A License Scheme for a Global Federated Language Service Infrastructure

Christopher Cieri Denise DiPersio

Linguistic Data Consortium
3600 Market Street
Philadelphia, PA USA 19104
{ccieri, dipersio}@ldc.upenn.edu

Abstract. Language service infrastructures are an efficient means for hosting tools and services and processing data, but they can cause complications for licensing language resources. This paper describes the proposed license scheme for the US Language Application (LAPPS) Grid -- an open grid incorporating diverse tools, services and resources – and suggests that the LAPPS Grid license approach can be extended to a global federated language service infrastructure.

1 Introduction

Language service infrastructures, often referred to as grids, have risen to prominence in the natural language processing and human language technology communities, capitalizing on the advantages of cloud computing for processing large amounts of data. The idea is that grids reduce the burden of tool acquisition, integration and hosting by presenting them as services and coordinating their input and output requirements, while the grid infrastructure rapidly builds and executes workflows and pipelines from resources and services. How that framework interacts with licensing constraints is a question that has received some attention and approaches vary across grids. As interest in a global language service infrastructure gains traction, the question becomes how license conditions on multiple resources and tools combined in complex workflows across different platforms can be rationalized to support grid interoperability on a large scale. Solving that problem requires the community to rethink traditional language resource and tool distribution schemes, some of which carry a host of use restrictions, in an environment based on open access and cross-platform integration.

Researchers and organizations that rely on language resources (LRs) are well acquainted with the class of use restrictions under a set of finite standard license arrangements. In that scenario, users take time to integrate the LR into a local workflow before acquiring the next resource. Unless grid developers create a mechanism that coordinates licensing issues while constructing workflows, they risk exacerbating intellectual property issues while they ameliorate tool integration problems. In the sections below, we present steps for handling licensing constraints within a language service grid with a proposal for implementation in a globally-federated language service infrastructure.

2 Web Service Complexities

Web-based language services implement and combine data sets and tools in new ways that may not fit comfortably under established intellectual property law and existing contracts. In a traditional license model, a data center or data provider gives a user the right to process data, but prohibits the user from sharing the LR with others. To the extent that moving the resources over the web for processing could be considered a kind of "redistribution" -- albeit not in the sense of the original license condition -- it is not clear that all copyright holders would consider web processing a permitted use. Shared software in a service grid presents challenges as well. How will users know any license terms or that attribution is required when they are working in an organic grid pipeline where source code is not visible and the command line is not needed?

Add to this the fact that service grids are characterized by multiple stakeholders. *Grid operators* are responsible for the software and servers that support the infrastructure. *Service providers* control access to data and to software. *Users* avail themselves of grid services to access data and otherwise process it. Importantly, grid operators and service providers may or may not be the copyright holders of the software and data underlying the services they provide. Each stakeholder's view of intellectual property protection may vary depending on what is provided by whom to whom. Users generally favor less restrictions than providers do. Some operators and providers may be compensated. Operators will likely want to track user behavior. Service providers can impose multiple conditions including attribution and restricted use (e.g., research) and at the same time, use the data they process for their own research or system development purposes. Moreover, federated grids will have multiple grid operators each seeking to preserve the integrity of their particular infrastructure.

Furthermore, data and software are variously combined in these infrastructures in ways that produce varied effects on licensing. Examples follow in Figures 1 and 2 below.

Figure 1 summarizes three simple grid use cases. The first example illustrates users directing their owned or controlled data through an external service controlled by a second party (Provider 2). In the second scenario, a single entity who is not the user (Provider 2) controls the data and the processing. In the third instance, one external party (Provider 1) controls the data while another controls the software (Provider 2). The presence of multiple parties and actions in each example has the potential to affect licensing depending on the constraints introduced by each.

