

Überlegungen zu einem Betriebsmodell für etablierte wissenschaftliche Software

Thomas Schmidt (IDS Mannheim) &
Kai Wörner (Universität Hamburg)



Überlegungen zu einem Betriebsmodell für etablierte wissenschaftliche Software

- CFP: „Wer soll das bezahlen?“
- Arbeitstitel:
 - „(Re)kapitulation: Vom Ende und möglichen Neuanfang der Entwicklung einer wissenschaftlichen Software“
 - „Scheiße, es funktioniert“
- „Das angesprochene Problem ist ein sehr ernstes, denn zu viele gute wissenschaftliche Software verschwindet oder wird unbenutzbar, weil kein geeignetes Finanzierungsmodell gefunden werden konnte.“
[Kommentar eines Gutachters]
- Dieser Beitrag: keine Lösung, Fallstudie, erste Ideen



EXMARaLDA

- Erstellen, Verwalten, Auswerten, Publizieren audiovisueller Sprachkorpora
 - Gesprächskorpora
 - Spracherwerbsdaten
 - Soziolinguistische/Dialektologische Daten
- Transkription und Annotation (Partitur-Editor)
- Korpus-Management / Metadaten (Coma)
- Query/Analyse (EXAKT)

The screenshot shows the EXMARaLDA EXAKT 1.1 interface. The main window displays search results for 'DEMO-KORPUS' with a search filter 'ib(der|de|das|b)'. A table lists communication events with columns for #, Communication, Speaker, Left Context, Match, Right Context, Remarks, and Beruf[S]. Below the table, a detailed view of a selected event shows the original audio transcription and its translation into English.

#	Communication	Speaker	Left Context	Match	Right Context	Remarks	Beruf[S]
18	Rudi Völler: Wutausbruch	RV	! Wir ham heute .. Und da hat	der	Sebastian Kehl .. Wir ham + genei		Teamchef
19	Rudi Völler: Wutausbruch	RV	.. Wir sind Vize Weltmeister ..	das	muss n bisschen mehr kommen +	verz. enclitd.	Teamchef
20	Rudi Völler: Wutausbruch	RV	en .. Aber diesen Scheiß ..	der	da immer tabellar wird + Da so	Tea KWIC concordance	Teamchef
21	Rudi Völler: Wutausbruch	RV	irklich mal Gedanken machen, ob	das	/ + ob wir in der Zukunft so wei	wrong case	Teamchef
22	Rudi Völler: Wutausbruch	RV	sien machen, ob das + ob wir in	der	Zukunft so weitermachen können?		Teamchef
23	Rudi Völler: Wutausbruch	RV	los runter zu ziehen? - Das ist	das	Alterstetel + Und ich lass mir		Teamchef
24	Rudi Völler: Wutausbruch	RV	Alterstetel? - Und ich lass mir	das	nicht mehr so lange gefalten. D		Teamchef
25	Rudi Völler: Wutausbruch	RV	.. Ach, wieso müssen wir denn?	das	Mannschaften klar bezeichnen?		Teamchef
26	Rudi Völler: Wutausbruch	RV	ich hab doch ich hab doch	das	Mannschaft .. Neint ich hab doch		Teamchef
27	Rudi Völler: Wutausbruch	RV	Mannschaft .. Neint ich hab doch	das	Mannschaft ich hab doch die Ma		Teamchef
28	Rudi Völler: Wutausbruch	RV	ich hab doch ich hab doch	das	ich hab doch ich hab doch	repar. seq. unrec.	Teamchef
29	Rudi Völler: Wutausbruch	RV	ich hab doch ich hab doch	das	ich hab doch ich hab doch		Teamchef
30	Rudi Völler: Wutausbruch	RV	iden nur die Leute spielen ..	das	sich wirklich hundertprozentig		Teamchef
31	Rudi Völler: Wutausbruch	RV	st doch tuhet! Der Günter! Was	das	früher für n Scheiß gespielt ha		Teamchef

The screenshot shows the EXMARaLDA Coma interface. The main window displays a communication event titled 'Communication Rudi Völler: Wutausbruch'. The event details include the description, recording date, type, source, and transcription information. A list of participants is shown on the right side of the window.

