The META-SHARE Language Resources Sharing Infrastructure:
Principles, Challenges, Solutions

Stelios Piperidis

Athena RC/ILSP

Athens, Greece

E-mail: spip@ilsp.athena-innovation.gr

Abstract

Language resources have become a key factor in the development cycle of language technology. The current prevailing methodologies, the sheer number of languages and the vast volumes of digital content together with the wide palette of useful content processing applications, render new models for managing the underlying language resources indispensable. This paper presents META-SHARE, an open resource exchange infrastructure, which aims to boost visibility, documentation, identification, openness and sharing, collaboration, preservation and interoperability of language data and basic language processing tools. META-SHARE is implemented as a network of distributed repositories of language resources. It offers providers and consumers of resources the necessary functionalities for describing, storing, searching, licensing and downloading language resources in a single integrated technical platform. META-SHARE favours and aligns itself with the growing open data and open source tools movement. To this end, it has prepared the necessary underlying legal framework consisting of a Charter for language resource sharing, as well as a set of licensing templates aiming to act as recommended licence models in an attempt to facilitate the legal interoperability of language resources. In its current version, META-SHARE features 13 resource repositories, with over 1200 resource packages.

Keywords: language resources infrastructure, metadata-based documentation, legal framework

1. Introduction

No matter what technology or application one intends to build, be it, for example, a machine translation, speech recognition or information extraction engine, a substantial, sizeable data set together with the associated basic processing tools/services is indispensable. Data collection, cleaning, annotation, curation and maintenance proves to be very costly, however. As evidence from other domains (e.g. biotechnology, geodata) shows, data and tools become valuable through opening and sharing. Infrastructural issues – such as interoperability, resource sharing, easy access to language resources and technologies (LR&T) – were recurring messages in all fora organised by FLaReNet (www.flarenet.eu). Many language resource (LR) groups, initiatives and individuals have been advocating since some time the need of a language resource and technology infrastructure: an open resource infrastructure, which allows easy sharing of data and tools that are made interoperable and work seamlessly together, is felt essential. A preliminary set of basic principles and characteristics for such an infrastructure – as emerging from a set of concrete scenarios of its potential use was anticipated by the FLaReNet Steering Committee and made its way into the FLaReNet Blueprint and the Strategic Language Resource Agenda (Calzolari, Quochi & Soria 2011).

In this paper we are presenting META-SHARE (www.meta-share.eu), a sustainable network of repositories of language data, tools and web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources. META-SHARE is developed in the framework of the META-NET Network of Excellence (www.meta-net.eu). Eventually, it aims to be an important component of a language technology marketplace for Human Language Technology (HLT) researchers and developers, language professionals (translators, interpreters, content and software localisation experts, etc.), as well as for industrial players, especially SMEs, catering for the full development cycle of HLT, from research through to innovative products and services.

In section 2, we provide an overview of the META-SHARE design and the infrastructure's architecture. We continue, in section 3, with a concise presentation of the metadata schema used for the description of language resources, and, in section 4, we present the proposed legal framework underlying resource sharing in, and through, the infrastructure. In section 5, we draw the main conclusions and outline future plans.

2. META-SHARE

META-SHARE aims to build an open, integrated, secure and interoperable exchange facility for language resources (data and tools) for the Human Language Technologies domain and other applicative domains (e.g., digital libraries, cognitive systems, robotics, etc). Following current methodological shifts in language resource creation, the term "language resources", in the META-SHARE context, includes language data sets and basic language processing tools.

While language spans all realms of human activity and therefore anything (from distant past language recordings in any language to unregulated blog language and tweets) can be considered useful part of language data,

META-SHARE places a strong focus on language data that are important in language technology development for building applications useful to the EU citizen, primarily in his everyday communication and information search needs. While META-SHARE in its full blown incarnation will be interested and can contain information about all possible forms of and in any language, the first pilot implementations focus on EU languages and major world languages, in fact the languages of the major (business and trade) partners of the EU.

2.1 Design Principles

The goal of META-SHARE is to create an infrastructure of networked repositories that are viewed as one global unified space of language and language-related resources (datasets and basic language processing tools).

