CLARIN-D AP6 report: Legal and ethical issues
May 2012 – April 2013

May 2013
1 Introduction .......................................................................................................................................... 3
2 Legal Information Platform .................................................................................................................. 3
3 Publications ......................................................................................................................................... 4
4 Presentations ......................................................................................................................................... 4
5 Committees and CLARIN Representation ......................................................................................... 5
6 CLARIN Legal Issues and Submission Policies .................................................................................. 5
7 Plans for 2013 and beyond ................................................................................................................ 6
Introduction

AP6, or Work Package 6, focuses on legal and ethical issues regarding the usability and re-usability of language resources in CLARIN-D. The overall goal of AP6 is to provide legal information and resources for the benefit of CLARIN-D partners, as well as the language science and digital humanities communities in general. The tasks and goals of AP6 are guided by the outline and milestones laid out in the CLARIN-D Vorhabenbeschreibung (project description).

Heading up work on AP6 at the Institut für Deutsche Sprache are Erik Ketzan (Wissenschaftlicher Mitarbeiter or Research Associate since 2011) and Pawel Kamocki (the same, since 2012). Mr. Ketzan previously worked as an attorney in the United States and has published scholarship on copyright law and language technology. Mr. Kamocki is currently completing his degree in copyright law at Paris Descartes University and also holds a Masters degree in linguistics.

Legal Information Platform

IDS has continued building a comprehensive Legal Information Platform, currently available on the CLARIN-D Wiki (http://de.clarin.eu/de/home/arbeitspakete/ap-6-rechtliche-und-ethische-fragen and http://de.clarin.eu/mwiki/index.php/Legal_Information_Platform), to serve as a resource on legal issues for CLARIN-D researchers. The Legal Information Platform currently runs to over 30 pages of entries on various legal topics that have been identified as most relevant to the language science and digital humanities communities, much of it original research by Ketzan and Kamocki, and provides general information, analysis, and best practices. Entries created in Year 1 have been polished and updated:

- When to consult institutional and private lawyers
- Licensing issues
- Copyright and author’s rights under German and international law
- Copying and distribution exceptions for scientific research under the Urheberrechtsgesetz (UrhG, German Copyright Act)
- Term/length of copyright protection
- Database rights under the EU Database Directive (96/9/EC) and its incorporation in German law, plus a discussion of how this differs from international approaches
- Data privacy issues under the Bundesdatenschutzgesetz (BDSG) and international law, including standards for the handling of personal data and anonymization

Meanwhile, new entries have been added on:

- Building corpora from open resources (Wikipedia, etc.)
- Building corpora from online text in general
- Building corpora from online forums
• Best practices for dealing with publishers and authors
• Planned changes to German, EU, and UK law
• Language Science Legal Issues Bibliography
• Sample disclaimer forms for spoken interviews for oral corpora
• Orphan works and the Orphan Works Directive
• Sharing software - open source licenses
• Sharing non-software - open source licenses

The Legal Information Platform will grow in the future to include extensive discussion of these and other issues that affect language science projects within CLARIN-D and digital humanities in general.

Useful documents, such as sample contracts and forms, will continue to be added in the coming months. Examples of such documents include license agreements and disclaimers for participants in recorded interviews or dialogues.

Publications

Content created, developed, and fine-tuned on the Legal Information Platform is, in stages, finalized as formal publications. This past year, Ketzan co-authored a book with chapters on legal issues, and Ketzan and Kamocki together published on data privacy issues for oral interviews. More publications are in the works and in various stages of completion.

  o Contains chapters on copyright and other legal aspects of data management

  o Contains legal best practices for language scientists to build spoken corpora

Presentations

• E. Ketzan, "An Introduction to Legal Issues in eHumanities and Developments in EU Law," eHumanities Kickoff Workshop 2013, University of Leipzig, Germany, 9.4.2013.
We are taking the first steps toward a set of legal submission guidelines for resources within CLARIN, and will continue work on integrating legal information in metadata as well as legal policies for long-term archiving of data. It will be the task of the newly formed Legal Helpdesk

Committees and CLARIN Representation

Witt, Ketzan, and Kamocki sit on the Legal Issues Committee of CLARIN, and were represented by Kamocki at the first Legal Issues Committee in Sofia, Bulgaria, in October 2012.

Issues being discussed by the Committee include the CLARIN Server Provider Federation Agreements, deposition agreements and EULAs for the three categories (PUB, ACA and RES) of resources within CLARIN, and coordinating CLARIN’s position statement when taking part of legislative dialogues, particularly regarding “academic use” exceptions to copyright law and positions on changes to licensing law.

In last year’s M12 report, the AP6 team wrote that, “IDS will continue to contribute to talks with publishers in advancement of CLARIN-D goals. In addition, IDS plans to conduct research, present and publish on the creation of EU-wide fair use-type of legislation, particularly for scientific research.” Toward this goal, the AP6 team is currently active in the foremost European-wide initiative to reform licensing for text and data, namely the current European Commission initiative, “Licenses for Europe”, which aims to make improvements to, among other things, laws and licenses in the EU that affect text and data mining for researchers (including language scientists). “Licenses for Europe” includes representatives from numerous stakeholders in the legislative reform debate, especially publishers. Ketzan and Kamocki have attended meetings in Brussels as representatives of the CLARIN Legal Issues Committee and advocates of the interests of CLARIN-D researchers, and will continue to do so in 2013.

Additionally, Kamocki represents CLARIN at the Research Data Alliance, an international group that aims to accelerate and facilitate research data sharing and exchange.

CLARIN Legal Issues and Submission Policies

Members of the AP6 team assisted with internal CLARIN and CLARIN-D legal issues, including aiding in the finalization of the CLARIN Service Provider Federation Agreements in conjunction with the AP3 team, and the implementation of the Persistent Identifiers (PID) and CLARIN Identity Provider (IdP), also in conjunction with the AP3 team.

We are taking the first steps toward a set of legal submission guidelines for resources within CLARIN, and will continue work on integrating legal information in metadata as well as legal policies for long-term archiving of data.

to communicate these submissions policies to researchers in conjunction with the AP7 (Support and Help Desk) team.

**Plans for 2013 and beyond**

Throughout 2013, IDS will continue to expand the Legal Information Platform and continue revising its contents as formal publications, for the benefit of both CLARIN-D partners and the research community. Papers are currently underway on numerous legal topics, including: orphan works, personal data in linguistic resources, liability of service providers, the implied license doctrine (comparative law analysis), language corpus as a database, whether IP law allows reproducibility of linguistic research, the importance of patent law for language researchers, and the protection of ideas in academic standards and in copyright law.

IDS will continue representing CLARIN-D and CLARIN at national and international legislative hearings, such as the European Commission's "Licensing for Europe", German implementation of the Orphan Works Directive, and others. IDS will push for statutory exceptions that make the use of resources easier for researchers, as well as solutions that simplify, streamline, and lower the cost of licensing text and data for research. In 2013 we have made significant first steps in the stated goal of our M12 status update: "The eventual goal is to give CLARIN-D partners a place at any table that shapes legislation drafting and lobbying efforts in this direction."

IDS will continue to assist in the incorporation of legal issues in technical solutions, such as:

- Legal requirements for access to resources through IdP (identity provider) software and access for various user groups
- Connecting to AAI (Authentication and Authorization Infrastructure)
- Distributed user- and license-based authorization
- Automatic detection of abuse, etc.
- Incorporation of legal/license issues into metadata