



D-SPIN

**D-SPIN Report R3.3:
Case Studies – Intermediate
report**

March 2010

D-SPIN, BMBF-FKZ: 01UG0801

Deliverable R-3-3: Case Studies – Intermediate Report

© All rights reserved by BBAW on behalf of D-SPIN

Editors: Frank Binder, Ralf Gehrke, Lothar Lemnitzer

1 Summary	4
2 Summary of the status of the study on using language resources for computer-assisted language learning (CALL)	5
2.1 Introduction	5
2.2 Content overview	5
2.3 Conclusion and outlook	7
3. EURObabel	9
4. In-depth interviews with humanities researchers	11
5. Concluding remarks	14
Appendix 1: study on using language resources for computer-assisted language learning – (preliminary) Table of Contents	15
Appendix 2: Guidelines for the in-depth interviews	17
Einleitungsteil	17
Gezielte Einzelfragen:	17
Prototyp	19

1 Summary

D-SPIN is the German counterpart of the European Research Infrastructure project CLARIN (Common Language Resources and Technology Infrastructure, <http://www.clarin.eu/>). The ultimate objective of CLARIN and D-SPIN is to create a federation of existing digital repositories that include language data, and to provide uniform access to those data, wherever there are. CLARIN and D-SPIN aim also at providing existing language and speech technology tools as web services to retrieve, manipulate, enhance, explore and exploit the data. The primary target audience are researchers in the humanities.

Within the CLARIN federation, the focus of D-SPIN is on German resources, tools and their integration through web services. In addition, some of the project partners are preparing for their role of acting as service centers on the national and pan-European level. Besides these localization efforts, D-SPIN has a special focus in addressing potential users of the infrastructure with the preparation of training material and teaching activities.

This report covers the activities of D-SPIN work package 3: “Ressourcen- und Anwendungsplanung im Hinblick auf geisteswissenschaftliche Anwender”. The document, being one of D-SPIN’s M24 deliverables, captures the state as of March 2010. It provides an overview of the activities towards the case studies that have been defined in the proposal of the project:

- Computer Assisted Language Learning with special focus on Intelligent Computer Assisted Language Learning
- Recording and documentation of endangered languages (the BABEL project)

These case studies serve to analyze the resp. field of research and scientific practice with regard to the use of language resources. They will document the current situation in the fields as well as requirements of practitioners concerning the development of a language resources and tools infrastructure. The outcome should be well-defined scenarios, in which the role of D-SPIN is defined. Thus, the case studies will help to prepare model solutions in the implementation phase, which are based on the infrastructure which is currently being built.

In addition to these studies, which focus on a narrower field or research, we report on a series of in-depth interviews with researchers in the humanities. These interviews will give a broader perspective on both the current practice and the future needs of humanities researchers.

Finally, we will document, in two appendices, the table of contents of the CALL study in preparation and the questionnaire we use for our in-depth interviews.

2 Summary of the status of the study on using language resources for computer-assisted language learning (CALL)

2.1 Introduction

A case study on using language resources for Computer-Assisted Language Learning (CALL) and Intelligent Computer-Assisted Language Learning (ICALL) is being conducted at the University of Giessen. The study is based on three reports, written specifically for this study by scientific staff members of the research network “Educational Linguistics” at the Justus Liebig University Giessen. The contributing authors focus on the usage of language resources in their respective fields. The resulting integrated report, which is being edited currently, comprises 75 pages of text. A preliminary summary of this report was presented at the workshop “Kumulatives Arbeiten mit Textarchiven” (cumulative usage of text archives) at BBAW Berlin on March 26th-27th, 2009.

2.2 Content overview

The report is organized along the following topics:

1. Corpus linguistics
2. Learner corpora
3. Corpora and foreign language teaching
4. Using language resources for ICALL

The report’s first section on corpus linguistics primarily deals with concordance analysis and performing keyword analyses with concordance tools like WordSmith or AntConc, or web-based tools like the “Compleat Lexical Tutor”. All the methods mentioned in this section operate on the surface form of the text, with all limitations implied. Usage scenarios from this field of research will be applied to languages other than English, in order to provide relevant input for D-SPIN. One of the central findings of this part of the report is that in order for language resources to be used appropriately in didactic scenarios, teachers and tutors need sufficient training and guidance towards “corpus literacy”. This stresses the importance of training measures as part of the D-SPIN project and its followers.