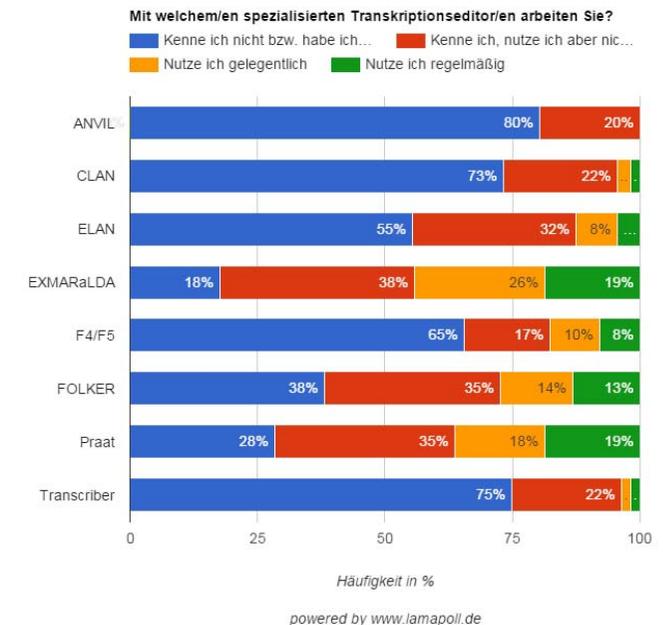
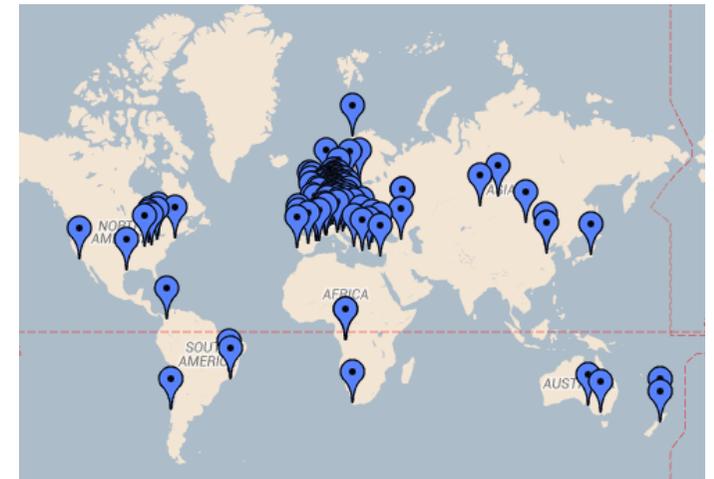
Name	Transkription erstellt in
Rudi Völler: Wutausbruch	Januar 2004
Hilge Schneider: Arbeit...	?
Huber Fichte: Interview	November 2003
Hilge Schneider: Tropf...	?
Studio Braun: English T...	August 2008
Deutsche Bahn: Rossau	August 2008
Studio Braun: Forum Wa...	Oktober 2008
Sigalène Royal: Débat	November 2008
Fernando Savater: Her...	November 2008
Paul McCartney: Interview	Dezember 2001
Seif Iasi: Experimentauk...	November 2008
near story	April 2009
Henry Fröhlich: My Theory	Januar 2009
Hart aber fair	Juli 2009
Anne Will: Halbes Wahr...	Dezember 2009
Goepferle: Cello	Januar 2010
Nyuyen Ngoc Ngan	Januar 2010
Tella-Telenor	September 2000

The screenshot shows the EXMARaLDA Partitur-Editor 1.4 interface. The main window displays a transcription and annotation of an audio segment. The transcription is shown in a table format with columns for time intervals and the corresponding text. The annotation is shown in a table format with columns for time intervals and the corresponding phonetic or linguistic annotations.

Time Interval	Transcription	Annotation
0 [00.0]	And then you see ehm a maan in maybe	
1 [01.9]	louder	
2 [01.9]	louder	
3 [03.2]	And then you see ehm a maan in maybe	his fittles. ((takes breath 0,5s))
4	And then you see ehm a maan in maybe	his fittles. ((takes breath 0,5s))
5 [09.2]	And so h	
6 [10.4]	And so h	

Verbreitung / Verwendung

- 500-1000 Downloads pro Monat, seit vielen Jahren
- 100-200 Support-Anfragen jährlich über CLARIN-Helpdesk
- Nutzerstudie „Mündliche Korpora“: 19% nutzen EXMARaLDA regelmäßig, 26% gelegentlich, nur 18% „kenne ich nicht / habe ich noch nie genutzt“
- Basis mehrerer großer Korpora
- Regelmäßige Anfragen für Schulungen, von vielen verschiedenen deutschen Universitäten, aus 15 verschiedenen Ländern
- Als Best-Practice empfohlen in DFG-Handreichungen



Wer hat es bisher bezahlt?

