

Legal and Ethical Issues

with respect to LRT and e-infrastructures

Marc Kupietz

in collaboration with

Christina Bankhardt

 INSTITUT FÜR DEUTSCHE SPRACHE

D-SPIN Scientific Advisory Board Meeting
Garching, 2009-10-18

D-SPIN

CLARIN

Overview

1 Introduction

2 Legal Background

- Intellectual property rights (mainly written LR)
- Personal rights (mainly spoken LR)
- Additional de facto restrictions

3 Implications for D-SPIN/CLARIN

4 Further Plans

1 Introduction

2 Legal Background

- Intellectual property rights (mainly written LR)
- Personal rights (mainly spoken LR)
- Additional de facto restrictions

3 Implications for D-SPIN/CLARIN

4 Further Plans

Aims of CLARIN WP7 & D-SPIN AP7

General goal

Make Language Resources as easily and freely available as legally, ethically and de facto possible.


Subgoals

- best practice guidelines readily usable by the community
- model contracts for legal relations
- specification of requirements on the technical infrastructure
- disseminate a deeper understanding to the community

Responsibilities

- University of Helsinki for CLARIN-WP7 / European layer
- Institut für Deutsche Sprache for D-SPIN-AP7 / German layer
 - first-hand experience with legal issues
 - challenging amount of written and spoken language resources
- close collaboration indispensable
 - 1 to ensure that German peculiarities and specific interests find their way into the CLARIN framework
 - 2 to make the network of legal relationships work

Division of labor & workflow

- D-SPIN-AP7 delivers input:
 - national legal peculiarities
 - existing license terms, contracts
 - user needs
 - technical requirements
 - CLARIN-WP7
 - gathers, clusters, canonizes
 - develops legal relations framework, contracts
 - takes care of technical requirements
 - D-SPIN-AP7 examines, validates and localizes
- 

Current status of work in AP7

Comparatively small work package
(8.5 PM distributed over 3 years)

- 1 analysis of the legal background and its German peculiarities



- 2 analysis of existing needs & practices



- 3 development of solutions wrt to CLARIN/D-SPIN framework



↑
today

report R7.1 focuses on 1

1 Introduction

2 Legal Background

- Intellectual property rights (mainly written LR)
- Personal rights (mainly spoken LR)
- Additional de facto restrictions

3 Implications for D-SPIN/CLARIN

4 Further Plans

The Problem

- the majority of digital research resources in linguistics and the social sciences and humanities is subject to a variety of legal and ethical restrictions
- this is not going to change in the near future

Specific problem in linguistics and related disciplines

- not only its 'secondary' research data is protected by intellectual property rights (IPR)
 - i.e. published scientific research
 - but also its primary research data
 - i.e. recorded observations of language use
 - and rights belong to entities
 - outside the public sector
 - outside the scientific community
- open access models from other communities are not transferable
- but getting primary research data in most disciplines costs money (e.g. LHC, ...)

1 Introduction

2 Legal Background

- Intellectual property rights (mainly written LR)
- Personal rights (mainly spoken LR)
- Additional de facto restrictions

3 Implications for D-SPIN/CLARIN

4 Further Plans

Collision of basic rights:

- freedom of science and research



- guarantee of property
 - general personal right
 - right to informational self-determination
 - right to privacy of personal information

- the conflict cannot be solved in general
- but . . .
 - there are **specific** solutions for our community (and better ones with CLARIN/D-SPIN)
 - specific laws further regulate the conflict

German Copyright Act (UrhG)

some specifics

- copyright applies automatically
- is not transferable
- applies until 70 years after death of the author
- protected: . . . *works of language, such as literary works, speeches and computer programs* . . . (§ 2 UrhG)
- no qualitative requirements for the work, but a *height of creation* (threshold of originality) must be achieved
- *“The author has the right to decide whether and how the work is to be published”* (§ 12 UrhG, moral rights / originator’s personality right)

Bounds of German copyright

- temporary acts of reproduction (§ 44a UrhG): “Caching”
- right of citation (§ 51 UrhG)
- education and research (§ 52a UrhG):

It is permissible to make . . . published parts of a work, works of small size and individual articles from newspapers or magazines publicly accessible exclusively for a specifically defined group for their own scientific research, as far as this is done for the respective purpose which is non-commercial.

- § 137k UrhG:
§ 52a is no longer applicable after December 31, 2012.

