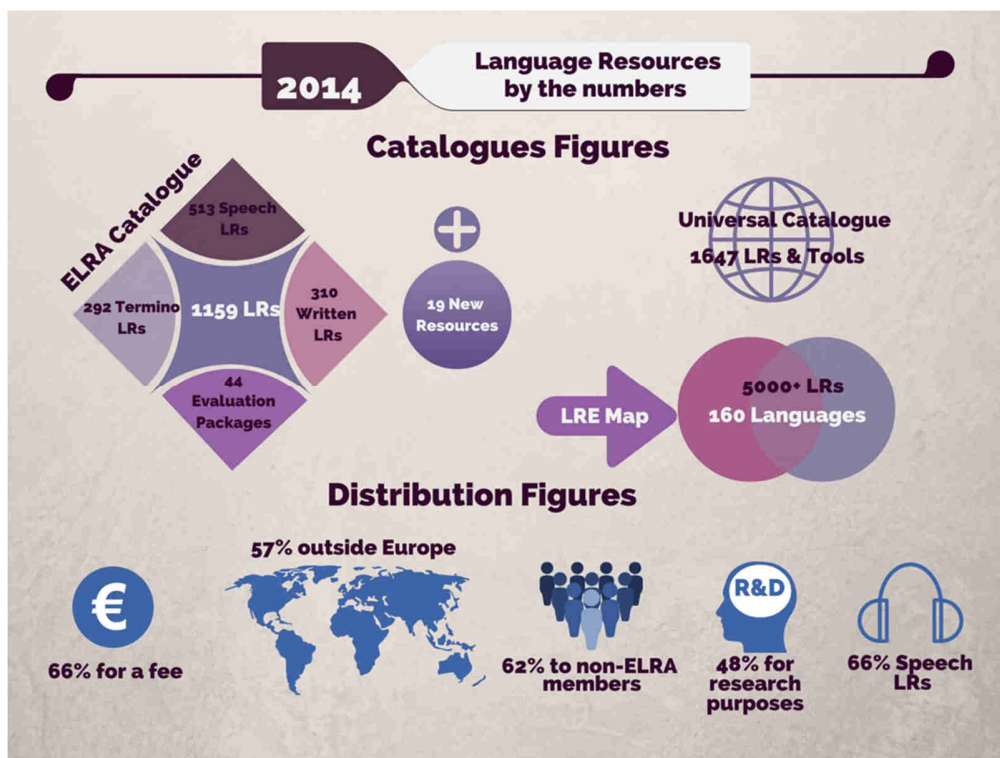
The initiative for LR & Tools identification, the LRE Map

Initiated by ELRA and FlareNet at LREC 2010, the LRE Map is a new mechanism intended to monitor the use and creation of language resources by collecting information on both existing and newly-created resources during the submission process. Nearly 2000 language resource forms have been filled in. Apart from providing a portrait of the resources used by the community, of their uses and usability, the LRE Map intends to be a measuring instrument for monitoring the field of language resources. The feature has been so successful that it has been implemented also at COLING 2010 and EMNLP 2010, while other major conferences have adopted it since then.

The LRE Map is now part of the LREC submission process which has been slightly updated for the LREC 2012 submission to consider the changes brought to the 2012 version of the Map. Further to the very high number of submissions to LREC 2012 (1027 for both oral and poster papers), the number of LR forms have increased substantially. Globally, 5000+ LR forms have been filled in within all the conferences which have adopted the LRE Map.

