

# ELRA Activities and Services

Khalid Choukri, Valérie Mapelli, H el ene Mazo, Vladimir Popescu  
ELDA/ELRA

9 rue des Cordeli eres, 75013 Paris, France  
Email: {choukri; mapelli; mazo; popescu}@elda.org

## Abstract

After celebrating its 20th anniversary in 2015, ELRA is carrying on its strong involvement in the HLT field. To share ELRA’s expertise of those 21 past years, this article begins with a presentation of ELRA’s strategic Data and LR Management Plan for a wide use by the language communities. Then, we further report on ELRA’s activities and services provided since LREC 2014. When looking at the cataloguing and licensing activities, we can see that ELRA has been active at making the Meta-Share repository move toward new developments steps, supporting Europe to obtain accurate LRs within the Connecting Europe Facility programme, promoting the use of LR citation, creating the ELRA License Wizard web portal. The article further elaborates on the recent LR production activities of various written, speech and video resources, commissioned by public and private customers. In parallel, ELDA has also worked on several EU-funded projects centred on strategic issues related to the European Digital Single Market. The last part gives an overview of the latest dissemination activities, with a special focus on the celebration of its 20<sup>th</sup> anniversary organised in Dubrovnik (Croatia) and the following up of LREC, as well as the launching of the new ELRA portal.

**Keywords:** Data Management Plan, Cataloguing, Licensing, LR Citation, LR Production, Events

## 1. Introduction

In 2015, ELRA celebrated its 20th anniversary. This was the right moment to take a look back and consider ELRA’s achievements. From these achievements, we could gather a number of skills that merit to be shared. Thus, this article begins with a presentation of ELRA’s strategic Data and Language Resource (LR) Management Plan for a wide use by the language communities.

Always following the current trends on adapted ways of cataloguing and licensing LRs, ELRA has been active at making the Meta-Share repository move toward new developments steps, supporting Europe to obtain accurate LRs within the Connecting Europe Facility (CEF) programme, promoting the use of LR citation, creating the ELRA License Wizard web portal dedicated to legal issues. These works are presented in the following sections.

During the last two years, we have also produced (or participated in the production of) written, speech and video resources, commissioned by public and private customers. In parallel, ELRA (through the ELDA agency) has also worked on several EU-funded projects centred on Machine Translation in eCommerce, and on maintaining the Meta-Share repository.

The last part gives an overview of the latest events and dissemination activities, with a special focus on the celebration of its 20th anniversary organised in Dubrovnik (Croatia) and the following up of LREC. More is also said on the developments that were carried out to finally launch a newly designed ELRA portal.

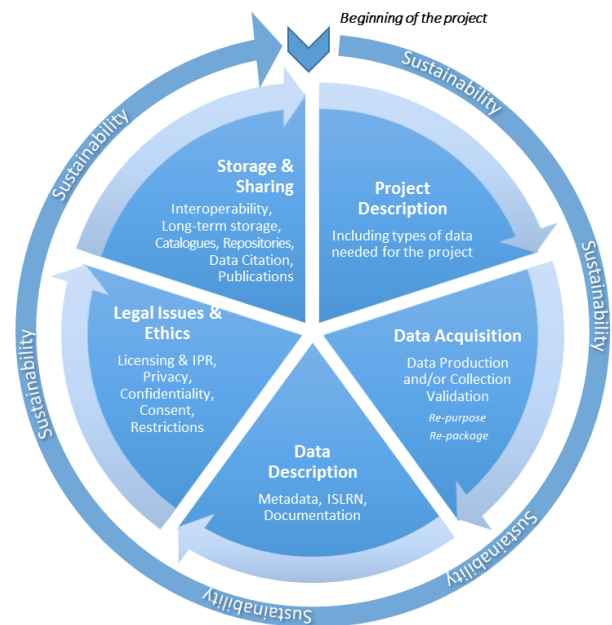
## 2. Data Management Plan

Data Management is one of ELRA’s key activities which encompasses LRs discovery and integration to its LRs Catalogue, LRs production, and IPR issues clearance prior to distributing the LRs. Dissemination on LRs and their uses, including the evaluation of HLT systems, is also part of ELRA’s missions.