It is designed as a **network of distributed repositories** of language resources, including language data and basic language processing tools (e.g. morphological analysers, pos taggers, etc). Repositories can have a local or non-local (central) role. **Local repositories** are set up and maintained by organisations participating in the META-SHARE network, storing their own resources. **Non-local (central) repositories** are also set up and maintained by organisations participating in the META-SHARE network acting as storage and documentation facilities for resources either developed in organisations not wishing to set up their own repository, or they are donated or they are considered orphan resources, etc. Language resources are described according to the META-SHARE metadata schema. Actual resources and their metadata reside in the local repositories. Each repository undertakes the responsibility to a) maintain an inventory (a local inventory) with all the metadata records of their resources, b) export them, c) allow their harvesting. Meta-data records are harvested and stored in the META-SHARE central servers. Central servers share metadata, create, host and maintain a central inventory including metadata-based descriptions of all resources available in the distributed network. Each central server effectively hosts a copy of the central inventory.

Data and tools can be both open and with restricted access rights, free and for-a-fee. META-SHARE targets existing but also new and emerging language data, tools and systems required for building and evaluating new technologies, products and services. In this respect, reuse, combination, repurposing and re-engineering of language data and tools play a crucial role.

Distinct user profiles have been defined, including related authorisations which enable certain actions and ensure the security of transactions. Users may be registered or non-registered, where the former may be divided into end users, providers or administrators of a META-SHARE node. With the exception of non-registered users, every user is given a specific profile containing the information about their rights and obligations.

Consumers of language resources will be able to:
a) register and create a user profile,

b) log-in to the repository network (single sign-on),
c) browse and search the central inventory using multifaceted search facilities (simple and advanced search),
d) access the actual resources by visiting the local (or non-local) repositories for downloading them,
e) get information about the usage of specific resources, their relation (e.g. compatibility, suitability, etc) to other resources as well as recommendations,
f) download resources accompanied by easy-to-use licensing templates
g) provide feedback about resources
h) exploit additional functionality, such as web services

Providers of resources will also be able to:
i) create, store and edit resource descriptions by using a metadata editor implementing the META-SHARE metadata model
j) get support through mapping services from an existing metadata schema into the META-SHARE model
k) upload actual resources directly or by contacting support staff for large volume resources
l) get reports and statistics on number of views, downloads, types of consumers, etc. of LRs , as well as feedback from consumers
m) get support for making available additional functionality (e.g. web services)

META-SHARE has opted for platform independent, **open source solutions** for its implementation. Likewise, all software generated by META-SHARE is open source, released under a BSD licence, available on github at https://github.com/metashare/META-SHARE.

Organisations wishing to set up their own repository will be invited to use the META-SHARE repository implementation. However, in the future, other software solutions can also be accepted, as long as basic responsibilities of creating an inventory, exporting and allowing harvesting of metadata are borne.

2.2 Network structure

The META-SHARE network provides a multi-layer infrastructure that will:
1. make available possibly high quality LR and related metadata over the META-SHARE network,
2. ensure that such LR and metadata are properly preserved and maintained,
3. make a minimum set of services available to all META-SHARE members and users,
4. promote the use of widely acceptable standards for language resource building ensuring the maximum possible interoperability of language resources,
5. allow associated members to easily index their LR over the META-SHARE network, and
6. allow any potential user of the LR to easily and legally safely acquire the requested LR.