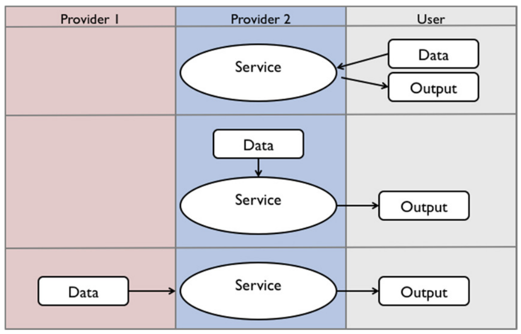

Figure 1: Simple Configurations of Web Services

Figure 2 sketches more complex use cases in which data passes through multiple services. The data may or may not be owned or controlled by the user, while the services are likely controlled by many separate parties as well. Examples of the first two use cases, which show data that is processed through multiple services, might have as its output translated speech that was first transcribed from audio and followed by translation of the transcribed text. The input speech can be controlled by the user (e.g., in voicemail transcription) or by an independent party (e.g., translated newswire). In the third case, multiple services operate on the same data that depend on inputs from other providers for operations on specific languages, such as language identification systems.

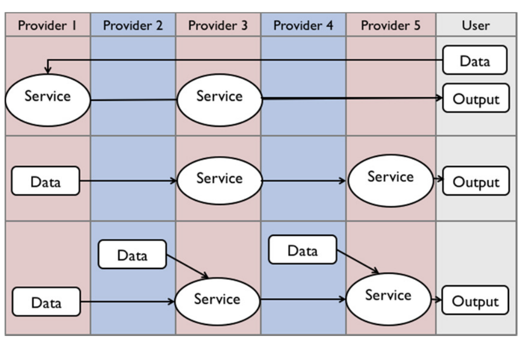

Figure 2: More Complex Web Service Configurations

Moreover, the fact that no party controls the entire system adds another layer of complexity for licensing. Each stakeholder is likely distinct, there are many of them, and even more in a global federated grid, It is expected therefore that each may act in its own interest which probably does not align with the interests of others in the grid community.

3 Approaches to Grid Licensing

There are multiple approaches to grid licensing. One may constrain service and data providers by requiring as a condition of grid participation, that resources are available to particular users under specified terms as in the case of The Language Grid (National Institute of Information and Communications Technology (NICT), Kyoto University). But what if grid service providers are not the owners or developers of the resources? For example, the US NSF-funded Language Application (LAPPS) Grid contains services based on NLTK (Natural Language Toolkit) (Bird, Klein, Loper 2009) and the Stanford Toolkit (Manning et al., 2014). Because the LAPPS service providers do not own those tools, they cannot directly license them to LAPPS users. A solution could be to provide the resources and their underlying agreements and constrain users to comply with the terms. A third alternative assumes that all parties are responsible for their actions during grid operations and no controls are imposed on providers or users. A fourth option restricts providers and users.

The licensing approaches used by existing grids are not easily discovered. They can be gleaned from the grids themselves in a few cases, from papers or web pages in others or by implication based on the licenses used. They are described below from available information.

META-SHARE is a membership-based infrastructure of networked repositories that contain language data and language processing tools. It does not provide grid services as they are described here. It is designed as an infrastructure for data providers and data users to promote resource description and sharing. Those language resources are available under three license types: all combinations of the Creative Commons licenses; META-SHARE Commons Licenses, based on the Creative Commons model, for resources available to META network members only; and “No Redistribution” licenses that prohibit users from redistributing a resource regardless of use, leaving control of distribution to the resource owner.[1] Its metadata catalog is publicly available under a Creative Commons license.

The META-SHARE license types permit a range of controls as seen above, some of which also include the payment of fees. META-SHARE presents the elements of each license group as a table of characteristics, in fact the model of our Table 2 below. The META-SHARE license scheme does not address cumulative rights, that is, what rights attach to any derivative works. Instead, members are asked to deposit any derivatives in the network under the same license as the original resource.