Through the Data Management Plan (DMP), ELRA outlines how data should be handled during a research project and beyond. The DMP covers the entire data

lifecycle (see Graph 1) and describes the policy that should be implemented to manage data efficiently throughout the project and ensure data is sustainable. In order to refine the sustainability assessment, ELRA is using the 20 impacting factors and means identified by FlaReNet<sup>1</sup> and whose combination lead to a global sustainability.

All phases are documented: project description, data acquisition, data documentation and metadata, data licensing, data curation and sharing.



Graph 1: Data Management Plan lifecycle

### 2.1. Project description

The nature of the project must be described, as well as the institutions involved. Any information that can provide useful context information should be added. The nature of the data should be specifically described as well as their use within the project, and reference to documented existing data policies should be included.

<sup>1</sup> <http://www.flarenet.eu>

## 2.2. Data acquisition

The DMP covers all data required by the project. They can be new data to be produced within the project and/or existing data to be collected and can be used/re-used, repurposed, repackaged. ELRA has a long-time experience in Data Production. The LRs production process described below complies with the sustainability rules adopted within FlaReNet and includes the following phases:

- Definition of specifications: description of data, use of standards and best practices, metadata, planning and costs, quality assessment, validation procedures, IPR issues, management of ethics, privacy, confidentiality, packaging, the promotion in case of distribution and sharing.
- The production phase: actual production of the data, quality controls on a regular basis, validation.
- The post-production phase: allocation of a unique identifier (ISLRN), attribution of the appropriate license, documentation, selection of the appropriate storage, bug reporting, maintenance, dissemination.

## 2.3. Data description

ELRA has often been commissioned by external partners or customers to identify data for specific purposes and within a given timeframe. On the basis of accurate requirements that can cover all or part of the following criteria: data type, languages, modality, access and availability, copyright and user license type, conditions for re-use, re-purpose, re-package, documentation.

In order to facilitate the discovery of the data needed for the project, the criteria should be as specific as possible and should also cover all aspects of the data use. In particular, all the aspects that can ease the clearance of IPR issues and the rights negotiation with the owners and/or the right-holders. The storage and the distribution or the sharing should also be envisaged at an early stage.

## 2.4. Legal issues & ethics

The legal aspects impact in a critical way all the phases of the DMP, from the data creation to its storage and use, even more if sharing the data is part of the plans. Today, a large set of licenses exists for data sharing that range from the LR-dedicated licenses by ELRA and by Meta-Share to the more permissive license sets such as Creative Commons.

With the support of legal specialists, ELRA has gained a long experience in managing the IPR issues and supporting the LR licensing. In order to find and define the best license for a specific use, ELRA helps right-holders to answer critical questions such as:

- Who owns the data? Is there a copyright?
- Under which license can we share the data?
- Should we adopt a LR-specific license, a well-known license, a large-scope license?
- Should we use the data as is? Are we allowed to create derivatives?
- Is the access to data restricted and why?

Legal issues relating to the data itself, such as ethics, privacy, confidentiality, data anonymization, should also require special attention and should be handled properly.

## 2.5. Storage & sharing

As a data centre, ELRA has had to face the issues of archiving and maintaining LRs for two decades now. In order to guarantee a stable and on-going access to all the data, a secure and sustainable infrastructure, the ELRA catalogue<sup>2</sup>, has been put into place. Recently, ELRA has been involved in the setup of distributed LR repositories in Europe which sustain the access to data.

Long-term and short-term storage as well as data maintenance, backup processes, use of archiving tools and security management must be anticipated for both the data curation and the data access and sharing.

Making the data visible and discoverable is ensured by a sustainable storage, but it is also supported by proper metadata and the use of a unique identifier such as the International Standard Language Resource Number (ISLRN)<sup>3</sup>.