The current structure includes 3 layers of membership, entailing different obligations and rights:
a) META-SHARE Members, b) META-SHARE Members acting as managing nodes, c) META-SHARE Associate members.
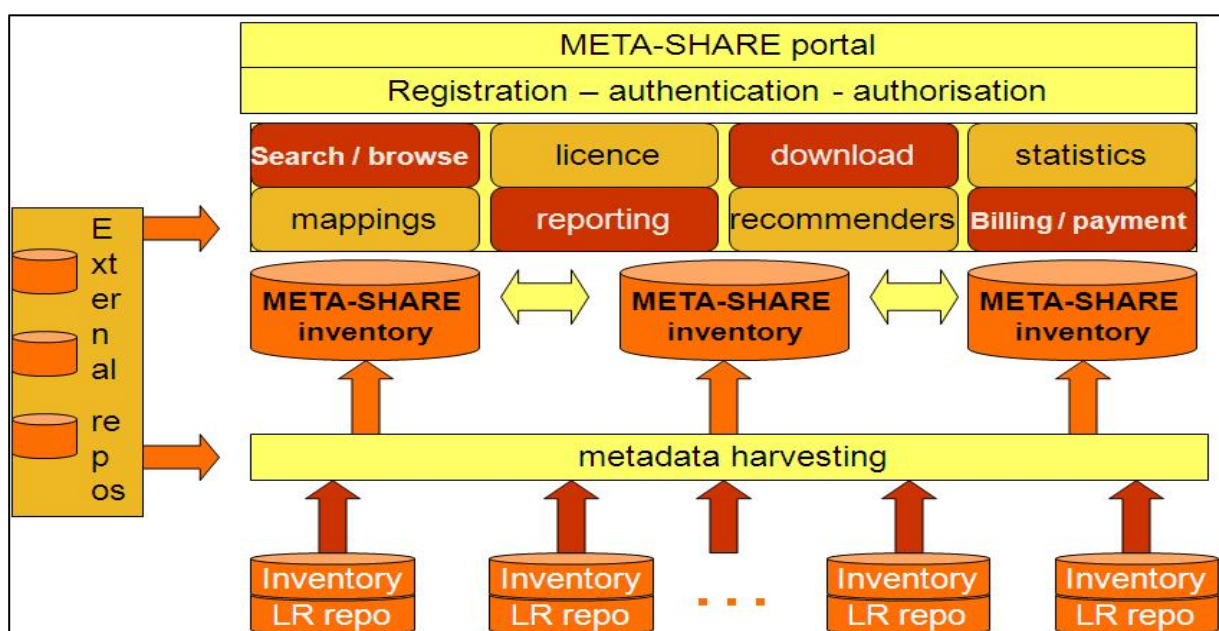
Figure 1: META-SHARE architecture

Members may be all organisations that undertake to a) follow the META-SHARE metadata schema, b) store the LRs and the respective metadata in their repositories, c) maintain an inventory (a local inventory) with all the metadata records of their LR, d) export the metadata records of their LR, and e) allow their harvesting.

Members are able to f) search the central inventory using multifaceted search facilities, g) access the actual resources by visiting the local (or non-local) repositories for downloading them, h) get information about the usage of specific resources, their relation (e.g. compatibility, etc) to other resources as well as recommendations, i) providers of resources can get statistics on number of downloads, types of consumers, etc.

Managing nodes have the obligation to provide a core set of services critical to the whole of the META-SHARE network. They do not have any additional privileges regarding the access to LRs. Currently the role of initial managing nodes is assumed by: Athena RC, CNR, DFKI, ELDA, FBK. New managing nodes are introduced provided they are able to meet the organisational and technical standards necessary to provide core services and have agreed to the Common Service Level Agreement.

Associate members are all organisations or individuals wishing to make their metadata catalogue available to the META-SHARE network but do not wish to follow the META-SHARE specifications with respect to the LRs, use the META-SHARE Core Services, or follow the META-SHARE guidelines regarding sharing and distribution of LRs and metadata. Their obligations are: to make their metadata schema available for mapping to the META-SHARE schema, and subsequently to export and allow harvesting of their metadata.

In addition, *the whole community of consumers or providers* can still search for or donate / deposit resources without being members, as long as legal provisions are met.

At this stage, it is foreseen that each META-SHARE node running a local repository will assign a repository manager responsible for maintaining the inventory and ensuring that uploaded resources are in conformance with intellectual property rights (IPR) and Data Protection rules and regulations as well as any policies the respective organization has in place.

3. Language resources formal description

The META-SHARE metadata model (Gavrilidou et al, 2012) builds upon previous initiatives so as to be *easily, fully and immediately adopted* by the target community.

In the design of the META-SHARE metadata model, central is the principle of a **minimal core subset of metadata**; the elements that form this minimal set are considered indispensable in the process of language resource description and are, thus, obligatory. The minimal level of description is the one at which interoperability with other schemas and typologies will take place.

The META-SHARE metadata schema includes **elements** (most can be linked to ISOcat Data Categories), as well as **relations** used to link together resources that are both included in the META-SHARE repository (e.g. original and derived, raw and annotated resources, a language resource and the tool that has been used to create it etc).

The schema comprises all elements and relations required for the description of language resources, including related tools and services; it refers to any kind of information, including identification parameters, administration information (creation, distribution, licensing), technical information required for their

manipulation, information as to the production and usage (intended and actual), etc.