PANACEA (Platform for Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources for Human Language Technologies) was a European project whose object was to create an infrastructure to acquire, produce,

[1] http://www.meta-net.eu/meta-share/licenses accessed 15 December 2014.

update and maintain language resources needed for machine translation systems. Described as a factory, PANACEA acts like a grid in that it offers chained web services (workflows, tools) for processing data. A unique aspect of the platform is its capability to develop data sets on demand by crawling the web; those corpora can then in turn be processed through PANACEA's web services. [2]

The PANACEA licensing strategy is two-fold: (1) the cluster of open source tools comprising the web services are available under various open source software licenses (e.g., Apache 2.0, BSD, GPL); and (2) data sets developed on the platform or provided by users are governed by a non-commercial research only license. In both cases, it is the responsibility of the resource provider to "clear" intellectual property rights for tools and data even if the provider is not the owner of the resource. For the data sets developed by harvesting web sites, PANACEA consortium members undertook to obtain research rights to the source material; any materials for which permission was not obtained were not included among PANACEA resources.

Users can try out the platform on an experimental basis but must register for extended access.[3] The PANACEA project ended in 2012, and the project consortium committed to operating the platform for an additional two years.

The Language Grid developed by NICT is a closed system whose resources and services are available to members only under conditions established by the resource or service provider. There are three use categories: non-profit, research and commercial. License text appears in the resource description when available. When a workflow is executed, the licenses that pertain to the selected tools and data are displayed.[4] One can also browse the available language services which include for each service the "purpose of use," that is, research and/or non-profit.[5] The Language Grid has federated with like infrastructures in Thailand, Malaysia and China that operate under a common Service Grid Agreement. (Ishida, et al. 2011).

Bosca et al. (2012) describe **Linguagrid** as "open to different operators (Universities, Research institutes, Companies) with configurable service access policies: free, restricted to registered users, research or commercial licensing".[6] Linguagrid is administered by CELI, University of Trento (Italy). It is built on the Language Grid infrastructure and presumably employs that grid's license scheme.

CLARIN (the Common Language Resources and Technology Infrastructure) is a networked federation of European data repositories and service centers accessible to users in the participating countries.[7] Its diverse licensing options include those rooted in the Creative Commons licenses with clauses to constrain LRs by user group (e.g., META-SHARE members, academic users). For some resources, papers about them must be reported to the providers and in a variation on a share-alike condition, any derived resources are to deposited in the CLARIN repository.

---

[2] http://panacea-lr.eu/en/project/ accessed 15 December 2014.
[3] http://myexperiment.elda.org/ accessed 15 December 2014.
[4] http://langrid.org/en/index.html accessed 15 December 2014.
[5] http://langrid.org/operation/en/service_list.html accessed 15 December 2014.
[6] http://www.linguagrid.org/ accessed 15 December 2014.
[7] http://www.clarin.eu/ accessed 15 December 2014.

The LAPPS Grid model for license management is described in detail in Section 6. It is open to all users and accommodates a range of license types as well as fees. Since many licenses constrain behavior that occurs post-grid, the LAPPS license scheme is designed to block obvious and immediate violations of licenses, make users aware of constraints that affect future behavior and secure their agreement to relevant terms. Thus, constraints accumulate as the pipeline is constructed and are presented to users prior to the execution of the workflow. Most constraints are presented as notifications which users acknowledge before the workflow begins. A smaller set of constraints are presented as requirements and block the workflow until their conditions are satisfied.

4 Dimensions of Constraints on Language Resource Use

License constraints vary along a number of aspects, starting with the object licensed. Software licenses generally pertain to using software and derivative works of the software, and data licenses regulate the use of the data and derivative works of the data. None of the software licenses reviewed for this paper placed limitations on the use of their output, which is often data. On the other hand, data licenses can and do impose restrictions on using processed data.

The LRs used in web services may be owned by the user, by someone else, or they may be in the public domain. Copyrighted LRs may carry various restrictions: on the use (commercial use, creating and using derivative works); on the user (research labs, non-profit organizations; commercial organizations); on sharing (with whom and how, including attribution and license requirements such as share-alike). There are less common restrictions as well. For instance, we are aware of at least one corpus that requires training in the treatment of human subjects prior to use.