Dissemination is also a key factor to visibility. In the HLT domain, ELRA has contributed to improve the information dissemination of LRs and somehow impact their sustainability by setting up the LRE-Map<sup>4</sup>. This map, now a standard feature of the LREC Language Resources and Evaluation Conference submission process, aims at collecting information (metadata) about the LRs in conjunction with scientific and technical publications.

## 3. Cataloguing and licensing

### 3.1. “e-Meta-Share” objective and workflow

ELRA’s involvement in the Meta-Share infrastructure<sup>5</sup> remains strong and so is the objective to establish a link between Meta-Share repository and e-commerce possibilities. The challenge is now to build on the acquired expertise new assets (Repository and Workflow Management).

ELRA has been working jointly with the ILSP (Institute for Language and Speech Processing, Greece) to discuss the follow-up of the repository. Recent discussions focussed on how we can move from Meta-Share towards an “e-Meta-Share”. Some upgrades and further developments were initiated at both places (see also section 4.6.2). As far as the metadata is concerned, the following elements were updated:

- addition of the ISLRN number
- improvement of the licensing module: move to new license versions (CC v4.0 and MSNoReD v2.0), deletion of MSCCommons, description of non-standard licenses and terms of use/service.

As said above, e-commerce possibilities are envisaged. These includes in particular the integration of a full e-commerce module (order management, delivery, invoicing and payment facilities), an e-licensing module (automatic filling in and electronic signature), and ISLRN automatic submission.

### 3.2. Cataloguing LRs in Europe

ELRA carries on its identification and negotiation services to release new LRs to meet the HLT community’s expectations. At the end of the first quarter 2016, the ELRA Catalogue consists of almost 1200 LRs shared into Speech

<sup>2</sup> <http://catalogue.elra.info>

<sup>3</sup> <http://www.islrn.org>

<sup>4</sup> <http://www.resourcebook.eu>

<sup>5</sup> <http://www.meta-share.eu>

LRs, Written Corpora, Terminological Resources and Evaluation Packages. Since LREC 2014, ELRA announced the availability of various LRs, in particular the ones coming from projects ELRA was involved in, like the REPERE and MAURDOR Evaluation Packages, or Khresmoi manually annotated reference corpus. Lastly, ELRA also released a first set of Pashto language corpora, produced within the PEA TRAD project supported by the French Ministry of Defence (DGA).

Beyond the day-to-day work of identifying and disseminating LRs through the catalogue, ELRA has constantly offered its investigation services to obtain LRs that could be used for specific purposes within those projects. In this line, ELRA is now involved in the Connecting Europe Facility (CEF) Programme. During the Riga Summit, held late April 2015 in Latvia, the European Commission launched the SMART 2014/1074 European Language Resource Coordination (ELRC)<sup>6</sup>. The objectives of this 2-year project are to:

- improve availability and simplify access to language resources (LRs) relevant for MT,
- establish an observatory for language resources across EU Member States and CEF associated countries,
- raise awareness among stakeholders about the value and use of data for automated translation
- clarify legal and commercial issues related to the data.

Targeted data are those produced by the public sector across all 30 European countries participating in the CEF programme, which can be made available for re-use through the EU Open Data portal, with suitable copyright protection.

The project is coordinated by Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI) and the other members of the European Language Resources Coordination Consortium (ELRC) are ELRA, TILDE and ILSP.

To support the development of the Automated Translation platform of the Connecting Europe Facility (CEF.AT), ELRA and the other three partners of the ELRC committee are organizing local workshops in each of the 30 European countries to raise awareness about the importance of data sets, thus preparing the ground for collecting the LRs needed for CEF.AT.

To answer questions that may arise on the technical or legal aspects related to the use, production, collection, or sharing of Language Resources, ELRA developed a Technical and Legal Helpdesk, available free of charge throughout the project, to provide an answer to any question on the technical or legal aspects related to the use, production, collection, processing, and sharing of LRs.