ELRA ANNUAL ACTIVITY REPORT 2014

LANGUAGE RESOURCES: CATALOGUES & DISTRIBUTION

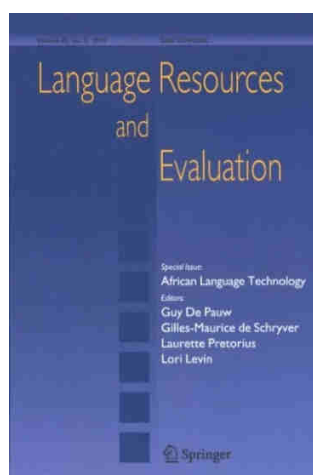


New LRs in 2014 & Distribution

Category	LR ID	Description
Speech LRs	ELRA-S0366	LECTRA (LECTure TRANscriptions in European Portuguese)
	ELRA-S0367	CORAL Corpus
	ELRA-S0368	Nepali Spoken Corpus
	ELRA-S0369	CLIPS_MT_MANUAL
	ELRA-S0370	MoveOn Speech and Noise Corpus
	ELRA-S0371	PortMedia French and Italian corpus
	ELRA-S0372	GlobalPhone Thai Pronunciation Dictionary
	Written LRs	ELRA-W0076
ELRA-W0077		English-Nepali Parallel Corpus
ELRA-W0078		NE3L named entities Arabic corpus
ELRA-W0079		NE3L named entities Chinese corpus
ELRA-W0080		NE3L named entities Russian corpus
Monolingual lexica		
ELRA-L0089		Macedonian lexicon of toponyms (MACPLEX_TOPO)
ELRA-L0090		Macedonian lexicon of proper nouns (MACPLEX_PROPER)
ELRA-L0091		Macedonian lexicon of derived adjectives (MACPLEX_ADJDERV)
ELRA-L0092		Macedonian lexicon of participles (MACPLEX_ADJPARTIC)
ELRA-L0093	Macedonian lexicon of compound words (MACPLEX_COMP)	
Terminological resources:		
ELRA-T0375	ACL RD-TEC: A Reference Dataset for Terminology Extraction and Classification Research in Computational Linguistics	
Evaluation	ELRA-E0042	CLEFeHealth 2013 Evaluation Package

ELRA ANNUAL ACTIVITY REPORT 2014

LANGUAGE RESOURCES AND EVALUATION JOURNAL



The LRE Journal, published by Springer, is the first publication devoted to the acquisition, creation, annotation, and use of LRs, together with methods for evaluation of resources, technologies, and applications. The ELRA members are granted complimentary access to the journal through the society on the condition that they subscribe to the publisher's table-of-contents alert service.

In 2014, the following regular issues, edited by Nicoletta Calzolari and Nancy Ide, have been published:

- Volume 48, n° 1, Special Issue on **Resources for language learning** guest-edited by S. Sharoff, S. Spina and S. Johansson-Kokkinakis
- Volume 48, n° 2, Summer 2014
- Volume 48, n°3, Fall 2014
- Volume 48, n° 4, 2014: a Special LREC 2012 Issue

For LREC editions, the LREC organisers and the JLRE Editors have agreed with Springer on some special conditions for subscription to the JLRE to be offered to LREC participants.

SOME KEY PROJECTS



The **MLi Support Action** is working to deliver the strategic vision and operational specifications needed for building a comprehensive European MultiLingual data & services Infrastructure, along with a multi-annual plan for its development and deployment, and foster multi-stakeholders alliances ensuring its long term sustainability. More information can be found on the following website: <http://mli-project.eu>



Focusing on Medical Information analysis and retrieval, the project KHRESMOI which stands for *Knowledge Helper for Medical and Other Information users* addresses both challenges of trustworthiness and complexity levels in online health information. The project is funded under the EU 7th Framework Programme ICT theme. The consortium, consisting of 12 partners from 9 European countries, will develop a web-based multi-lingual multi-modal search and access system for biomedical information and documents over the next four years. The system will allow querying in several languages, in combination with visual image queries. It will return translated document summaries linked to the original documents. This project continues under internal funds.

ELRA ANNUAL ACTIVITY REPORT 2014



META-SHARE is one of the objectives within META-NET, it aims to be “a sustainable network of repositories of language data, tools and related web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources.”

(<http://www.meta-net.eu/meta-share>). In its aim to achieve such a network, work is being carried out for the specification of a metadata schema which builds upon currently available schema, knowledge and expertise and provides a unified schema capable of handling the requirements of the community. These requirements comprise both the description of Language Resources and that of tools or technologies.

To allow an easy understanding of the META-SHARE licences, ELRA is developing a legal Wizard to help the right-holders share/distribute their resources under the right licence. This will be also usable by users to better understand their rights and duties.

FLaReNet Follow-up



FLaReNet is a project funded by the European Commission under the eContentplus programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable. The project is now finished but ELRA agreed to take over the management of the national contact points. In addition, ELRA agreed to work on the idea to host and maintain a repository of best practices and standards.

ELRA ANNUAL ACTIVITY REPORT 2014

ISLRN, THE INTERNATIONAL STANDARD LANGUAGE RESOURCE NUMBER

<http://www.islrn.org>

Every object in the world requires a kind of identification in order to be correctly recognised. Traditional printed material like books, for example, have generally used the International Standard Book Number (ISBN), the Library of Congress Control Number (LCCN), the Digital Object Identifier (DOI) and several other numeric identifiers as a unique identification scheme. Book identifiers allow us to easily identify books in a unique way. Other domains make use of several other identifier schemes. For instance, it is not rare to run into an International/European Article Number (EAN), which is a universal barcoding system for everyday products. Each of these schemes seems to have been the output of some specific need or circumstance within a domain.