The second section on learner corpora provides an overview of the field. Resources such as the German learner corpus “FALKO” and the “International Corpus of Learner English” (ICLE) are presented and their relevance for research in second language acquisition as well as for designing language teaching materials is discussed. As an example for a relevant research method in second language acquisition, the authors mention Contrastive Interlanguage Analysis (CIA), which relies on

language resources in the form of learner corpora. Furthermore, learner corpora can be subject to “immediate pedagogical use”, where learners themselves analyze their own interlanguage in order to make progress towards their target language. One of the characteristics of learner corpora is their rich and “non-standard” annotation. Users of these corpora must be enabled to exploit and use this annotation and to add their own layers of annotation to the primary data. These requirements pose high demands on a language resource infrastructure. In reverse, the gain for researchers in this field will be high.

The third section discusses the use of corpora for foreign language teaching. It is organized as a review of publications on that topic with a focus on access to data and their use in research and teaching contexts. Again, basic look-up functionality for keywords in context and contrastive analysis of local learner corpora with larger collections of the target language are the most demanded functionalities. With regard to the teaching of pronunciation and intonation, multimodal and sound archives are indispensable resources. Insufficient teacher training and the lack of easily accessible, efficiently usable resources are mentioned as crucial obstacles for the wider use of corpora in language teaching. Again, training and guidance are of utmost importance to enable researchers as well as practitioners to properly use available or to provide their own resources to their colleagues and students.

The section on using language resources in ICALL commences with a discussion of how the concept of CALL has developed during the last decades under the influence of various approaches in learning theory, such as behaviorism, cognitivism, and constructivism. In contrast, ICALL has been a rather technically motivated business, closely tied to advances in artificial intelligence, natural language processing, machine learning and machine translation. One particular issue in ICALL is the generation of feedback to learners through an automated error analysis based on error-annotated learner corpora.

The text pinpoints several attempts that have been made in the previous decades to apply NLP-tools such as taggers and parsers, lexical resources, word lists and corpora in ICALL-systems. It also outlines some of the difficulties which have arisen. Both the field of language learning and teaching and the field of natural language processing have been advancing independently of one another. Transfer of knowledge and resulting improvements of the applications have rarely occurred directly. Rather, there has always been a complex interplay of advances in language learning theory and instruction design, causing frequent changes in requirements on the side of the domain of application, with rather isolated creation of resources and technology-driven prototypes that provided unsatisfactory solutions to those moving-targets.

However, the authors express their confidence that despite the diverse approaches of individual ICALL projects, it is possible to specify general requirements for language resources that are to be used within ICALL applications. Making standard resources and basic functionality available through web service APIs could help future ICALL projects, since there would possibly be less need for them to create their own resources for integration into their ICALL systems. It remains to be seen, to which extent the availability of such standard resources reduces the costs and efforts invested into research and development of ICALL-technology and -applications.

In summary, the domain of ICALL poses the highest demands on a language resource infrastructure because in this domain the use of sophisticated NLP tools and richly annotated language data as well as lexical resource is without an alternative. Sharing resources, tools and knowledge is therefore an immediate demand. In the other domains the benefit of using sophisticated tools is not that obvious, nor is the impact of the linguistic annotation of language resources. Much needs to be done to raise the awareness of researchers and practitioners in the field. Some of the works which are reported however show that computer-aided language teaching will profit from a rich and easy accessible landscape of (multilingual) language resources, if, and this is the other important point, the infrastructure is manageable for the user who in most cases is not a computer expert.

2.3 Conclusion and outlook

Based on these texts, some common patterns of how applications based on language resources are used in the covered fields are visible. These patterns will be used as input for the specification of use cases, which will be integrated as a separate section of the report. It is to be expected that these use cases will, in a first attempt, cover very common and basic ways of using language resources. In that sense, they will complement existing activities in requirements engineering for the CLARIN/D-SPIN infrastructure. One issue of the specification of such use cases is a fair estimate of the load which they will cause to providers of language resources and tools.