An additional complexity lies in the fact that neither the law nor most licenses distinguish between derivative works (which are typically restricted) and transformative uses (which are typically not restricted). The difference can be illustrated with simple examples from human language technology and natural language processing tasks. Transcribing audio from a copyrighted news broadcast constitutes a derivative work subject, at least in the US, to copyright as well as any license restrictions on the source audio. In contrast, a unigram frequency list based on the transcript is deemed to be a highly transformed work not subject to such limitations.

Many licenses prevent commercial organizations from accessing an LR or using it to develop commercial technology. The motivation in some instances is to encourage direct negotiations with the provider for commercial access which can include a fee. User types typically distinguished by LR licenses include academic institutions, not-for-profit organizations, governments and commercial entities. Cases of pre-commercial technology development may receive different treatment. A licensing model must also recognize those organizations that have executed a required, specific license for a particular resource and those that have not executed the required agreement. Licenses can track users by enumeration or by features. The Linguistic Data Consortium (LDC) maintains databases of all users, all required licenses and the organizations that have executed each license. This is an example of licensing by enumeration. Tracking licenses by organization type (e.g., non-profit organizations) is an example of licensing by feature.

Existing grid licenses in general do not address the use case where service providers wish to benefit from user activity. For instance, a translation service that computes n-grams from processed text that are used to improve the provider's models -- in addition to translating the input text as requested by the user -- raises the question of whether the user can permit, or consent to, such use by the provider.

5 Combining Licensing Constraints

For some combinations of license constraints, users should be notified that a specific workflow is blocked or requires agreement to a set of conditions. Clear cases of the former are those in which some input data requires a specific license that the user had not executed or in which some processing service required a fee that the user had not yet paid. With respect to the latter, a commercial organization should be warned by the grid when it wants to use an LR with a non-commercial restriction and should be required to click-through its assent to that condition before activating the workflow.

In the United States (and likely elsewhere), copyright law and individual licenses commonly associated with LRs do not directly address questions relevant to web-based language services. For example, the notion of "fair use" under US copyright law is not defined, but rather depends on a case-specific analysis under the four-factor statutory criteria. Accordingly, it is expected that laws will be of little help in developing a way to assess the effect on any given workflow of a combination of constraints.

For example, what license attaches to the output of a workflow that uses two LRs, one which permits commercial use and another that does not? We posit a pipeline that consists of a language recognition service that identifies the language of the input and routes it to a machine translation service. If the language identification service relies on an LR that cannot be used commercially, can the resulting translation be sold if the input data and the translation system permit commercial applications? We may think this is acceptable, but would our thinking change if the data used by the translation engine was restricted to research pueposes? Is the answer different if the input text cannot be used commercially but other components in the pipeline could?

Another thorny area is the derivative work-transformative work continuum. Should an LR with a no derivatives element in its licensing contract be blocked from further processing on the assumption that such processing might be a derivative use? As shown in Table 3 below, the LAPPS Grid license model does not block processing on those grounds, but provides the user with notifications about any conditions on derivative and transformative uses.

Of some comfort perhaps is the fact that grid licensing is not so different from traditional LR license schemes in that users agree to a set of conditions and providers are not generally informed about the planned use. The gaps in the law referred to above are present in both instances. And in both, users are expected to abide by any applicable agreements and conditions. From a data center perspective, we can say that the language research community generally acts responsibly in that regard. The noteworthy difference in the web language service environment is that the analysis of multiple license terms and users' acquiescence to them happen on the fly, raising the concern that some users may miss the import of the license. Hence the need for careful planning in the grid license infrastructure to include user-friendly license information and click-through options as well as any necessary authentication mechanisms.

6 The Language Application Grid

We now consider the resources implemented in the LAPPS Grid as a model for a license management solution.

To date, the LAPPS Grid has used 27 unique software packages (programs, toolkits, APIs, libraries) covered by the nine licenses summarized in Table 1.

