The ELRA License Wizard

Valérie Mapelli, Vladimir Popescu, Lin Liu, Meritxell Fernández-Barrera, Khalid Choukri

ELDA/ELRA

9 rue des Cordelières, 75013 Paris, France

Email: {mapelli; popescu; lin; meritxell; choukri}@elda.org

Abstract

To allow an easy understanding of the various licenses that exist for the use of Language Resources (ELRA's, META-SHARE's, Creative Commons', etc.), ELRA has developed a License Wizard to help the right-holders share/distribute their resources under the appropriate license. It also aims to be exploited by users to better understand the legal obligations that apply in various licensing situations. The present paper elaborates on the License Wizard functionalities of this web configurator, which enables to select a number of legal features and obtain the user license adapted to the users selection, to define which user licenses they would like to select in order to distribute their Language Resources, to integrate the user license terms into a Distribution Agreement that could be proposed to ELRA or META-SHARE for further distribution through the ELRA Catalogue of Language Resources. Thanks to a flexible back office, the structure of the legal feature selection can easily be reviewed to include other features that may be relevant for other licenses. Integrating contributions from other initiatives thus aim to be one of the obvious next steps, with a special focus on CLARIN and Linked Data experiences.

Keywords: web configurator, licensing, Language Resource distribution agreement

1. Objectives

From the very beginning of the association, ELRA[1] has come across all types of legal issues that arise when exchanging and sharing Language Resources (LRs). 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 licenses for both the research and the industrial sectors to use the resources available in the ELRA Catalogue of Language Resources[2]. ELRA contracts were drawn up between ELRA and the resource provider and/or ELRA and the resource user (known as a Distribution agreement for the former, and a VAR-Value Added Reseller or End-User agreement for the latter).

Recently, its strong involvement in the META-SHARE infrastructure[3] 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 ones created by the Creative Commons.

To allow an easy understanding of the various licenses that exist for the use of Language Resources (ELRA's, META-SHARE's, Creative Commons', etc.), ELRA has developed a License Wizard[4] to help the right-holders share/distribute their resources under the appropriate license. It also aims to be exploited by users to better understand the legal obligations that apply in various licensing situations.

2. Methodology

This work was initiated in the META-SHARE initiative[5] within the Meta-Net Network of Excellence[6]. In Deliverable 6.1.3 META-SHARE: Licenses, Legal, IPR and Licensing Issues[7] (Choukri et al., 2012), a number of legal features were extracted from different existing licenses and considered as comparable features between those licenses. Moreover, a list of existing licenses was compiled with a comparison of legal features used in each identified license.

2.1. Investigation on legal features

For the purpose of the License Wizard, the features from the META-SHARE Deliverable were exploited so as to be the basis of the future web configurator. Such features were gathered by legal experts, in particular from the Creative Commons community, as well as experts of the Language Resources field who have to deal with legal questions for the use of Language Resources on a day-to-day basis. The features used in the Wizard are defined below:

- **Implicit versus Explicit:** "implicit" only requires a notification on the web site whereas "explicit" requires a licensing process (with hardcopies, electronic signatures, etc.);
- **Use type:**
  - internal purposes (research, education, not shared with anyone);
  - for research purpose (internal research) ;
  - for research & technology development (including development of products, applications that could be exploited, commercial as well as for non profit;
  - for evaluation of technologies;
  - and/or for any other use;
- **Production of Derivative Works:** which could be a derived resource as well as a derived technology (e.g. technology based on that particular resource or resources combining that one);
- **Distribution of Derivative Works:** making the derived resources available and/or technologies to third parties;

- **Commercial exploitation:** this can be the commercial exploitation of the resource itself or any derived resource and/or product/service;
- **Redistribution of the resource as is (integrity):** the license may allow to redistribute copies of the original resource without any changes (respect to the integrity of the resource or "as is"), it may allow to include it in a bundle or a larger combination of resources, it may also allow it to be encapsulated in a technology (without the authorization to decompile the technology so as to reconstruct the data);
- **Attribution** or **acknowledgment of ownership:** the license may require that the authorship of the resource is acknowledged and the creator mentioned;
- **Irrevocable:** the license may be concluded once for all and users can keep using them as long as they do not breach the clauses but some conditions could be set up to revoke them;
- **Perpetual:** the license is valid perpetually (no duration limit), in fact in most cases, jurisdictions require that contracts have an indication of duration and hence some licenses simply state "as long as the copyright lasts";
- **Other features:** the license may include clauses concerning:
  - Warranty (of the accuracy, completeness, correctness, merchantability or fitness for a given purpose of the resource);
  - Liability and responsibility;

- The Obligation to monitor distributions (users), vis-à-vis suppliers;
- The Jurisdiction.