We must be aware, however, that it would be out of scope for the current phase of the D-SPIN project to develop full-fledged applications that could serve the needs of end users. Rather, the project provides basic resources and services that shall reduce the costs of research and development of such applications. Given that perspective, we need to extract from these use cases how existing and future applications can best benefit from the emerging language resources infrastructure. The definition of such use cases based on the findings of the report will be the subject of subsequent work in this work package.

Furthermore, as has become obvious especially in the field of language teaching, many requirements are not of purely technological, but rather of organizational nature: Training efforts are needed, along with access regulations, so that resources and applications can be used with acceptable effort and at affordable costs.

The full report will be distributed early enough before the meeting in December to the D-SPIN Advisory Board and the project partners.

3. EUROBabel

Very important for D-SPIN are collaboration and information exchange with major international projects which are concerned with the development and the analysis of linguistic corpora. For the current phase, the project EUROBabel (preparing language material in endangered languages) has been chosen as partner. Analyzing its approach and requirements will be very helpful in evaluating the tools and methods that are being developed by D-SPIN.

The EUROCORES programme EUROBabel (“Better Analyses Based on Endangered Languages”) of the European Science Foundation consists of 5 international, linguistically oriented projects (<http://www.esf.org/activities/eurocores/running-programmes/eurobabel.html>):

1. Alor-Pantar languages: origins and theoretical impact
(Project leader: Dr. Marian Klamer, Leiden) (<http://www.alor-pantar.org>)
2. The Kalahari Basin area: a ‘Sprachbund’ on the verge of extinction
(Project leader: Prof. Tom Güldemann, Berlin)
(http://www2.hu-berlin.de/asaf/Afrika/Forschung-EN/Kalahari_Basin.html)
3. Ob-Ugric languages: conceptual structures, lexicon, constructions, categories - An innovative approach to creating descriptive resources for Khanty and Mansi
(Project leader: Prof. Elena Skribnik, Munich) (<http://babel.gwi.uni-muenchen.de>)
4. Referential Hierarchies in Morphosyntax: description, typology, diachrony
(Project leader: Dr. Katharina Haude, Cologne) (<http://www.rhim.uni-koeln.de>)
5. Endangered sign languages in village communities
(Project leader: Prof. Ulrike Zeshan, Preston, Lancashire)
(http://www.uclan.ac.uk/iscri/islands/deaf_villages.php)

A short description of the individual projects can be found under

<http://www.esf.org/activities/eurocores/running-programmes/eurobabel/projects/list-of-projects.html>.

EUROBabel and, more generally, the community of researchers in the field of endangered language studies, was chosen as a target group for the investigation of resource infrastructure requirements for several reasons.

The programme is interesting from the point of language resources because the various projects will generate a large collection of heterogeneous types of data, e.g., spoken utterances, gestures, sociolinguistic information, anthropological data. Research will cover many areas of linguistics and beyond, such as morphosyntax, grammar, and analysis of genetic relations between languages, convergence processes. And also the results will be manifold, consisting of electronic grammars,

electronic dictionaries, electronic editions of rare old publications as well as presentations of the collected materials and research results.

The projects will generate a lot of audio and video material and a great amount of structural data which will have to be stored and maintained permanently and securely (curation aspect). As the focus of the programme is on endangered languages, lost or irrecoverable data would probably be irreplaceable.

There is also need for sophisticated metadata descriptions to make the material accessible and usable for the research community beyond the researchers which are directly involved in the programme. This will require networking with other researchers and initiatives to assure long-term accessibility and usability (sustainability aspect).

In addition to this, EUROBabel is still in an early phase, so information exchange with D-Spin can be profitable for both sides.

At the Launch Conference of EUROBabel (Berlin, September 11th-13th 2009), D-Spin was presented by Ralf Gehrke (Frankfurt), and first contacts were established. Besides this presentation, two initiatives were present to offer their experience and their infrastructure to EUROBabel: Prof. Peter Austin from the “The Hans Rausing Endangered Languages Project at SOAS” and Dr. Angela Terril from the “The Documentation of Endangered Languages Programme (DOBES) by the Volkswagen Foundation”. While Dr. Terril offered a CLARIN consistent approach, Prof. Austin favored individual resource structures not restricted by standards. There was some discussion about these two approaches, but no decision yet from the side of the project leaders.