Finanzierung / Projektzusammenhänge

- 2000-2011: SFB 538 Mehrsprachigkeit
- ab 2011: HZSK Hamburg (CLARIN/LiS) / IDS Mannheim (AGD)
- Mischfinanzierung: Universität Hamburg, Qualitätsoffensive der Stadt Hamburg, DFG-SFB, DFG-LiS, BMBF-CLARIN, Leibniz
- Software-Entwicklung immer als Mittel zum Zweck, ein Arbeitspaket unter mehreren
 - Computergestützte Erfassungs- und Analysemethoden multilingualer Daten (SFB)
 - Etablierung eines Schwerpunkts Mehrsprachigkeit und gesprochene Sprache am HZSK (LiS)
 - Archiv für Gesprochenes Deutsch (IDS)
 - Eine web- und zentrenbasierte Forschungsinfrastruktur für die Geistes- und Sozialwissenschaften (CLARIN)

Sonderforschungsbereich
Mehrsprachigkeit 

 INSTITUT FÜR
DEUTSCHE SPRACHE

 Bundesministerium
für Bildung
und Forschung



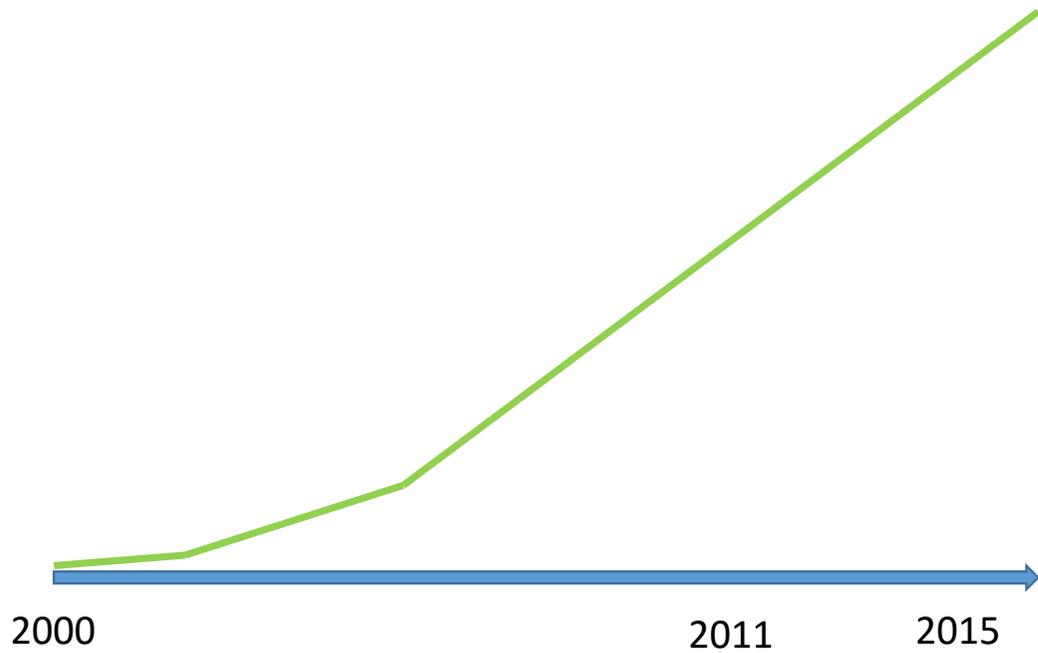



CLARIN-D

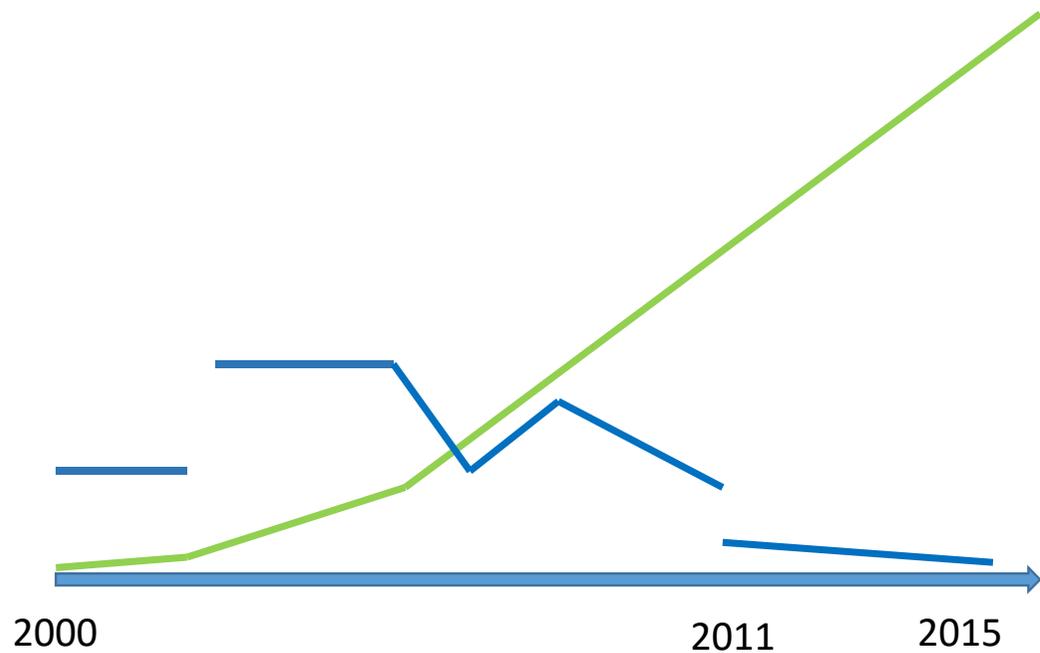
Nutzung/Entwicklung

Anzahl der Nutzer

- 2 zu Beginn
- Hunderte nach ca. 5 Jahren (Basisfunktionalität)
- Tausende aktuell



Nutzung/Entwicklung



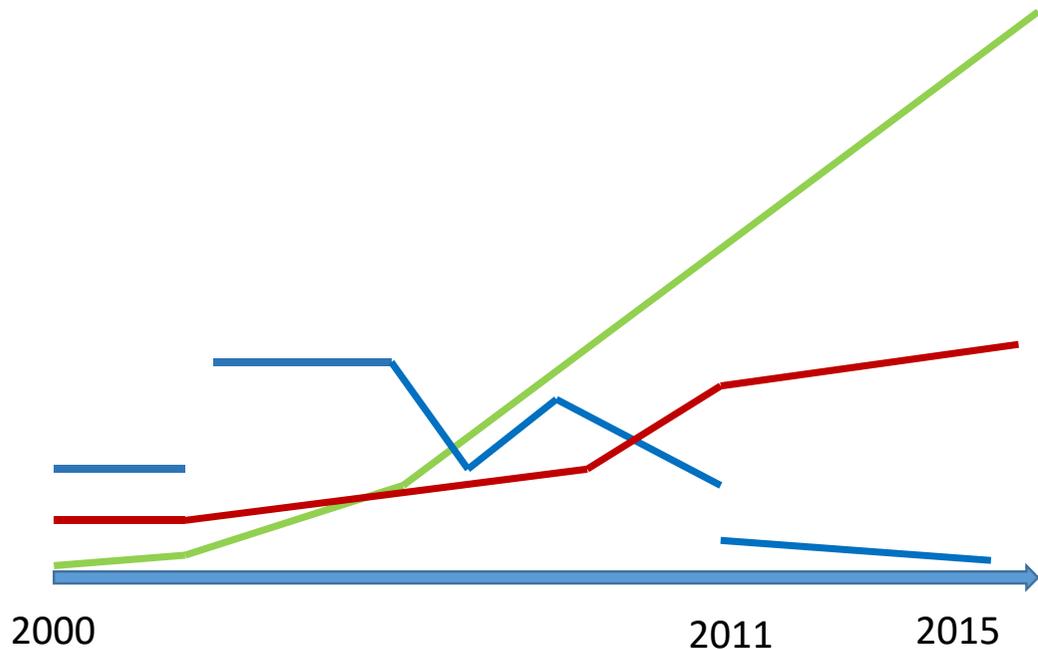
Anzahl der Nutzer

- 2 zu Beginn
- Hunderte nach ca. 5 Jahren (Basisfunktionalität)
- Tausende aktuell

Wöchentlich investierte Entwicklungszeit (inkl. Maintenance)

- ca. 30h zu Beginn (initiales Design, Prototypen, 1 Person)
- ca. 60h 2003-2005 (Ausbau, 2 Personen)
- Danach weniger, schwankend, je nach aktueller Priorität im Projekt (Datenkuration im SFB, bis zu 4 Personen)
- Seit SFB-Ende deutlich weniger, aktuell ca. 1h
- Regelmäßige Updates bis 2011, seitdem nur noch sporadisch/"inoffiziell"/auf Zuruf