In order to accommodate flexibility, the elements belong to two basic levels of description: a) an initial level providing the basic elements for the description of a resource (minimal schema), and b) a second level with a higher degree of granularity (maximal schema), providing more detailed information on each resource.

These two levels contain four classes of elements: 1) the first level contains Mandatory (M) and Condition-dependent Mandatory (MC) elements (i.e. they have to be filled in when specific conditions are met), while, 2) the second level includes Recommended (R, i.e. LRs producers are advised to include information on these elements) and Optional (O) elements.

The minimal schema with the mandatory elements will be the *sine qua non* condition for **interoperability** between the META-SHARE metadata model and the other models; mappers / converters will cater for migration from one to the other based on the set of mandatory elements.

Following the ISOcat DCR model, elements are grouped together into semantically coherent "**components**", which, in turn, can include other components.

Obviously, certain components (e.g. the identification, contact and the rights component) are common to all types of language resources, as the elements they contain can be used to describe a resource regardless of type.

The content component, the annotation component etc., being modality dependent, differ across types. The modality of each type determines the description component, which is used not only for resources but also for resource parts. Thus, elements needed for the description of *audio* resources, for example, are only used for the specific *mediaType*.

The core of the model is the *resourceInfo* component**,** which contains all the information relevant for the description of a LR. It subsumes components and elements that combine together to provide this description.

The set of components that are common to all LRs are: *identificationInfo*, *contactPerson*, *versionInfo*, *distributionInfo*, *validationInfo*, *creationInfo*, *usageInfo*, *metadataInfo*, *resourceDocumentationInfo* and *relationInfo*.

Full details on the use of the model are included in the Documentation and User Manual of the META-SHARE metadata model, which is available at http://www.meta-net.eu/meta-share/metadata-schema/. A dedicated user forum, currently available through the metadata editor environment, as well as a metadata helpdesk have been established in order to help LR providers wishing to describe and document their resources in the details and use of the model.

3.1 Recommendations on standards towards interoperability

Building on a report ("CLARIN Standardisation Action Plan") first put together by the CLARIN preparatory phase project (www.clarin.eu) together with FLaReNet,

Figure 2: META-SHARE User Forum

ELRA (www.elra.info) and META-NET, a new document has been revised and updated by FLaReNet with relevant standards for the broader LT community, also addressing those that are typically used in industry, at different levels of granularity. This new document, "The Standards' Landscape Towards an Interoperability Framework" is meant to serve as a general reference guide for all LR providers and users as well as the broader language technology community. The document lists and briefly explains standards proposed and endorsed by META-SHARE for LR encoding and representation. It lists the standards recommended for use in encoding and representing mainly language datasets that will be made available through META-SHARE.

These standards are also referenced in the form of controlled vocabularies used for assigning values to specific elements of the META-SHARE metadata model. It is made clear that this document serves as an information source and a recommendation for following standards and best practices, while it does not entail any actual technical work in the form of conversion of existing resources so that they comply with the proposed standards. By its nature, this document, available at www.meta-net.eu/meta-share/metadata-schema and http://www.flarenet.eu/sites/default/files/FLaReNet_Standards_Landscape.pdf, is a living document that will be constantly updated to reflect progress and developments in the field.

4. Legal Framework

META-SHARE has formulated its basic, high-level, principles in the form of a Charter, the Language Resources Sharing Charter. It is a form of declaration aiming to solicit consensus from the language resources and language technology practitioners on the principles of sharing and its prerequisites. The aim of the Charter (www.meta-net.eu/meta-share/METASHARE_Charter.pdf) is to give a clear signal to language resource providers and users, market players, policy makers and the general public that in the digital world LRs should be shared and further re-used with the minimum possible transaction costs and efforts and under clear and easy to understand rules. The term Language Resources (LRs) in this Charter

includes all language-related digital assets including, without limitation, raw data, processed data, metadata and any other kind of data sets as well as language processing tools, technologies and language related services.

4.1 Licensing Language Resources

Recent surveys show that currently the HLT field seems mature for a move towards openness and sharing of resources. However, licences of resources vary substantially making it difficult for researchers to handle the legal issues around a resource.