The goal for the next three months is to identify demands and requirements of the projects involved, making use of the questionnaires developed in Berlin. The present moment seems especially suitable for integrating problems and necessities newly discovered during the starting phase of EUROBabel. On the other hand, the phase of the projects is still early enough to incorporate changes and adaptations of norms.

4. In-depth interviews with humanities researchers

In addition to the online survey – see D-SPIN Report 3.1 for details – the D-SPIN group at BBAW, in collaboration with the D-SPIN group at the University of Frankfurt, is conducting explorative in-depth expert interviews with specialists and key persons of different fields within the humanities. We are conducting interviews with a small list of experts in the following areas:

- *Historical semantics*: Prof. T. Gloning (Gießen), Prof. Bernhard Jussen (Frankfurt)
- *Literature*: Prof. Anne Bohnenkamp-Renken (Frankfurt)
- *Art History*: Prof. Thomas Kirchner (Frankfurt)
- *Work edition*: Regina Roth (Marx-Engels Ausgabe)
- *Historical edition*: Bärbel Holtz (Acta Borussica)
- *Cultural History*: Claudia Sedlartz (Berliner Klassik)
- *Ancient Oriental Studies*: Prof. Cancik-Kirschbaum
- *Classical Studies*: Wolfram Brunschön and Roland Wittwer (Corpus Medicorum Graecorum/Latinorum)
- *Philology*: F. Martin (Corpus Vitrearum Medii Aevi, BBAW)
- *Linguistics*: Prof. Manfred Stede (Potsdam)
- *Library science*: Anton Stecker (Wolfenbüttel)

Some of the researchers are in their majority working in long-term-projects at the Berlin-Brandenburgische Akademie der Wissenschaften (BBAW) and have already experienced (technological) changes which affected their daily work.

Other researchers we recruited in the context of the so-called *LOEWE* programme (this acronym stands for: “Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz”) of the state of Hessen. Within this framework, the “Johann Wolfgang Goethe-Universität Frankfurt”, “Technische Universität Darmstadt”, “Freies Deutsches Hochstift / Frankfurter Goethe-Museum, Frankfurt” and “Städel Museum, Frankfurt” submitted a joint proposal for the installation of a “Schwerpunkt [=focus project] Digital Humanities”. The submitters represent the fields of Linguistics, Literature, History, Arts History and Computer Science. Within the intended project an intense collaboration with D-Spin is planned, especially with respect to the tools which should be deployed.

The collaboration with researchers working with language resources, but with research questions and concerns not primarily of a linguistic nature (as e.g. the subject of “Historical Semantics”) should be of considerable interest for D-Spin. The same holds with research across media types, which is the case with Arts History and their integrated text-image-corpora.

The interviews will commence with the main research topic of the interviewee. It will shed light on the whole work process. In other words: the interview is focused on the language resources involved in the research process, but they take a wider context into account. An important step, in particular for processes which lead to an edition, is the publication process of the results and the long-term maintenance of these results. It is assumed that a research infrastructure should also take care of this part of the research process.

The interview questions will cover the following topics:

- a) The expert’s own research and the role which language resources can play or could play in it, both as object of investigation and as “background resources” which are consulted during this investigation;
- b) The current scientific practice with respect to the use of language resources in the field at large;
- c) The expert’s vision of the research process in the future and what is required with respect to the availability and usability of language resources and tools. This includes legal issues with regard to resources which are used by the researcher as well as resources which the researcher produces and plans to provide to the community of researchers and/or a larger public.

An important part of the interviews will be a guided usage scenarios based on a prototypical implementation of a web-based application. Lacking a handy prototype built on the D-SPIN infrastructure, we decided to use, where this is appropriate, an implementation of a tool for collaborative and cumulative annotation of texts which is currently being developed by the German Text Archive (Deutsches Textarchiv-Prototype, <http://www.deutsches-textarchiv.de/>). Presentation and interactive work with such a tool will hopefully stimulate further ideas of the interviewee with regard to new tools for the humanities work place.