Nutzung/Entwicklung



Anzahl der Nutzer

- 2 zu Beginn
- Hunderte nach ca. 5 Jahren (Basisfunktionalität)
- Tausende aktuell

Wöchentlich investierte Entwicklungszeit (inkl. Maintenance)

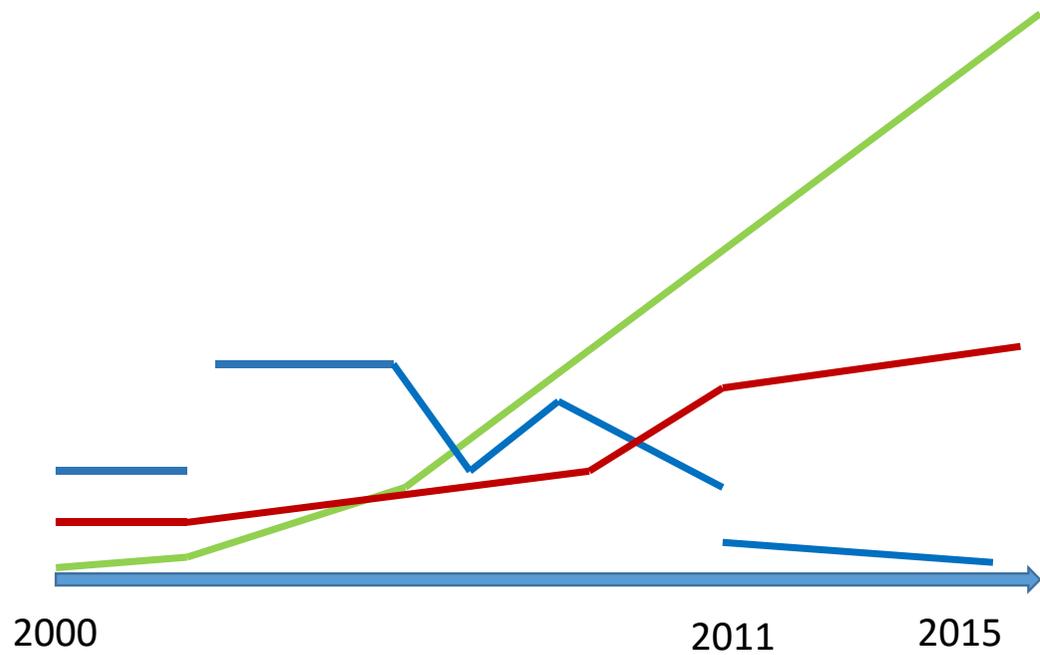
- ca. 30h zu Beginn (initiales Design, Prototypen, 1 Person)
- ca. 60h 2003-2005 (Ausbau, 2 Personen)
- Danach weniger, schwankend, je nach aktueller Priorität im Projekt (Datenkuration im SFB, bis zu 4 Personen)
- Seit SFB-Ende deutlich weniger, aktuell ca. 1h
- Regelmäßige Updates bis 2011, seitdem nur noch sporadisch/"inoffiziell"/auf Zuruf

Supportbedarf/-aufwand

- Individuelle Betreuung zu Beginn
- Wachsende Zahl an E-Mail Support-Anfragen
- Schulungen seit 2008
- CLARIN-Helpdesk seit 2011

Nutzung/Entwicklung

- ❖ Wachsende Nutzerzahlen
- ❖ Anhaltender Supportbedarf
- ❖ Schwindende Investition in Entwicklung/Maintenance



Aktuelle Aufgaben

Maintenance

- Anpassung an neue OS
- Unterstützung neuer Videoformate
- Touchscreens
- Umstieg SourceForge > Git
- Bugs

...

Ausbau (klein)

- ✓ neue Importfilter
- ✓ Usability-Adaptionen

...

Ausbau (groß)

- Anbindung an CLARIN-Webservices (WebLicht)
- Kompatibilität mit ISO/TEI Standard
- Neue Annotationsmethoden
- Neue Datentypen
- Übertragung in webbasierte Architekturen

...

Dokumentation / Support / „Marketing“

- Schulungen / Email-Anfragen
- Nachführen von Manuals etc.
- Internationalisierung
- Präsentation auf Konferenzen / Fachmessen

...