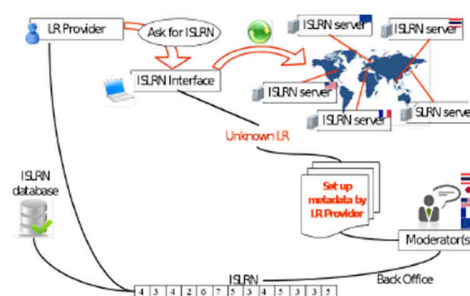
After having reviewed existing identification schemas, ELRA and a large number of HLT key-players (NLP12) have concluded that there is a need to establish a specific identifier for Language Resources (LRs). The **International Standard Language Resource Number (ISLRN)** is a universal identification schema that provides a unique identifier using a standardized nomenclature. Each Language Resource with an ISLRN can therefore be recognized with its proper reference whether it is used in applications in R&D projects or in product evaluation and benchmark, or cited in documents and scientific publications.

The ISLRN does not replace local and specific identifier, it is not a legal deposit nor an obligation, rather a best practice: a Language Resource distributed by several data centers with a “local number” or published in several meta-catalogues can still be allocated a unique ISLRN.

ISLRN ATTRIBUTION PROCESS

ISLRN PROCESS

The current ISLRN attribution workflow is pictured below. The moderation is handled by ELRA.



A 13-DIGIT NUMBER

ISLRN: xxx-xxx-xxx-xxx-x
 Random number Checksum

Since the LR identifier delegates semantics of its content to metadata descriptions, it was decided to use a 12-digit random number as the new LR identifier, followed by 1 digit for a checksum number. The checksum number is encrypted from the preceding numeric identifier and version information.

ELRA ANNUAL ACTIVITY REPORT 2014

HOW DOES IT WORK?

Two types of external players may interact with the ISLRN Portal: Visitors and Providers.

- ➔ Visitors may browse the web site and search for the ISLRN of a given language resource by its name or by its number if it exists. They can also sign up to become Providers.
- ➔ Providers are registered and own credentials. They can request an ISLRN for a given language resource. A provider has the possibility to become certified, after moderation, in order to be able to import metadata in XML format.

You are a Visitor, you can:

- Identify a language resource according to its ISLRN
- Identify an ISLRN by the name of a language resource
- Get information about ISLRN, FAQ, Basic Metadata, Legal Information
- View last 5 accepted resources (“What’s new” block on main page)
- Sign up to become a provider

You are a Provider of Language Resources, you can:

- Log in
- Request an ISLRN according to the metadata of a given resource
- Request to become a certified provider so as to import XML files containing metadata
- Import one or more metadata descriptions in XML to request ISLRN(s) (only for certified providers)
- Edit pending requests
- Access previous requests
- Contact a Moderator or an Administrator
- Edit Providers’ own profile

Example of a resource identified under ISLRN

Reference	NetDC Arabic BNSC (Broadcast News Speech Corpus)
Date of Submission	Jan. 24, 2014, 4:30 p.m.
Status	accepted
ISLRN	663-177-513-755-1
Resource Type	Primary Text
Media Type	Audio
Source	http://catalog.elra.info/product_info.php?products_id=13
Language	Arabic
Description	The NetDC Arabic BNSC (Broadcast News Speech Corpus) is a corpus developed by ELDA in the framework of the European-funded project Network of Data Centres (NetDC). The project was done in collaboration with the LDC (Linguistic Data Consortium) [...] A phonetic lexicon in Arabic SAMPA has also been included.
Version	1.0
Distributor	ELRA

ELRA ANNUAL ACTIVITY REPORT 2014

INSTITUTIONS HAVING REQUESTED ISLRNS

Both ELRA and LDC have completed the allocation of ISLRNs to each of the Language Resources in their respective catalogues. Language Resources common to both catalogues have been allocated a single ISLRN.

The Joint Research Centre (JRC), the European Commission's in house science service, was the first organization to request the attribution of ISLRN to its Language Resources.

In 2014, more than 20 providers have submitted ISLRN allocation requests for their Language Resources. In total, **2047 requests** have been processed and **1885 ISLRN** numbers have been allocated (including LRs from ELRA and LDC).