META-SHARE creates a "space" within which different LRs may be shared under specific licensing terms, where services of a certain level are offered and a number of standards is followed. The idea of a membership based network is also critical in order to differentiate the concept of sharing vis-à-vis that of opening. Sharing is used to describe the situation where the LRs are shared within the network, whereas opening would not limit the application of the rules to a specific class of users or creators. The structure of the META-SHARE network is such that provides a set of benefits to the members of the network and in that sense invites more members to be part of this platform where a minimum quality of services, security and resources is ensured. It is very important to note that the metadata catalogue remains open to the public, so that all interested parties are able to identify different types of LRs, either in the form of data or technologies/services, they wish to acquire.

The proposed META-SHARE model licensing scheme, with a firm orientation towards the creation of an openness culture and the relevant ecosystem for LRs, is organised on the axes of
- Creative Commons licences,
- META-SHARE Commons licences,
- META-SHARE No redistribution licences,
- Existing standard open source software licences,
which are briefly presented below.

These ready-to-use licensing schemes will be as quick and easy to apply as possible, in particular for new resources. The rights of use of the resource, any possible restrictions as well as rights and restrictions on the original raw data are under the control and responsibility of the resource owners and the repository in which the resource resides. Resources should ideally be available in the public domain. The copyright conditions of the initial raw resource should be known, should be ideally copyright-free or accompanied by one of the permissive licences. Likewise, processed and derivative resources (annotated web or other text, lexica extracted from parallel text) should be open at least for academic/research purposes, allowing their re-use, reengineering and repurposing.

4.2 Creative Commons licences

Creative Commons licences (http://creativecommons.org/) starting with Creative Commons Zero (CC-0) and all possible combinations along the Creative Commons (CC) differentiation of

rights of use are the first level of applicable legal licence templates. META-SHARE favours the open distribution of LRs to the widest possible audience. Members and IPR owners are asked to consider this option before resorting to more restrictive sharing instruments. In META-SHARE then, when open distribution is selected by the LR owner, the CC licences are used. META-SHARE suggests using the CC unported licences (the latest running version, i.e. 3.0) to avoid any issues with regards to the choice of different national CC licences.

4.3 META-SHARE Commons licences

A second layer includes META-SHARE Commons Licences, a fully developed CC-based licensing tool that allows META-SHARE members and Extraneous Depositors to make their resources available to other network members. This class of licences supports the creation of a sharing and collaborative resource building culture, a real challenge for the LR community through the years, and one of META-NET's objectives to drive change and innovation in this respect.

META-SHARE Commons licences cover all expected combinations of licensing attributes, including the distribution of the original resource(s), inside the META-SHARE Network. They are effectively CC-like licences, applicable only within the META-SHARE network, if the LR providers wish to restrict licensing and distribution to the network circle. In that sense we have six MSComs licences: MSComs Attribution, MSComs Attribution NonCommercial, MSComs Attribution NonCommercial ShareAlike, MSComs Attribution Non Commercial No Derivatives, MSComs Attribution No Derivatives, MSComs Attribution ShareAlike.

The decision to use the CC licence texts in order to develop these documents was based on the decision to follow the always useful advice to use existing, well documented, court proven, community supported licences than creating a new one.

Although these licences are based on the respective standard CC texts, and leaving aside the "within META-SHARE only" provision, these licences include the Sui Generis Database Rights in the set of rights covered. This approach has been chosen for the following reasons: (a) in many European jurisdictions, many LRs (datasets) are thought to be covered by the Sui Generis Right for Databases and therefore if we need to have jurisdiction-insensitive instruments we have to take this always into account, and (b) the forthcoming version of the CC licences is expected to cover the Sui Generis Database Right and we have to be able to keep our META-SHARE Commons counterpart licences as compatible to future versions as possible.

An open issue that remains to be discussed and resolved through the development of the META-SHARE network operation is the management of the cumulative rights created by derivative production based on the META-SHARE Commons Licences. This issue should be so resolved as to not undermine the free sharing of LRs. A

step towards this end is a soft norm, a sort of recommendation, that members should ideally be prompt to redeposit derivative work in META-SHARE under the same provisions to those of the initial LR. A system of derivative tracking should then be able to take care of the management of rights.