Another point of reference for our activities is a comparable subproject within the Dutch CLARIN framework that we are going to establish with the primary researcher(s) of this project. As far as we

know, they also plan a series of in-depth interviews based on a questionnaire. However, their focus is on the field of linguistics.

We will propose to organize a common workshop of both projects where we will present the findings of the interviews, exchange ideas and formulate a user-driven strategy for the further development of the language resources and tools infrastructure. We also want to discuss issues of trans-national cooperation.

5. Concluding remarks

D-SPIN Work package 3 reviews the needs of humanities (and, to some degree, also social science) researchers with regard to language resources and tools from three different angles:

1. Two case studies analyze a relatively narrow field of research in depth and will lead to **concrete scenarios** which, once they are implemented, can serve as show cases for the usability of the D-SPIN infrastructure;
2. A quantitative analysis of the **current practice** of the use of language resources via a questionnaire (results have been reported during the last meeting of the Scientific Advisory Board, October 2009)
3. A qualitative analysis of the **future needs** of individual researchers which as a group represent a much broader spectrum of the humanities. This will lead to a workshop in which the findings of these interviews will be presented. They will also be presented on the next meeting of the Scientific Advisory Board in December 2010. Currently, this is still work in progress.

The international embedding of our activities is currently far from ideal. This is partly due to the reluctance of the EC to fund, within the preparatory phase of CLARIN, activities which are similar to ours. Of course we have observed the activities of CLARIN WP 3 and take them into account.

A measure to encounter this problem is to seek allies in other countries where national projects are being funded. We will start with the Netherlands and will try to organize a common workshop. CLARIN will of course play a prominent role in this workshop.

Appendix 1: study on using language resources for computer-assisted language learning – (preliminary) Table of Contents

Gliederung der Gießener Studie „Computer-Assisted Language Learning unter dem Aspekt sprachlicher Ressourcen und deren Nutzung in anwendungsbezogenen Workflows“
Stand: 26. 03. 2010

1 Zu Korpora und Korpuslinguistik allgemein	5
1.1 Korpus und Korpuslinguistik	5
1.1.1 Korpus - eine Definition	5
1.1.2 Korpus-Typen	5
1.1.3 Korpuslinguistik	7
1.1.4 Einführungen in die deutsche Korpuslinguistik	8
1.2 Korpus-Tools	11
1.2.1 AntConc	12
1.2.2 Compleat Lexical Tutor	12
1.2.3 WordSmith Tools	12
1.3 Wie "baue" ich mein eigenes Korpus?.....	12
1.3.1 Textdaten	12
1.3.2 Audio-Daten/Video-Daten	13
1.4 Arbeiten mit Korpora.....	13
1.4.1 Wahl des Korpus	13
1.4.2 Die Nutzung von Konkordanzprogrammen	14
1.4.3 Funktionen von Konkordanzprogrammen	15
1.4.4 Kriterien zur Auswahl von Korpora	18
1.5 Anwendungsmöglichkeiten von Korpora	18
1.5.1 Soziolinguistische Studien	19
1.5.2 Studien zu Varietäten	21
1.5.3 Analyse von phraseologischen Routinen	22
1.5.4 Konkordanzanalysen mit dem Compleat Lexical Tutor (CLT)	23
1.5.5 Erstellung von spezifischen Vokabellisten	24
1.5.6 Analyse des eigenen Stils	24
1.5.7 Korpora und Literaturanalysen	25
1.5.8 Korpora und Lernalterssprache	27
2 Lernerkorpora.....	27
3 Korpora im Fremdsprachenunterricht	30
3.1 Die geschichtliche Entwicklung von Korpora im Sprachunterricht .	30
3.2 Data-Driven Learning (DDL)	31
3.2.1 Definition	31
3.2.2 Ausprägungen des DDL	31
3.2.3 Der Einfluss der Korpuslinguistik auf den Sprachunterricht	32
3.3 Einsatz von Korpora bei der Unterrichtsvorbereitung und im (DaF-) Unterricht	33
3.3.1 Beiträge zu Rolle und Möglichkeiten des Einsatzes von Korpora im Fremdsprachenunterricht	33
3.3.2 Beiträge zum Einsatz von Korpora bei der Vermittlung von grammatischen, phonetischen und Wortschatzkenntnissen	44