Legal ways of providing written LR in Germany

- 1 use texts that stay under the threshold of originality
- 2 use the research bound
- 3 use the caching bound
- 4 wait 70 years after all authors have died
- 5 provide your own texts
- 6 use license contracts

Ways of providing written LR in Germany

actually used (results of 2009-09-28 AP7-workshop at the BBAW in Berlin)

- 1 stay under the threshold of originality
 - web-based corpora with shuffled sentences
- 2 use the research bound
- 3 use the caching bound
 - all web-search engines, some web-based LR
- 4 wait 70 years after all authors have died
 - historic corpora
- 5 provide your own texts
 - probably to some degree – extensible in CLARIN/D-SPIN (publishers could join)
- 6 **use license contracts**
 - most language resources

License agreements

between LR-providers and copyright holders should contain provisions for:

- purpose of the license agreement
- subject matter (kind of data)
- data preparation
- data storage
- long-term preservation
- types of use
- size of visible/exportable passages of a text
- end user group
- end user license agreement (EULA)
- liability
- protective measures against misuse
- term of a contract/right of revocation

1 Introduction

2 Legal Background

- Intellectual property rights (mainly written LR)
- Personal rights (mainly spoken LR)
- Additional de facto restrictions

3 Implications for D-SPIN/CLARIN

4 Further Plans

Collision of basic rights:

- freedom of science and research



- guarantee of property
- general personal right
 - right to informational self-determination
 - right to privacy of personal information

- the conflict cannot be solved in general
- but . . .
 - there are **specific** solutions for our community (and better ones with CLARIN/D-SPIN)
 - specific laws further regulate the conflict

German Federal Data Protection Act (BDSG)

- § 40 II BDSG Processing and use of personal data by research institutions
»Personal data shall be anonymized as soon as the research purpose permits this.«
- § 40 IV BDSG:
Bodies conducting scientific research may publish personal data
 - 1. only if the subject has consented or*
 - 2. if it is indispensable for the presentation of research findings on contemporary events (»Ereignisse der Zeitgeschichte«).*

Legal way(s) of providing spoken LR in Germany

- 1 use declarations of consent & anonymized data
- 2 (use totally anonymized data)
 - hardly achievable for audio data (voice)

Declarations of consent

Requirements (and problems)

- must be signed before recording
- responsible body of the research project has to be named
- information on the research project, purpose, use and processing of the records
- time of deletion should be stated explicitly
- indication of voluntary participation
- possibility of revocation has to be stated
- in spite of anonymisation revocation must be possible

1 Introduction

2 Legal Background

- Intellectual property rights (mainly written LR)
- Personal rights (mainly spoken LR)
- Additional de facto restrictions

3 Implications for D-SPIN/CLARIN

4 Further Plans

Additional de facto restrictions

In general, legal potentials **cannot** be fully exploited, because . . .

- *quod licet iovi, non licet bovi*
- LR-builders must gain and retain their reputation as trustworthy partners of LR-donors:
we depend on a good relationship with our informants
- most resources can only be obtained after signing a contract
→ extension of fair/academic use bounds might not help
- legal opinions are expensive and ultimately not reliable
 - in legal gray areas, different judges will decide differently
- of ethical reasons
- CLARIN/D-SPIN cannot advise to enter legal gray areas

1 Introduction

2 Legal Background

- Intellectual property rights (mainly written LR)
- Personal rights (mainly spoken LR)
- Additional de facto restrictions

3 Implications for D-SPIN/CLARIN

4 Further Plans

Implications for LR (in D-SPIN/CLARIN)

the legal and property situation results in a . . .

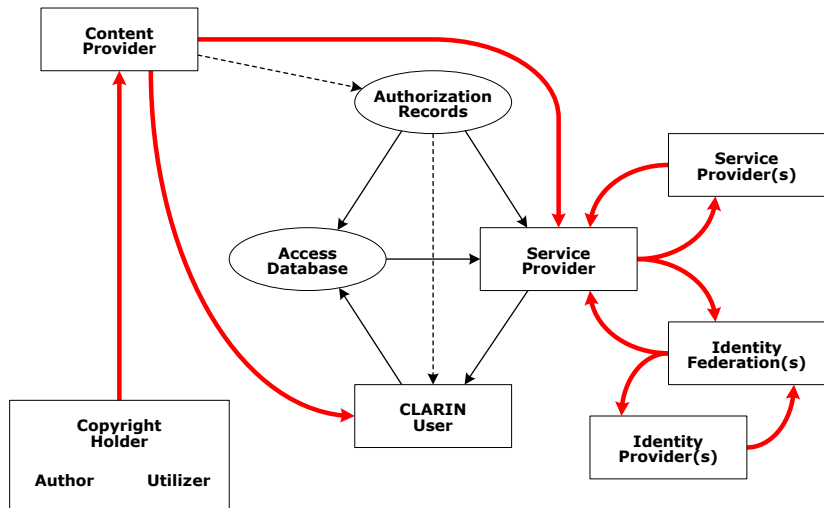
- trade-off between different properties of a LR:
 - quality (degree of distortion, shuffled/unshuffled, . . .)
 - size (some but few LR are freely usable)
 - accessibility (downloadability, authorized groups)
 - re-usability (for different purposes)
 - costs
- not all properties can be optimized at the same time
- different resources will use different emphases/weightings depending on the respective research purposes
- different license models are necessary
- canonization only possible to some degree