4.4 META-SHARE No Redistribution licences

The third legal layer is a set of licenses that allow use and exploitation of the resources while permitting the LR owner to have full control over the resource distribution. These "No Redistribution" licences will effectively help get "closed" resources safely out to the community. These licenses also come with a commercial flavour and are accompanied by their for-a-fee version.

This class of licence templates responds to a strong trend among META-SHARE members to build a new branch of licences that would cater for both the need to have the LRs themselves not redistributed and the wish to have legal tools that would support commercial uses in a sharing environment. These licences are also based on standard CC templates. The features on which these licences base their merit are the following:

- they disallow any distribution of the original resource, entailing that the owner retains strong control over the spread of the LR;
- they all have a Commercial and a Non-Commercial variety, allowing for LRs to be both useful in promoting the state-of-the-art in research and providing revenue possibilities for their owners, if and as necessary;
- they all have a "free" and a "for-a-fee" version, providing the requested and necessary freedom to owners to either get some profit out of derivative works or just good willingly promote the use of robust and market-ready LRs.

The set of licences developed essentially includes all combinations of CC licences, naturally excluding the share-alike feature due to the NoRedistribution element inherent in this class of templates.

4.5 Software and services licensing

Software tools and network services are ideally provided though one of the standard **open source** licenses or, if necessary, under a custom commercial license. Open source licenses proposed by META-SHARE include BSD, GPL, LGPL, Apache and AGPL. Other open licences are not excluded, but LR owners are being advised to always keep in mind certain selection criteria that will allow their LRs to be legally easy and straightforward to share.

4.6 Depositor's Agreement: META-SHARE entry point and license selection

The Depositor's Agreement (DA) is the legal document by power of which the Repository is authorised by the Depositor (normally the LR IPR owner) to include a specific LR in its archives and to make it available in any possible form and means, in compliance with the licensing agreement(s) chosen and appearing in the DA

Appendix. The DA assures that the LR owner has all the necessary rights and authorisations to deposit the LR in META-SHARE and obliges the META-SHARE repository accepting the LR to take all measures possible and necessary so that the resource stays always available, is not used in any unauthorized way and is curated in a way that makes it technically easily accessible.

It is evident that the DA is triggered and signed when an LR owner (either a member or non-member) decides to let a META-SHARE non-local repository host and share his/her LR(s). In the case of local repositories, having joined the network will be the equivalent step.

The hosting non-local repository may use another member's or an altogether external service to host and share the LR, provided that it has the necessary means and authority to respect its contractual obligations coming from the DA.

4.7 Further Issues in Licensing

For all types of licenses used in META-SHARE a related table is provided that can immediately act as a helper grid to allow users easy understanding of each licence's features and therefore assist them in choosing the right legal tool (licence) to use. These tables are available to interested LR providers through www.meta-net.eu/metashare/licenses. A dedicated section of the user forum, currently available through the metadata editor environment, as well as helpdesk facility have been set up for assistance in understanding, choosing and using any of these licence and other legal tools.

As said above, the rights and restrictions of use of the resource are under the control and responsibility of the resource owners. The repository in which the resource resides acts mainly as a facilitator for the search-and-get procedure, while providing guidelines and metadata curation activities.

Metadata elements regarding rights of use, availability and distribution are among the elements of the minimal metadata schema and are harvested by the central META-SHARE servers so that users (essentially language resource consumers) can promptly understand what they are allowed to do with a specific resource. Users are asked to always refer to these metadata before using any of the resources on META-SHARE.

It is assumed that Copyright on the metadata that accompany all LRs in META-SHARE is owned by the LR owner who, by depositing the LR (or just the metadata), agrees that the metadata are available under a permissive licence. A Creative Commons Attribution Licence is currently being discussed for metadata licensing. .

It should be noted that the EC is issuing an Open Metadata Licence, while it also plans to issue an Open Data Licence for the EU Data Portal by summer 2012. In addition, Creative Commons is currently consulting on version 4.0 of its licencing suite. As a result, the META-SHARE licensing recommendations will be adjusted accordingly as necessary. What is worthwhile noting further is that the principles already applied in META-SHARE, being in accord with the Open Knowledge Foundation, make the

endorsement of migration to the upcoming new licences versions possible, if necessary and desired by the META-SHARE community.