3.3.3	Skizzen von Unterrichtseinheiten auf Basis von Korpora	52
4	Nutzung sprachlicher Ressourcen im ICALL	54
4.1	Phasen des CALL: Einfluss von Lerntheorien	54
4.1.1	Behavioristisches CALL	55
4.1.2	Kommunikatives CALL	56
4.1.3	Konstruktivistisches CALL	56
4.2	Qualitätssicherung von Software	56
4.2.1	ISO/IEC 9126-1	57
4.2.2	ISO 13407	57
4.2.3	Qualitätssicherung von CALL-Anwendungen	58
4.2.4	Evaluationsdatenbanken	59
4.3	Parser-basiertes CALL	60
4.4	Lexikalische Ressourcen und Korpora in CALL	62
4.4.1	Lexikalische Ressourcen in CALL	63
4.4.2	Korpora in CALL	64
4.5	ICALL-Systeme im Überblick	67
4.5.1	BRIDGE Tutor und MILT (1995, im Einsatz an der ARI)	67
4.5.2	etutor (seit 1997 in ständiger Entwicklung, kommerziell)	68
4.5.3	PromisD (2003, Prototyp)	68
4.5.4	DiBEx (seit 1999, Prototyp)	69
4.5.5	NN-System (2007, Zürich, Prototyp)	69
4.5.6	e-GramD (2007, L1 Primarstufe, Prototyp)	69
4.5.7	Ein Tagger-basiertes Studentenmodul (frei zugänglich)	70
4.5.8	ProGram (1999-2007) und ProGram-iK (2003-2006, aktiv im Einsatz)	70
4.5.9	ALLES (EU-Projekt 2002-2005, aktiv im Einsatz)	71
4.5.10	DUO (Deutsch-Uni Online, kommerziell)	72
4.6	Weitere natürlichsprachliche Anwendungen	73
4.6.1	CLAT (aktiv im Einsatz)	73
4.6.2	Metis I (EU-Projekt 2002-2003) und Metis II (EU-Projekt 2004-2007)	74
4.6.3	Google Übersetzer (aktiv im Einsatz)	75
4.7	Sprachliche Ressourcen und Tools im Überblick	76
4.7.1	DWDS (DFG-Projekt 2000-2003)	76
4.7.2	Fehler-Projekt	77
4.7.3	FALKO	78
4.7.4	HaGenLex	78
4.7.5	MPRO (Morphologie-PROgramm)	78
4.7.6	COSMAS-II	79
4.8	Fazit	79
5	Szenarien für den Einsatz von Sprachressourcen	80
5.1	Erstellung von Korpora für fremdsprachendidaktische Zwecke	80
5.2	Fremdsprachendidaktische Nutzung von Lernerkorpora	80
5.3	Fremdsprachendidaktische Nutzung von allgemeinen Sprachkorpora	80
5.3.1	Lerner80	
5.3.2	Lehrende	80
5.3.3	Lehrwerk-Autoren	80
5.4	Nutzung von Sprachressourcen bei der Entwicklung von CALL-Systemen	80
6	Literatur	80
Internetquellen		87

Appendix 2: Guidelines for the in-depth interviews

Einleitungsteil

Könnten Sie kurz schildern, welches Ihr Interesse an einem elektronischen Arbeitsplatz der Zukunft ist?

Schildern Sie kurz typische Forschungsfragen, denen Sie in ihrer wissenschaftlichen Arbeit nachgehen und bei denen Sie auf die Nutzung von elektronischen Textsammlungen, linguistischer Korpora *oder anderer Sprachressourcen (Wörterbücher etc.)* angewiesen sind

- Wenn eine Forschungsfrage genannt wird: Nachfrage: Wie würden Sie dabei (bei der Forschungsfrage) methodisch vorgehen?
- Sehen Sie bei den von Ihnen vorgeschlagenen Methoden generalisierbare Aspekte (z.B. wäre das für Kollegen von Ihnen auch anwendbar)?