Any contributor from public administrations may submit aligned corpora from translation memories, terminology resources, lexica or dictionaries to the ELRC Committee at any exploitation stage: simple internet links to websites (Sources), raw data, fully-packaged data (Language Resources). The ELRC-Share repository was developed to support the documenting and storing of all LRs collected.

As a complement to the collection of existing LRs, ELRA and the ELRC Committee guarantees to perform basic data-cleaning, pre-processing, formatting, conversion and alignment on data where required, as well as a quality review of all collected LRs.

Overall, the project targets to gather at least 200 new Automated Translation DSI relevant LRs across EU and CEF associated countries.

### 3.3. LR Citation and impact factors

Language Resource citation is nowadays considered as an important issue. In particular, having a persistent identifier for each existing LR can be considered as an important parameter to compute a “Language Resource Impact Factor” (LRIF). As a first benefit, the merits of LR producers can be better recognized through the use of such an identifier. Moreover, identifying unique LRs will help analyse the LR field evolution in a more efficient way by reducing the duplicated information which may give a wrong idea of the real situation.

With the objective to associate LRs to a unique identification number, ELRA has been promoting the use of the International Standard Language Resource Number (ISLRN). The ISLRN portal was launched in 2014 and its use is increasing since then. Its promotion continues (Choukri et al., 2016). Up to March 2016, 2100 LRs were allocated an ISLRN number, not only ELRA’s and LDC’s catalogued Language Resources, but also the ones from other important organisations like the Joint Research Centre (JRC) and the Resource Management Agency (RMA) who expressed their strong support to this initiative. ELRA is also encouraging the idea of using a BibTeX entry for bibliographical references that would take into account Language Resources items. A stripped-down form of ELRA’s BibTeX proposal is already being used for the LREC 2016 conference papers.

### 3.4. ELRA License Wizard

From the very beginning, ELRA has come across all types of legal issues that arise when exchanging and sharing Language Resources. The association has devoted huge efforts to streamline the licensing processes while continuously monitoring the impacts of regulation changes on the HLT community activities and trying to liaise with all the institutions involved for common actions. The first major step was to come up with a few licenses for both the research and the industrial sectors to use the resources available in the ELRA catalogue. Recently, its strong involvement in the META-SHARE infrastructure focused on designing and drafting a small set of licenses, inspired by ELRA licenses but also accounting for the new trends of permissive licenses and free resources, represented in particular by the Creative Commons.

Within this context, ELRA has developed a License Wizard (Choukri et al., 2016) to support the right-holders in finding the appropriate license under which to share/distribute their Language Resource and clarify the legal obligations applicable in various licensing situations. Currently, the user can choose among the several licenses that exist for the use of Language Resources: ELRA, Creative Commons and META-SHARE.

The License Wizard works as a web configurator that allows:

- Right holders/Users to select a number of legal features and obtain the user license adapted to their selection.
- Right holders to define which user license they would like to select in order to distribute their Language Resources.

<sup>6</sup> <http://lr-coordination.eu/>

- Right holders to integrate the user license terms into a Distribution Agreement that could be proposed to ELRA for further distribution into the ELRA Catalogue of Language Resources.

The ELRA License Wizard<sup>7</sup> portal was launched in April 2015.

## **4. Production activities**

### **4.1. Collection and transcription of speech data in American English**

From 2014 till the beginning of 2015, ELDA undertook a significant spoken language production project, in American English. The goal was to produce several tens of hours of transcribed speech, coming from several hundreds of speakers.

In order to accomplish the goals of the project in a short amount of time, ELDA set up a full-fledged semi-automatic speech transcription platform, leveraging open-source Automatic Speech Recognition tools in order to take advantage of the distributed computing capabilities in today's PCs, while ensuring a thorough statistically-based speech quality control, based on forced-alignment outlier detection, followed by manual post-editing.

Facing the various logistical challenges, ELDA has delivered fully transcribed recordings of several hundred speakers.