5. Discussion

We have presented the general principles and design of the emerging META-SHARE infrastructure for language resource sharing and exchange. Such an infrastructural attempt is clearly not a one-off process. It is rather a long-term endeavour by which language resources are recognised as important assets that can boost research, technology and innovation through wide availability, pooling, openness and sharing.

At this point the META-SHARE infrastructure features 13 repositories of resources, one of them hosted by ELRA including most of the resources of the respective catalogue. In total, more than 1200 LR packages currently populate the infrastructure.
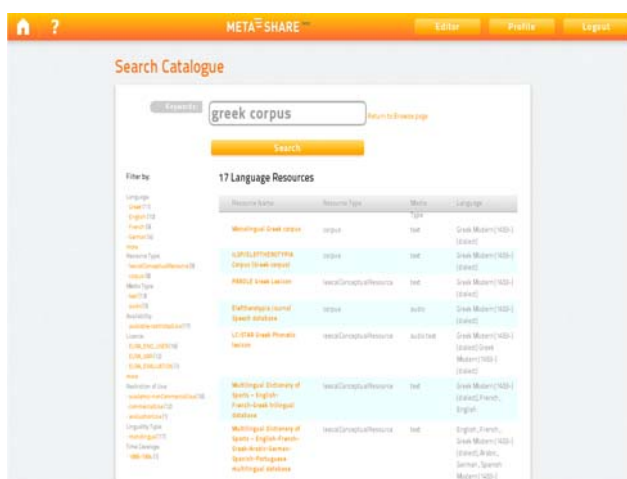

Figure 3: Searching for LRs and filtering results

Population efforts are going on using the public release of META-SHARE version 2. Resources emerging from the ongoing efforts of the LRE Map, as well as the upcoming Language Library in the framework of this LREC2012 will find their way in the respective dedicated META-SHARE repositories.

Particularly challenging in this process seem to be the issues of metadata standardisation, harmonisation and compatibility of the licensing frameworks and, of course, wide availability and openness.

Standardised metadata is indeed a moving target; the metadata world is extremely complex, consensus on the degree of detail, controlled vocabularies, etc. is hard to reach. We have made all efforts to propose an as flexible model as possible, with a minimal and maximal schema, providing mappings and links to ISOcat DCR (http://www.isocat.org/) (ISO 12620, 2009) and Dublin Core (DC, http://dublincore.org/).

Technically, we have achieved to generate the platform's object model automatically from the metadata XML schema (Federmann et al, 2012).

To maximize legal interoperability, we have adopted the widely used CC and Open Software Foundation licences, while all additional licensing templates that proved

necessary have been built on the same principles.

Wide availability, openness and sharing, as indispensable factors of a data economy underlying the attempt to boost a multilingual digital single market, are challenges of a different level. Considering only the data and content at large, generated by the public authorities in the European Union, it is estimated that if such data and content become open, a new market approaching 1.7% of the European Union's Gross Domestic Product can be created. Language technology applications and services apparently have their own share in this multilingual environment. META-SHARE is following closely the developments in the framework of the Digital Agenda for Europe and aims at contributing to all its dimensions regarding use, re-use, re-purposing and exploitation of language resources.

Like for all similar initiatives, success is largely dependent on the level of participation of, interaction with and use by the language technology community at large.

6. Acknowledgements

7. References

Calzolari, N.; Quochi, V. and Soria, C. (2011). *The Strategic Language Resource Agenda.* FLaReNet. http://www.flarenet.eu/sites/default/files/FLaReNet_St rategic_Language_Resource_Agenda.pdf

Federmann, C.; Georgantopoulos, B.; Girardi, C.; Hamon, O.; Mavroeidis, D.; Minutoli, S. and Schröder, M. (2012). META-SHARE v2: An Open Network of Repositories for Language Resources including Data and Tools. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC2012)*, Turkey.

Gavrilidou, M.; Labropoulou, P.; Desypri, E.; Piperidis, S.; Papageorgiou, H.; Monachini, M.; Frontini, F.; Declerck, T.; Francopoulo, G.; Arranz, V. and Mapelli, V. (2012). The META-SHARE Metadata Schema for the Description of Language Resources. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC2012)*, Turkey.

ISO 12620. (2009). Terminology and other language and content resources -- Specification of data categories and management of a Data Category Registry for language resources. http://www.isocat.org.