Führen Sie Lehrveranstaltungen durch? **Wenn ja:** Welche Fragenstellungen werden in Ihren Lehrveranstaltungen typischerweise behandelt, bei deren Bearbeitung elektronische Textsammlungen oder linguistische Korpora von Nutzen (*Ergänzung wie oben*) sind?

Gezielte Einzelfragen:

Welche technischen Geräte und welche Computerprogramme nutzen Sie in diesem Zusammenhang?

Web-Applikationen

Spezialisierte Software: Konkordanzprogramme, Editionsprogramme

Editoren

Betriebssystem(e)

Haben Sie einen guten Überblick über Ressourcen / Werkzeuge, die für Ihre Forschungen relevant sind bzw. sein könnten?

Frage zur Annotation von Datensammlungen:

- a) Müssen Sie für Ihre Forschungsfragen Datensammlungen annotieren?
- b) Nutzen Sie annotierte Datensammlungen?
- c) Stellen Sie Ihre Eigenannotationen Kollegen zur Verfügung oder könnten Sie sich vorstellen dies zu tun
- d) *Welche Garantien benötigen Sie, um Ihre Daten zu teilen? Bestehen Bedenken, dass andere sich mit Ihren Daten/Ressourcen bzw. daraus gewonnenen Forschungsergebnissen „schmücken“?*

Ist Zusammenarbeit mit Fachkollegen auf einer gemeinsamen Datenbasis

- a) in Ihrem Gebiet verbreitet?
- b) praktizieren Sie diese selbst?
- c) Gibt es Fragestellungen, die ohne diese Zusammenarbeit gar nicht lösbar wären?
- d) Sehen Sie Potenzial für die interdisziplinäre Zusammenarbeit, d.h. über den Kreis Ihrer Fachkollegen hinaus

[Hintergrund: beispielsweise Arbeit an einer großen Edition, Text-, Bildsammlungen]

Können Sie sich erinnern, schon einmal an eine technisch bedingte Grenze bei der Bearbeitung einer Forschungsfrage gestoßen zu sein? [Hintergrund: zum Beispiel: Daten lagen in einem nicht

nutzbaren Format vor, notwendige automatische Bearbeitungsschritte konnten aus verschiedenen Gründen nicht durchgeführt werden,...]

- Konnten Sie solche technischen Probleme lösen bzw. lösen lassen? Wie haben Sie das getan?

- Haben Sie aus solchen Gründen heraus schon einmal auf die Bearbeitung einer Fragestellung ganz verzichtet?

- Kennen Sie Kollegen, die vor solchen Hürden gestanden haben? Was haben Ihre Kollegen daraufhin getan?

Prototyp

Können Sie sich vorstellen, den Prototypen für Ihre Forschungsfragen zu nutzen?

Was gefällt Ihnen gut?

Was könnte man verbessern bzw. wäre unbedingt nötig?

[Prototyp ist Hilfsmittel, um den Interviewten dazu zu inspirieren, seine Gedanken über seinen idealen wissenschaftlichen Arbeitsplatz zu äußern: Tut er dies nicht über die og. Fragen, kann man folgende Fragen nachschieben:

- Welche technischen Geräte und Computerprogramme wären unverzichtbar?

- Welche technischen Geräte und Computerprogramme würden Sie sich wünschen? *(Vielleicht etwas weniger technisch? Arbeitsmittel? Arbeitsgeräte? Programme? Ressourcen?)*

- Welche für Ihre Untersuchungen nützlichen Werkzeuge fehlen?

- Welche Werkzeuge und Daten sollten Ihnen, in einer idealen Welt, zur Verfügung stehen?

- Gibt es typische wissenschaftliche Arbeitsschritte, bei denen Sie sich vorstellen könnten, diese zu automatisieren? *Inwiefern würde dies Ihre Arbeit erleichtern?*

Mit wem, denken Sie, sollten wir über diese Fragen ebenfalls sprechen?