Table 1: LAPPS Grid Software by License

License	Software
Apache 2.0	Language Grid, NLTK, ANC2G0, UIMA, OAQA, Uimafit, guava-libraries, ActiveMQ, AnyObject, Jaxws-maven-plug-in, Jetty, OpenNLP
BSD	Hamcrest, NERsuite, CRFsuite (in NERsuite)
CDDL 1.1	Jaxws-rt
CPL 1.0	MALLET, AGTK, JUnit
Eclipse 1.0	logback (v1.0), Jetty
HTK-Cambridge	HTK
MIT	Mockito, libLBFGS (in NERsuite), GIZA (v3)
Python	NLTK
WordNet	Genia tagger library (in NERsuite)

The LAPPS Grid includes a small number of data sets. Those include the Manually Annotated Sub-Corpus (MASC), an open,resource that can be used by anyone for any purpose,[8] and portions of LDC's Gigaword corpora, distributed under LDC's standard license model.[9]

Many of the constraints imposed by those licenses fall into recognizable categories summarized in Table 2

[8] http://www.anc.org/data/masc/ accessed 9 January 2015.
[9] https://catalog.ldc.upenn.edu/LDC2011T07 accessed 9 January 2015.

Table 2: LAPPS Grid Licenses and Common Constraints

License	Redistribution	Use	Derivative Use	Attribution	Share Alike	$
Apache 2.0	Yes	Commercial	Commercial	Yes	No	N
BSD	Yes	Commercial	Commercial	No	No	N
CDDL 1.1	Yes	Commercial	Commercial	Yes	Yes	N
CPL 1.0	Yes	Commercial	Commercial	No	No	N
Eclipse 1.0	Yes	Commercial	Commercial	Yes	Yes	N
HTK-Cambridge	No	Commercial	Commercial	No	No	N
MIT	Yes	Commercial	Commercial	No	No	Y
Python	Yes	Commercial	Commercial	Yes	No	N
WordNet	Yes	Commercial	Commercial	Yes	No	N
LDCFP Member	No	Commercial	Commercial	No	No	N
LDCNFP Member	No	Research	Research	No	No	N
LDCNon-member	No	Research	Research	No	No	Y
CC-Zero	Yes	Commercial	Commercial	No	No	N
CC-BY	Yes	Commercial	Commercial	Yes	No	N
CC-BY-SA	Yes	Commercial	Commercial	Yes	Yes	N
CC-BY-ND	Yes	Commercial	None	Yes	No	N
CC-BY-NC	Yes	Research	Research	Yes	No	N
CC-BY-NC-SA	Yes	Research	Research	Yes	Yes	N
CC-BY-NC-ND	Yes	Research	None	Yes	No	N
GPL (v2,3)	Yes	Commercial	Commercial	Yes	Yes	N

These many licenses have in common the constraints and values summarized in Table 3.

Table 3: LAPPS Grid Common License Constraints and Values

Constraint	Values
Redistribution	Yes/No
Use	Commercial/Research Only
Derivative Use	Commercial/Research Only/None
Transformative Use	Commercial/Research Only /None
Attribution	Yes/No
Share Alike	Yes/No
Fee	Yes/No
Other Constraint	--

Grid operators have less flexibility with respect to licensing conditions than providers under the historical distribution model. In the latter case, any fees are generally required in advance and it is not unusual for providers to condition resource delivery on a signed license or click-through consent. Or users may receive the LR and its license on the understanding that the user's consent to license terms is deemed made when the resource is used. Also, as mentioned earlier, most licenses address future events, such as redistribution, derivative works, attribution and share-alike. Thus, a key consideration for a grid licensing model is for it to accommodate those kinds of license provisions in real time as a workflow is built and executed. We address this in the LAPPS Grid by establishing two classes of enforcement, **requirement** and **notification** (summarized in Table 4). For required actions, a pipeline is blocked until conditions are met. Otherwise, users are presented with accumulated conditions before the pipeline is executed. Actual licenses must be made available as well since summarizing license terms is not a legal substitute for the subject license.