Implications for D-SPIN/CLARIN

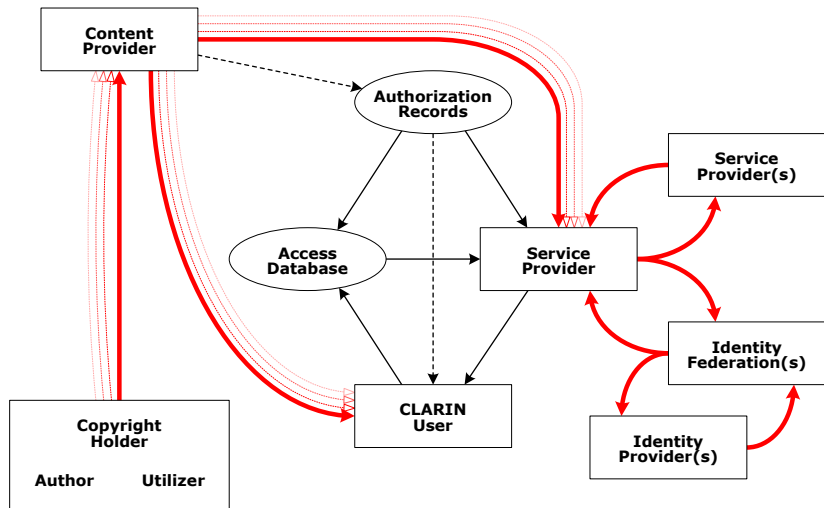
to make LR as available as legally, ethically and de facto possible we will need . . .

- to deal with resources under many different license models
 - to be receptive for present and future LR
- to deal with different types of use
 - linguistics / SSH, non-/ commercial – (priorities?)
- a complex network of contracts
 - to propagate license restrictions
 - regulate the legal relationships between all entities
- a technical infrastructure, that . . .
 - maps the contractual network
 - controls authorization (identity and license based)
 - facilitates the integration of new resources
 - provides technical means to detect misuse
 - makes the best of the license terms

CLARIN legal entities/roles and their relationships (1)



CLARIN legal relationships (2)



What will be the benefits of CLARIN/D-SPIN for LR providers and users with respect to legal issues?

- agreement templates and best practice guidelines
 - can be used independently of CLARIN/D-SPIN
 - will reduce the overhead of LR building projects
 - will ensure a compatibility of agreements
 - will help to make LR as (re-) usable as possible
- the legal and technical infrastructure together will
 - provide a framework for authentication and authorization
 - reduce the overhead for making LR available
 - help to reduce costs (but not to 0)
 - help to keep LR available after projects have ended
 - solve some severe problems
 - publication of “derived resources”: e.g. manually annotated parts of corpora

1 Introduction

2 Legal Background

- Intellectual property rights (mainly written LR)
- Personal rights (mainly spoken LR)
- Additional de facto restrictions

3 Implications for D-SPIN/CLARIN

4 Further Plans

What we have done so far

- 1** analysis of the legal background and its German peculiarities
 - results in report R7.1
 - peculiarities coordinated with CLARIN WP7
- 2** analysis of existing needs & practices
 - collected and analyzed existing agreements
 - evaluation of results from Berlin-workshop: “needs and practices” in progress
- 3** development of solutions wrt to CLARIN/D-SPIN framework
 - example agreements for different relationships developed
 - concepts for technical solutions in discussion

Threads to follow (1)

what we promised and will do:

- 1 localize (model) contracts from CLARIN-WP7
- 2 localize best practice models from CLARIN-WP7
- 3 organize another workshop
(with focus on discussion of proposals, dissemination)

Threads to follow (2)

what we will also (try to) do:

- 1 offer practical advice to the community
- 2 establish the task permanently at the IDS
- 3 develop an “ethical codex” for spoken LR
- 4 publish example contracts
- 5 develop further concepts for technical solutions
- 6 develop a checklist/registry of license terms
- 7 develop a manifesto to amend copyright law
- 8 co-edit a special issue of a major corpus linguistics journal

Thank you for your attention!

Questions?

Comments?

Suggestions?

Disagreements?