### **4.2. Multi-layer annotation in English, French and Arabic**

In 2014, ELDA was involved in a multi-layer annotation project, in English, French and Arabic, whereby tweets have been annotated at a morphological level, while also tagging the named entities therein. Thus, gold annotations for a selected sample of the corpus have been established for data in the three languages, at the PoS and NE levels.

The particularity of this project with respect to other resource production projects undertaken by ELDA in the past is that ELDA now fully handles all LR production tooling in-house, thus leveraging existing open-source software by creating the embryo of a text-based LR production platform, which is already in current use by the LR production team at ELDA, proved itself very useful and, most importantly, seemed to scale the size of the team and of the task very well. To this end, ELDA further enhanced its LR production platform to handle several repetitive corner annotation cases.

### **4.3. Named Entities in Arabic, English and French**

In 2013 and 2014, ELDA has produced three Named-Entity annotated corpora, addressing different domains (data from comments posted on internet, web articles and tweets) for the following languages: Arabic, English and French. These corpora are made up of about 200,000 words per language, out of which a part is selected for double annotation. Annotation guidelines have been created, which are based on those from early annotation projects but refined and with a richer granularity with the aim of covering the needs of the domains treated.

The Arabic data production has proven particularly challenging as its sources sometimes mix both Arabic and Latin scripts within the same text. This has had an impact both on the display of certain items in the data as well as on their annotation.

As a reminder, the objective of this project was to create three Named-Entity corpora for evaluation. The sources comprised very complex and spontaneous-writing data from comments posted on internet, web articles and tweets.

### **4.4. Reference corpora for the evaluation of Machine Translation**

In 2014, ELDA has worked on producing reference corpora for the evaluation of Machine Translation. This project has looked into spontaneous-writing data such as comments posted on internet and web articles, respecting their actual format, segmentation and data noise, which is quite challenging for the professionals producing the reference translations as well as for their later technological applications. The source languages have been Arabic, Chinese, Persian and Russian and 2 reference translations have been produced per language direction, targeting both English and French. This production has followed the strict guidelines and protocols of translation and quality control that ELDA has defined throughout its years of experience in parallel corpora production.

### **4.5. Twitter data annotation in French and German**

Started in 2013 and continuing through 2015, this project aims at producing two Tweet-data corpora, annotated with what are referred to as "opinions/feelings/emotions". The annotation specifications, created for this project, define annotation at different levels, also covering relations between elements inside the tweets. The languages targeted are French and German. Each corpus comprises about 15,000 tweets and 10% of these data are also considered for double annotation.

The single annotation part of the French corpus has been concluded, with only the double annotation remaining for the beginning of February. The German data is currently being annotated at ELDA.

### **4.6. Projects of workflow automation**

#### **4.6.1. MLI**

MLi is looking at needs, achievements and results from on-going and recently finished actions/initiatives (e.g., META-NET), as well as discussing with strategic stakeholders and Digital Service Infrastructures (DSIs) in order to collect that kind of information. A first jet of interviews and information collection is currently being documented in the action's deliverables. In particular, the analysis has looked so far at the needs and requirements of institutions such as MT@EC and Europeana, DSIs with very challenging multilingual needs involving the area of Machine Translation (MT).

ELDA/ELRA is leading WP3 (Resources), which looks at all aspects around Language Resources (LRs), both in terms of data and software, and with a focus on EU

---

<sup>7</sup> <http://wizard.elra.info>

languages as well as languages of major EU partners. The work carried out within this WP is objective-driven and focuses on the analysis of the current situation in Europe, taking into account its linguistic reality, and providing a thorough inventory of both existing resources and resource types required by the different stakeholders, thus identifying major gaps and potential sources of materials. An interesting point within this analysis is its MT language-pair approach, which allows us to look at needs but from a realistic, practical, gap-focused point of view in the multilingual scene. This is particularly complex when languages with limited or fragmentary LR availability are being taken into account.