2.2. Investigation on existing licenses

A number of licenses were identified and analysed within the META-SHARE deliverable (ELRA's, LDC's, NIST's, Creative Commons, Berkeley Distribution Software, GNU, Princeton's). Within the scope of the License Wizard, we focused only on a first set of licenses that consist of:

- **ELRA's agreements:** ELRA's agreements were originally built to facilitate the sharing of LRs between producers/owners of LRs and users. From a work carried out with the support of dedicated lawyers, three types of agreements are now used: End-User (for research purposes), VAR (Value-Added Reseller, for commercial purposes), Evaluation Packages End-User (for the evaluation of technologies). On the producers/owners side, a Distribution Agreement was also drafted;
- **Creative Commons:** these licenses come from the recent trend of public open use. To this aim, the Creative Commons consortium drafted 7 licenses that start with the Creative Commons Zero (CC-0) and all possible combinations along the CC differentiation of rights of use: CC-BY, CC-BY-NC, CC-BY-NC-ND, CC-BY-NC-SA, CC-BY-ND, CC-BY-SA;
- **META-SHARE:** this set of licenses was built within the META-SHARE project with the aim of being LR-

Figure 1: Screenshot showing the list of licenses resulting from the selection of two legal criteria

oriented standard licenses. Like the Creative Commons, they were constituted along the META-SHARE differentiation of rights of use: Commercial No Redistribution, Commercial No Redistribution For-a-Fee, Commercial No Redistribution No Derivatives, Commercial No Redistribution No Derivatives For-a-Fee, Non Commercial No Redistribution, Non Commercial No Redistribution For-a-Fee, Non Commercial No Redistribution No Derivatives, Non Commercial No Redistribution No Derivatives For-a-Fee.

3. Content of the ELRA License Wizard service

The License Wizard was developed as a web configurator that enables the following actions.

From the "Licenses" page, right holders/Users can select a number of legal features and obtain as an output the user license adapted to their selection. Figure 1 below shows the selection page where users may choose the legal features (on the left) that are adapted to their expectations. On the right side is given the list of found licenses corresponding to the selection. The compound criteria filters make searching licenses easier. Besides, those criteria are grouped and include respective definitions to ease their understanding and use of the filters. In the figure, the number of 16 results is found after the selection of two legal features: explicit license and "for research purpose" use type. Alternatively, current selection may be unselected by clicking on the green tags on the top right or even cleared up to start again with a brand new selection.

As an output of the above mentioned selection, the users get the following information:
1) A summary of the legal features selected as a "roadmap" of the user license that apply to their selection
2) A downloadable version for the corresponding user license

Users also get 3 options to proceed on the configurator:
1) Possibility to select as many user licenses as needed by processing the same selection from the beginning again and get all summaries and downloadable licenses in a "basket" that can be viewed at any moment during the selection. The selection is done by simply clicking on the 🛒 icon which allows to gather one or more licenses in "Your selection" page.
2) Possibility to correct the current selection (go back to previous steps)
3) Possibility to produce the distribution agreement (see Figure 2) that integrate the terms that correspond to the selected user licenses in order to enter into agreement with ELRA for the distribution of their own Language Resources

Once users have reached the distribution agreement production page, they are required to fill in a number of details concerning the provider of the resource (name, affiliation, contact details, etc.), the Language Resource name and description, as well as required distribution/pricing policy. Once all details are filled in, the corresponding Distribution Agreement is produced and can be retrieved by the users to submit to ELRA for further distribution.

4. Future work

For the time being, only the ELRA, Creative Commons and Meta-Share licenses have been added to the ELRA License Wizard. However, more licenses should be integrated at a later stage of the project in order to have the more exhaustive coverage possible.

Moreover, as the Wizard back office allows the adding and modification of legal features in a user-oriented way, the structure of the legal feature selection can easily be reviewed to include other features that may be relevant for other licenses. In particular, it would be of interest 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. For instance, within the META-SHARE project, the META-SHARE Commons set of licenses was produced to allow META-SHARE members and Extraneous Depositors to make their resources available to other META-SHARE network members only. However, the abandon of these licenses is now highly recommended, due to the fact that they are now well and fully covered within the latest Creative Commons version 4.0. Only META-SHARE No Redistribution Set of licenses remains applicable. Such recommendation should thus lead to the deletion of the META-SHARE Commons set of licenses in the Wizard. New developments by ILSP (Institute for Language and Speech Processing, Greece) and ELRA for a new version of the META-SHARE repository are already considering this move.