- Kleinere Feature Requests können noch relativ problemlos umgesetzt werden
- Innovative Entwicklungen nur noch schleppend oder gar nicht
- Erhebliche Verspätung bei kritischen Anpassungen
- Dokumentation hinkt immer weiter hinterher
- „Service“-Qualität lässt nach

Weiter so?

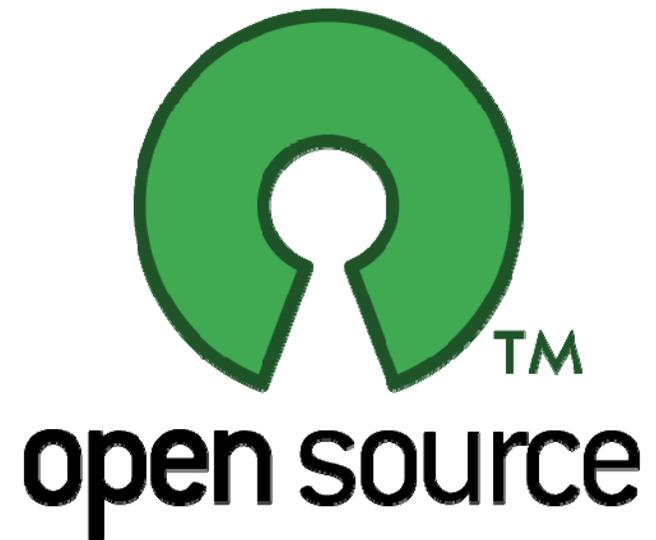
- Zunehmende Instabilität
 - Erhöhter Support-Bedarf bei
 - Ausbleiben von Anpassungen an Neuerungen
 - Zu langsamer Nachführung der Dokumentation
 - Nachlassende „Kunden“-Zufriedenheit
- ➔ Kapitulation: Abandonware (vgl. Transcriber)

- Bedeutung für universitäre Ausbildung (Curricula), laufende Forschung
- Bedeutung für einschlägige Datenzentren, Infrastrukturen, Best Practices, Standardisierungsprozesse
- Situation ähnlich für „Konkurrenten“ (ELAN)



Open Source?

- EXMARaLDA ist OpenSource
- ca. 10 punktuelle Beiträge innerhalb von 10 Jahren
- Organisch gewachsene Code-Basis
 - JDK 1.2 / kein DOM-Parser für XML in 2000
 - keine Refaktorisierung in der Zwischenzeit
- Wer sollte beitragen???