STEERING COMMITTEE

To guarantee its sustainability, free availability, and ethical modus operandi, the ISLRN attribution is trusted to a small group of organizations involved in LRs distribution and sharing issues:

- European Language Resources Association (ELRA)
- Linguistic Data Consortium (LDC)
- Asian Federation of Natural Language Processing (AFNLP)
- International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques (Cocosda)



Reach the ISLRN web site and request ISLRN for your own Language Resources:

<http://www.islrn.org>

ELRA ANNUAL ACTIVITY REPORT 2014

LREC, the Language Resources and Evaluation Conference

The Language Resources and Evaluation Conference is an international scientific event which aims at providing an overview of the state of the art, exploring new R&D directions and emerging trends, exchanging information regarding Language Resources and their applications, carrying out the evaluation of methodologies and tools, on-going and planned activities, industrial uses and needs, requirements coming from the e-society, with respect to both policy issues and technological and organisational ones.

LREC provides a unique forum for researchers, industrial players and funding agencies from across a wide spectrum of areas to discuss problems and opportunities, find new synergies and promote initiatives for international cooperation in support of investigations in language sciences, progress in language technologies and development of corresponding products, services, applications, and standards.

In 16 years LREC which is organized every other year has become the major event on Language Resources and Evaluation for Human Language Technologies. The conference programme is organised around parallel oral and poster sessions during the main conference (3 days) and workshops and tutorials (2 days before and 1 day after the main conference). Since 1998, the LREC conference has been organized every other year with an increasing success.

NEXT EDITION
LREC 2016 IN PORTOROŽ
23-28 MAY 2016
<http://www.lrec-conf.org/lrec2016>



PAST LREC EDITIONS
<http://www.lrec-conf.org>



LREC 2014 Reykjavik
1200+ participants



LREC 2012 Istanbul
1200+ participants



LREC 2010 Malta
1200+ participants



LREC 2008 Marrakech
1100 participants



LREC 2006 Genoa
800 participants



LREC 2004 Lisbon
950 participants



LREC 2002 Las Palmas
730 participants



LREC 2000 Athens
600 participants



LREC 1998 Granada
510 participants

ELRA ANNUAL ACTIVITY REPORT 2014

LREC 2016, 10TH EDITION



MAIN CONFERENCE: 25-26-27 MAY 2016 - **WORKSHOPS and TUTORIALS:** 23-24-28 MAY 2016

<http://www.lrec-conf.org/lrec2016/lrec2016.htm>



@LREC2016

The format of the 10th edition of LREC is the one adopted at the Reykjavik edition: two days for the pre-conference workshops and tutorials and on day for the post-conference.

The format of the Main Conference remains unchanged (on 3 days)

PROGRAMME COMMITTEE

- **Nicoletta Calzolari**, *Conference chair*, CNR, ILC "Antonio Zampolli", Pisa (Italy)
- **Khalid Choukri**, ELRA, Paris (France)
- **Thierry Declerck**, DFKI GmbH, Saarbrücken (Germany)
- **Marko Grobelnik**, Jožef Stefan Institute (JSI) (Slovenia)
- **Bente Maegaard**, CST, University of Copenhagen (Denmark)
- **Joseph Mariani**, LIMSI-CNRS & IMMI, Orsay (France)
- **Asuncion Moreno**, Universitat Politècnica de Catalunya, Barcelona (Spain)
- **Jan Odijk** – UIL-OTS, Utrecht (The Netherlands)
- **Stelios Piperidis**, Athena Research Center/ILSP, Athens (Greece)

PROGRAMME

The Scientific Programme will include invited talks, oral presentations, poster and demo presentations, and panels, in addition to a keynote address by the winner of the Antonio Zampolli Prize.

ELRA ANNUAL ACTIVITY REPORT 2014

Join ELRA

The annual membership fees are shown in the Fees column.

A Fidelity Programme has been implemented to reward the faithful members. The Miles column displays the number of miles (1 mile=1€) which is allocated to each member joining the association and/or remaining an ELRA member. Miles can be used to purchase resources, pay the membership fees or the registration to LREC.

	Fees (€)	Miles
Non-profit making organisations	750	200
European small/medium-sized companies (< 50 employees)	1000	250
European profit making organisations (>= 50 employees)	1500	375
Non-European profit making organisations	5000	1250

<http://www.elra.info/en/join-elra>



ELRA • 9, rue des Cordelières, 75013 Paris • France
Telephone: +33 1 43 13 33 33 • Fax: +33 1 43 13 33 30
info@elda.org • www.elra.info • www.elda.org