Furthermore, legal aspects are also being considered, providing guidance on data and software availability and constraints for the data sharing and reusing models to be defined.

ELDA is pursuing the updated project agenda by investigating precise LT needs for the eCommerce sector. Thus, B2C and C2C sites have been identified as potentially benefiting from MT. Namely, user-generated content consisting of user feedback, product appraisal and, more generally, opinions, would benefit from being translated automatically, as the cost of manually performing the translations is too high. Special attention is paid to the preservation of the pragmatic nuances of the user-generated texts: indeed, Machine Translation should alter neither the opinions conveyed, nor the particular stance of the user with respect to a particular topic, nor the style and register of the content.

ELDA's contribution to Work Package 2 on Natural Language Processing software initiatives, architectures and APIs has been well-received by the Work Package partners. ELDA has worked on assessing the Language Technology needs and opportunities in the eCommerce sector and is currently working on a final set of recommendations for using Language Technologies in the eCommerce sector in the European Union.

#### **4.6.2. CRACKER**

CRACKER is a Coordination and Support Action under the H2020 Programme from the European Commission. This action has just started and has held its kick-off meeting in Berlin last February 10th, meeting which has been organised by its coordinator Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI). The other members of the Consortium are: Charles University of Prague (CUNI), Czech Republic; Evaluations and Language Resources Distribution Agency SA (ELDA), France; Fondazione Bruno Kessler (FBK), Italy; Athena Research and Innovation Center in Information, Communication and Knowledge Technologies (ATHENA RC), Greece; University of Edinburgh (UEDIN), UK, and University of Sheffield (USFD), UK.

CRACKER aims at providing planned coordination and support to the European machine translation research community, which is suffering from the pressure of the current challenges and needs of the Digital Single Market. The principle behind this support action is that "sharing data, results and evaluation instruments" leads to a successful evolution and this collaborative work needs to go through the definition of priorities, planning, organisation of evaluations and then sharing of the outcome. For that purpose, CRACKER counts on a

Consortium of experts covering all these needs and it builds upon established networks, infrastructures and evaluation campaigns. ELDA/ELRA participates in this project providing its expertise in all relevant aspects of the language resource sharing landscape.

ELDA decided to undertake Meta-Share upgrades again, by working in close cooperation with the ILSP. The first step, which is currently underway, is to merge ELDA and ILSP's contributions and to publish them on the Meta-Share GitHub repository.

#### **4.6.3. Language Grid at ELRA**

ELRA / ELDA has set up a node of the Japanese Language Grid, in close cooperation with colleagues from the Ishida & Matsubara Laboratory of the Kyoto University, which is driving the Language Grid project. Thus far, a local instantiation of the node has been set up, a first, very naive, lookup-based, Morphological Analysis web service has been deployed at ELDA.

ELDA has also hosted an international Language Grid workshop on the 19th and 20th of November 2015, gathering participants from the USA, Japan, and Europe, which allowed all the participants to define and specify the interoperability between the various Language Grid nodes, as well as new, demand-driven, LR service types.

### **5. Events and dissemination activities**

#### **5.1. ELRA 20<sup>th</sup> Anniversary**

ELRA celebrated its 20th anniversary on October 12 and 13, 2015 in Dubrovnik (Croatia). This 2-day workshop aimed to engage in a substantial brainstorming on the LRs for the Next Decade by exploring the new paradigms in LR management (creation, sharing, etc.) and anticipating the needs of new technologies likely to be mature enough to go to market.

A number of organizations including AFNLP, CLARIN, COLING Committee, ISCA, Language Grid and LDC participated in the workshop, bringing the necessary expertise to work on a prospective analysis of the LR medium and long-term future. This event was also an opportunity to consolidate the coordination committee of the NLP12 institutions so that each of them can contribute to common concerns and initiatives.

The following themes structured the debates:

- Consolidation of citation practices,
- Requirements on common resources for evaluation purposes,
- Requirement to disclose resources used for assessment,
- Increase our influence on ISO and other standardization bodies activities

The objective of the 2-day workshop was to come up with a vision document on the 2015-2025 LR Landscape which will be shared with the community at large.