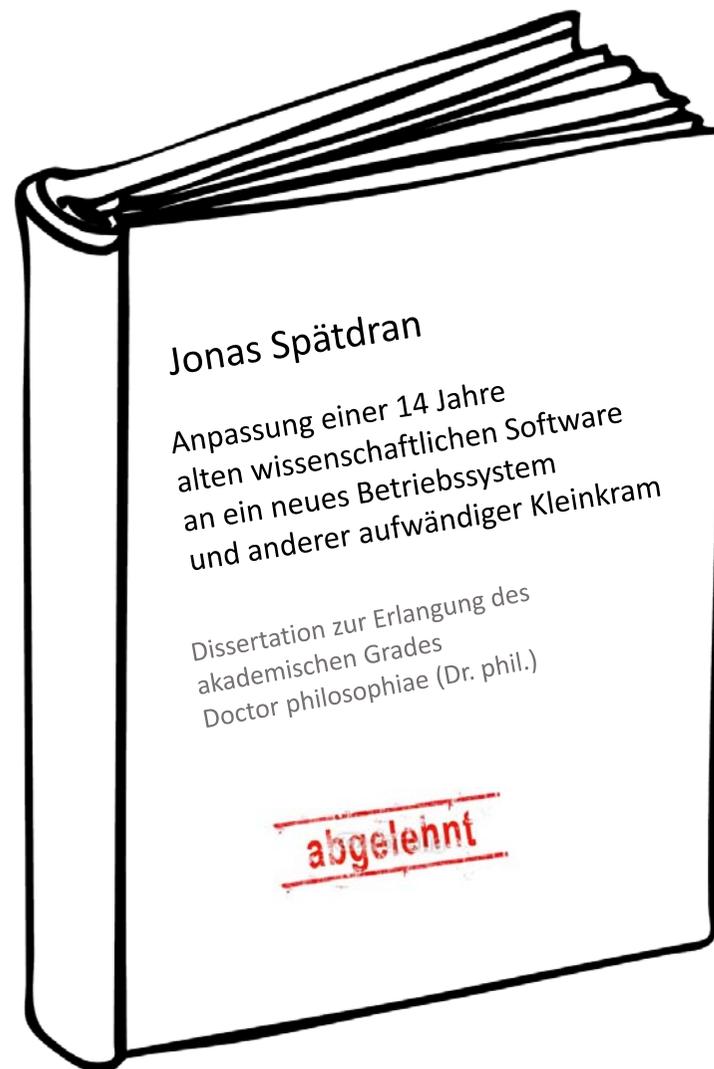
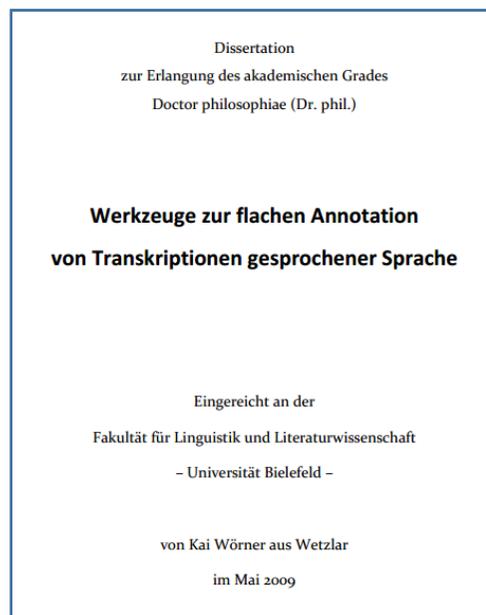
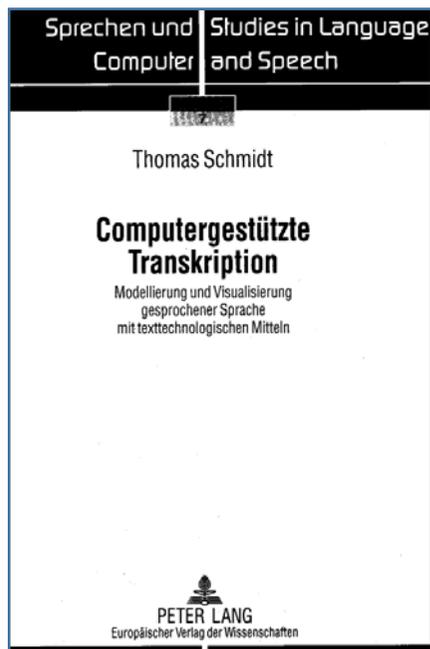


Neuer Drittmittelantrag?

- Keine dedizierten Förderprogramme für Entwicklung wissenschaftlicher Software
 - Softwareentwicklung muss im Antrag hinter anderer Fragestellung versteckt werden
- Maintenance ist keine Projektaufgabe
- Support ist keine Projektaufgabe
 - Dauerhafter Bedarf, nicht gut planbar
- Software-Entwicklung (in diesem Stadium) keine wissenschaftliche Tätigkeit (aber auch keine HiWi-Tätigkeit)
 - Entweder innovative Entwicklung oder Integration in laufendes System
 - Kaum geeignete KandidatInnen für Projektstellen



Bundesministerium
für Bildung
und Forschung



Kostenpflichtige Lizenzierung?

vgl. Qualitative Data Analysis

- MaxQDA – standard nicht-kommerzielle Lizenz, ca. 650 EURO
- atlas.ti, f4analyse, nvivo,
- Transana, Einzellizenz \$350

vgl. einfache („Consumer“) - Transkriptionsprogramme

- z.B. InqScribe - Voll-Lizenz \$99

Beispiele aus der Linguistik?

- Word Sketch Engine - £58 pro Jahr
- monoConc - \$85/2 Jahre
- wordSmith - £50

Marktanalyse, Business Plan, Lizenzmodell, Ausgründung

- ➔ Hohe Anfangsinvestition
- ➔ Rechtlich zulässig (jahrelange öffentliche Förderung)?
- ➔ Akzeptanz?
- ➔ Kunden außerhalb des akademischen „Marktes“?

„Kommerziell werden“

- ➔ Interoperabilität kann geschäftsschädigend sein
- ➔ Open Access-Idee in digitalen Infrastrukturen?
- ➔ Förderfähigkeit für Innovationen?

Das ideale Betriebsmodell...