The two types of enforcement, requirement and notification, are naturally implemented differently in the LAPPS Grid. Notification is treated similarly to a click-through license. Specifically, the software used to build grid pipelines queries each service as it is added to the pipeline for any licensing constraints. Those constraints may include Creative Commons primitives or the requirement to agree on the fly to specific licenses such as those listed in Table 1. The user is offered the opportunity to review each of those licenses and is notified that continued execution of the pipeline signals agreement with their terms. Requirement is implemented through a special module that connects the user to the organization responsible for enforcing the relevant constraints. The module passes to the authorizing organization the identifier of the resource requested and a token uniquely identifying the session. The authorizing organization may require the user to present login credentials, make payment or otherwise demonstrate that he has satisfied the constraint, after which it returns an approval or rejection that causes the pipeline to be executed or blocked respectively.

Table 4: LAPPS Grid License Constraint Enforcement

Constraint	Action
Redistribution	Notify
Use	Notify
Derivatives Use	Notify
Attribution	Notify
Share Alike	Notify
Fee	Require
Other Specific License	Require
Other Specific Constraint	?

7 A Federated Grid Licensing Model

We propose the framework in Figure 3 for a federated grid licensing model.

Users initiate their sessions by authenticating themselves in one of the federated grid frameworks. Resources and services are requested from the workflow management tools. For instance, in the LAPPS Grid, the Composer (Ide et al. 2014) displays available tools and services which are selected by the user in their preferred order which can include multiple parallel operations on the same data. The Composer directs resources to the appropriate service, taking into account varying tool input and output requirements. Using the LAPPS Grid Planner, users specify input and output requirements and a pipeline is then constructed.

Grid services are linked to the workflow managers, so users cannot implement in a pipeline any resources outside the grid. As the user builds a workflow, the management tools query license conditions from each requested resource or service; they may also query an API or data center regarding the user's satisfaction of license conditions. The pipeline is blocked if certain required conditions such as a fee or a signed license (Table 4) are not satisfied. If there are no required pre-conditions, a list of click-through licenses and their provisions are accumulated by the manger as the pipeline grows. The end result is a summary of restraints with links to the license texts with which the user must agree before processing can commence. Similarly, any service license conditions such as attribution or statements from a README file or in the command line are also displayed by the manager.

The success of this model depends on the existence a closed grid system where few management programs control each process. Some problems cannot be resolved, such as the distinction between derivative and transformative uses. Our proposed licensing scheme utilizes a conservative legal approach in that case, issuing appropriate warnings about uses that might be considered derivative.

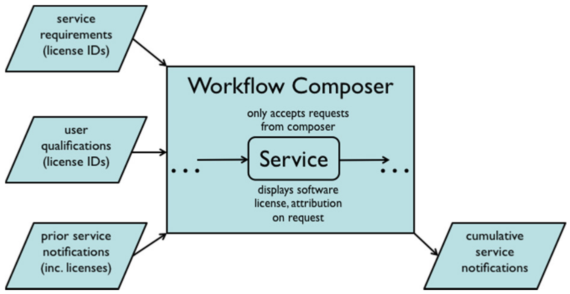

Figure 3: Federated Grid Licensing Model

Where formerly independent service grids are federated, we must also address the question of managing the variation in practice related to intellectual property that arises from their separate evolutions. In Section 3, we provided examples of how several extant grids approach licensing. In subsequent sections, we proposed a model for managing agreements between software and data service providers, on the one hand, and users on the other. Here, we continue by discussing the kinds of agreements that must be coordinated across all stakeholders in federated grids. We will set aside differences in local law, which are beyond the scope of this paper, focusing instead on differences in agreements.

Federated grids must decide whether participants should sign agreements developed specifically for the federation. As noted above, existing grids seem to differ with respect to how "membership" and related agreements are treated. If there is no single federation agreement, it will be necessary to address how to resolve differences in pre-existing grid agreements.