#### **5.2. LREC**

The 9th edition of LREC has taken place at the Harpa Conference Centre in Reykjavik (Iceland) on the week of May 26-31, 2014. Placed under the UNESCO patronage, the conference has been opened by Madame Irina Bokova, Director-General of the UNESCO on the Monday 26th. It was a very successful edition that brought together more

than 1200 participants from 60 countries, all continents, registered to the conference workshops and tutorials. The Antonio Zampolli Prize was awarded to Alex Waibel (CMU & KTH). A new feature was introduced: the “Shared your LRs” challenge which has been quite successful with 300+ LRs (data, tools, services) shared during the submission to the Main conference and to the Workshops. The 10th edition of LREC will take place at the Grand Hotel Bernardin in Portorož (Slovenia) on the week of 23 to 28 May, 2016. The conference website provides general information on the conference. Information will also be circulated on the LREC 2016 Twitter account (@LREC2016) using #LREC2016. The contact email for any question on the conference is [lrec@lrec-conf.org](mailto:lrec@lrec-conf.org).

### 5.3. New ELRA portal

The new ELRA-ELDA web Portal has been published mid-January 2015, with merged and updated information from both [www.elra.info](http://www.elra.info) and [www.elda.org](http://www.elda.org). Both addresses remain valid to reach the portal. From this single entry point, information is now available: access to the LRs Catalogue, to the information on ELRA membership (including the online form to join the association) and ELRA/ELDA services. The restricted areas are also available to all members of the association. ELRA is also on Twitter and can be followed @ELRANews.

## 6. Future work

On the catalogue’s side, a lot still needs to be implemented with respect to the development of an “e-Meta-Share” version of the Meta-Share repository. This will be a crucial issue for ELRA during the months that follow LREC 2016. As far as the ELRA License Wizard is concerned, more licenses should be integrated at a later stage of the project in order to have the more exhaustive coverage possible. We also would like to gather contributions from other initiatives concerned by LRs licensing like CLARIN or Linked Data. Besides, the representation of legal features in a standardised way would facilitate the HLT community to share their LRs. To this aim, the RDF format could also be explored for the representation of the licenses presented in the Wizard. Licenses may also be easily added or deleted depending on the evolution of their use.

The ISLRN being a requested field within the LREC 2016 submission, we expect that several other LRs will be allocated an ISLRN number by the conference date. With this expansion, this number aims to be a spreadly-used LR citation instrument within works referring to LRs. As a second dissemination phase, more work aims to be carried out both at the technical and the infrastructural levels of the web service. This includes in particular the implementation of an automatic submission to ISLRN from different large data catalogues to facilitate their submission into ISLRN, as well as the setting up of a networked and synchronised (mirrored) server to manage all ISLRN aspects and trusted to a small committee composed of some of the major international LR distribution and sharing institutions.

## 7. Bibliographical References

Choukri K., Liu L., Mapelli V., Popescu V. (2016). Language Resource citation: the ISLRN dissemination

and further developments. In Proceedings of LREC’16. Portorož, Slovenia, 2016.

Choukri K., Liu L., Mapelli V., Popescu V. (2016). The ELRA License Wizard. In Proceedings of LREC’16. Portorož, Slovenia, 2016.

## 8. Language Resource References

REPERE Evaluation Package, in ELRA catalogue (<http://catalogue.elra.info>), ISLRN: 360-758-359-485-0, ELRA ID: ELRA-E0044.

MAURDOR Evaluation Package, in ELRA catalogue (<http://catalogue.elra.info>), ISLRN: 364-018-517-901-2, ELRA ID: ELRA-E0045.

Khresmoi manually annotated reference corpus, in ELRA catalogue (<http://catalogue.elra.info>), ISLRN: 764-036-829-417-7, ELRA ID: ELRA-W0081.