- ✓ ... erlaubt weiterhin eine **kostenlose Distribution** der Software
- ✓ ... erlaubt weiterhin Entscheidungen nach Kriterien der beteiligten **wissenschaftlichen Communities**
- ✓ ... lässt einer **Leitung** ausreichend Raum für kurz-, mittel- und langfristige **Planungen**
- ✓ ... generiert finanzielle Mittel **bedarfsgerecht**
- ✓ ... generiert anteilig finanzielle Mittel für eine „**Grundfinanzierung**“
- ✓ ... erlaubt es, **Software-Entwickler** als solche zu bezahlen
- ✓ ... erlaubt es, **Support-Mitarbeiter** als solche zu bezahlen

Elemente

- ✓ Kostenpflichtige **Schulungen** für Lehrende, Forscherteams oder im Rahmen universitärer Lehrveranstaltungen
- ✓ Kostenpflichtige Unterstützung bei der **Aufbereitung (Kuration)** existierender Datenbestände für die Arbeit mit EXMARaLDA
- **Support-Verträge**
 - für **Fachbereiche**, die EXMARaLDA regelmäßig in der Lehre einsetzen
 - für **Forschungsprojekte**, die ihre Datengrundlage mit EXMARaLDA bearbeiten
 - für einzelne **Studierende und Promovierende**
- ✓ Partnerschaften mit **Forschungsprojekten**, die für ihre Arbeit spezielle **Erweiterungen** benötigen
- Partnerschaften mit **Fachverbänden** u.Ä., deren Mitglieder ein Interesse an der Weiterentwicklung von EXMARaLDA haben
- **Micro-Payment** auf freiwilliger Basis für einzelne Support-Fälle;
- **Crowd-Funding** für die Umsetzung aufwändiger Features, die für einen größeren Nutzerkreis von Interesse sind.
- **Eigene Drittmittelprojekte** für innovative Erweiterungen (bis Beta, danach Integration in übriges Betriebsmodell)

hzsk services

› hzsk services

› [Datenmanagement](#)

› [Schulungen](#)

› [Transkriptionsservice](#)

› [Mitarbeiter](#)

› [Referenzen](#)

› [Kontakt](#)

› [Impressum](#)

Linguistisches Datenmanagement für Forschung und Unternehmen

hzsk services bietet Beratung und technischen Support im Bereich des Datenmanagements, der Datenverarbeitung, -aufbereitung und -auswertung sowie der Erstellung und Verarbeitung sprachlicher Datenressourcen, insbesondere auch mündlicher Sprachkorpora, an. In individuellen Schulungen und Workshops vermitteln unsere Mitarbeiter methodologisches Fachwissen und technische Fertigkeiten für die erfolgreiche Durchführung von Forschungs- und Entwicklungsprojekten.

hzsk services ist ein an das [Hamburger Zentrum für Sprachkorpora \(HZSK\)](#) der Universität Hamburg angegliedertes, auf Wissens- und Technologietransfer spezialisiertes **Kompetenzzentrum** der [Hamburg Innovation GmbH](#).

Die Leistungen werden in der Regel als Gegenstand eines Werkvertrages, eines Forschungs- und Entwicklungsauftrages oder auf Honorarbasis erbracht.

2015 hzsk-services.de- This is a Free Drupal Theme

Ported to Drupal for the Open Source Community by [Drupalizing](#), a Project of [More than \(just\) Themes](#). Original design by [Simple Themes](#).

<http://hzsk-services.de/>

Preisschilder

EXMARaLDA
Download

0 EURO

1-tägige
EXMARaLDA
Schulung

500 EURO
+ RK

2-tägige
EXMARaLDA
Schulung

800 EURO
+ RK

projektspezifische
Erweiterung

50 EURO
Stundensatz

Support
Kontaktaufnahme

0 EURO

Support-Lizenz
Forschungsprojekt

? EURO

Support-Lizenz
Fachbereich

? EURO

Kuration 20h
komplexes
Videokorpus

15.000 EURO

Offene Fragen / Ungelöste Probleme

- Wie kommunizieren?
 - „Gestern noch 0 EURO, heute schon 800 EURO“
 - „Wir können nicht helfen, Sie haben nichts bezahlt“
- Anschubfinanzierung?
 - initiale Konzeption
 - Akquise
- Zielgröße?
- Rechtsform?

gwin

DFG

hzsk

Sonderforschungsbereich
Mehrsprachigkeit



INSTITUT FÜR
DEUTSCHE SPRACHE



CLARIN-D



Bundesministerium
für Bildung
und Forschung