Federated grids must also consider the basis upon which grid operators, service providers and copyright holders participate and how to deal with mismatches. For example, does a non-profit grid operator have any say as to whether providers may offer services for a fee? Along with that question comes the issue of what responsibility stakeholders assume by virtue of working together. If a service offered commercially becomes unavailable to the detriment of users, does the grid operator or service provider accept responsibility? Similarly, if any users, service providers, grid operators or software developers disrupt a grid, whether their home infrastructure or a federated grid, whether accidentally or intentionally, who assumes responsibility and what are the remedies? Finally, does any grid operator, service provider or copyright holder make any warranties of any kind relative to their offerings? Should disputes arise between grid users and providers, or between operators or federated grids, how are these disputes resolved and in which jurisdiction? This becomes especially important in the case of a dispute between a user and a remote grid operator.

Secondary issues include what information grid operators or service providers may collect from users and does that vary when the user comes from a remote grid? This will be particularly important in the case of unique, proprietary and business sensitive data. Also, in the event of changes to the grid or its hosted services, who is responsible for notification of the change and how does that information flow to other stakeholders? Finally, who decides whether a user is authorized to use the grid, does such authorization commute to federated grids, is it similarly revoked from all grids if revoked from any?

8 Conclusion

We discussed the challenges web language service infrastructures present for licensing language resources and how those challenges are addressed in the US LAPPS Grid. The LAPPS Grid license schema is based on a two-fold enforcement mechanism – requirement and notification. Under that model, most pipelines will be executed once a user agrees to the accumulated license provisions that attach to workflow components. A few pipelines that include resources with pre-use requirements such as a fee or signed license will be blocked until the condition is satisfied. This model protects intellectual property interests while permitting credentialed users to construct complex pipelines. Finally we proposed an extension of the LAPPS Grid license scheme to an open globally-federated language service infrastructure.

9 Acknowledgements

This work was supported by US National Science Foundation grants NSF-ACI 1147944 and NSF-ACI 1147912.

10 References

Arranz, V., Choukri, K., Hamon, O., Bel, N., Tsiavos, P. (2012) PANACEA Project D2.4, Platform Software, Project Tools + Resources, Licensing Policy and Exploitation Plan.

Bird, S., Klein, E., Loper, E. (2009) Natural Language Processing with Python. O'Reilly Media.

Bosca, A., Dini, L., Kouylekov, M., Trevisan, M. (2012) Linguagrid: A network of Linguistic and Semantic Services for the Italian Language. In Proceedings of the Eighth International Language Resources and Evaluation (LREC12), Istanbul, Turkey. European Language Resources Association (ELRA).

Cieri C., DiPersio, D. (2014) Intellectual Property Rights Management with Web Service Grids. OIAF4HLT Workshop: Open Infrastructures and Analysis Frameworks for HLT. The 25th International Conference on Computational Linguistics (COLING 2014), Dublin, Ireland.

Ide, N., Pustejovsky, J., Cieri, C., Nyberg, E., DiPersio, D., Shi, C., Suderman, K., Verhagen, M., Wang, D., Wright, J. (2014) The Language Application Grid. In Proceedings of the Ninth International Language Resources and Evaluation (LREC14), Reykjavik, Iceland. European Language Resources Association (ELRA).

Ide, N. and Suderman, K. (2014) The Linguistic Annotation Framework: A Standard for Annotation Interchange and Merging. Language Resources and Evaluation.

Ishida, T., Murakami, Y., Tsunokawa, E., Kubota, Y., Sornlertlamvanich, V. (2011) Federated Operation Model for Service Grids. The Language Grid, Springer, pp. 279-298.

Ishida, T., Nadamotoa, A., Murakami Y., Inaba, R., Shigenobu, T., Matsubara, S., Hattori, H., Kubota, Y., Nakaguchi, T., Tsunokawa, E. (2008) A Non-Profit Operation Model for the Language Grid. International Conference on Global Interoperability for Language Resources: 114-121.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D. (2014) The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

Piperdis, S. (2012) The META-SHARE Language Resources Sharing Infrastructure: Principles, Challenges, Solutions. In Proceedings of the Eighth International Language Resources and Evaluation (LREC12), Istanbul, Turkey. European Language Resources Association (ELRA).