In a more general perspective, other licensing platforms might be worthy of inspection, such as the "Licentia" web application[8] developed by the French National Computer Science Institute (INRIA), which is concerned with licensing any kind of data, with no consideration of their type, although mainly focused on software releases. This platform is also a license configurator, which functions in three modes:
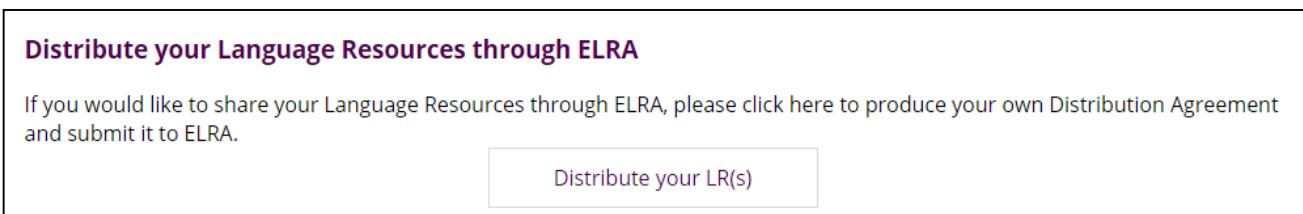

Distribute your Language Resources through ELRA

If you would like to share your Language Resources through ELRA, please click here to produce your own Distribution Agreement and submit it to ELRA.

Distribute your LR(s)

Figure 2: Screenshot showing the message offering to users to go ahead with the production of a Distribution Agreement

---

[8] http://licentia.inria.fr/

1. License selection, according to three criteria: permissions, obligations, and prohibitions. Each criterion can have one value among several predefined options (e.g. for the permissions, the user can choose between Commercialize, Derive, Distribute, Read, Reproduce, Sell). The user can also skip one or several criteria if she or he wishes to.

2. Compatibility checking, between a chosen license and choices of the three criteria listed above.

3. License visualization, in the RDF format, grounded into the W3C's Open Digital Rights Language (ODRL) ontology[9], as well as into the Licenses for Linked Open Data (L4LOD)[10] developed by the INRIA.

Although our aim was to focus on the license features and corresponding licenses dedicated specifically to the use of Language Resources, several ideas from Licentia seem interesting to us and might be adopted by the License Wizard, such as the license compatibility checking functionality.

Regarding the production of the Distribution Agreement, ELRA would like to add a more interactive process so that providers of LRs can register and interact directly with the distribution staff. The current Distribution Agreement allows to add only one resource at a time. It would also be useful for providers to be able to add more than one resource with corresponding distribution policy.

To a larger extent, the data flow exchange between the Wizard and other platforms could also be implemented in order to retrieve user information or resources information from other platforms or provide license information to other platforms. This license information may prove crucial for the management of Intellectual Property Rights within web services. Grid infrastructures present tools as services and allow to easily integrate them giving place to data processing pipelines. The big challenge is to keep track of specific licensing constraints affecting different LRs combined in a workflow and the resulting derivative LRs (e.g. LRs allowing only non-commercial use combined with LRs allowing commercial use). Making visible licensing information to web services is the first step towards automatic license management, thus reducing the risk of license violation. The US Language Application (LAPPS) Grid, for instance, introduces workflow managers which verify (or enforce) the fulfilment of license requirements (Cieri & DiPersio, 2014), either by blocking workflow execution (e.g. if the payment of a fee is required) or by informing the user of further requirements to be fulfilled (e.g. attribution requirements).

5. References

Choukri, Khalid, Piperidis, Stelios, Tsiavos, Prodromos, Patrikakos, Tasos, Gavrilidou, Maria, Weitzmann, John Hendrik, META-SHARE: Licenses, Legal, IPR and Licensing issues, Meta-Net Deliverable 6.3.1, February 24th, 2012.

Cieri, Christopher & DiPersio, Denise, Intellectual Property Rights Management with Web Service Grids, In Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT, pp. 93-100, Dublin, Ireland, August 23rd 2014.

---

[9] http://www.w3.org/ns/odrl/2/

[10] http://ns.inria.fr/l4lod/v2/l4lod_v2.html