nestor

CLARIN
Common Language Resources and Technology Infrastructure

Forschungsdaten in den Digital Humanities - Der Verbund nestor und Datenzentren am Institut für Deutsche Sprache und an der Universität zu Köln

Andreas Witt

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

IDS INSTITUT FÜR DEUTSCHE SPRACHE
Mitglied der Leibniz-Gemeinschaft

Gliederung

- nestor – Verbund
  - nestor: Partner und Arbeitsgruppen
  - die AG Zertifizierung
- Forschungsdaten in den Digital Humanities
  - Institut für Deutsche Sprache Mannheim
  - Das Data Center for the Humanities der Universität zu Köln
- Was können Infrastruktureinrichtungen leisten
  - Herausforderungen
  - Langfristigkeit
- Wie helfen Verbünde, wie nestor, CLARIN, DARIAH

VERANSTALTUNGEN
nestor-Praktikertag
Workshops
Vorträge

DIN-Normen

nestor-Schools
nestor-Handbuch

nestor

INTERNATIONALE KOOPERATION

DPC
NCDD
NDIIPP
EU-Projekte

PUBLIKATIONEN
Kriterienkatalog
Handbuch
Ratgeber
Expertisen
Infosheets
nestor edition
Newsletter

INFORMATIONSANGEBOTE
News
Veranstaltungskalender
Personen
Projekte

ARBEITSGRUPPEN z.B.
AG Elektronische Akte
AG Emulation
AG Formaterkennung
AG Forschungsdaten
AG Media
AG Zertifizierung

nestor - Partner

Landesarchiv
Nordrhein-Westfalen

nestor - Arbeitsgruppen

- AG Digitale Bestandserhaltung
- AG Elektronische Akte
- AG Emulation / GI FG Langzeitarchivierung
- AG Formaterkennung
- AG Forschungsdaten
- AG Kooperation der Archive
- AG A-V Medien
- AG OAIS Review
- AG Personal Digital Archiving
- AG Zertifizierung

nestor - AG Zertifizierung

- Auf der Basis der DIN 31644: "Kriterien für vertrauenswürdige digitale Langzeitarchive", hat die AG ein Zertifizierungsverfahren entworfen und betreut die Durchführung

- Aufgaben
  - Zertifizierung digitaler Archive
  - Fortentwicklung des nestor-Siegels

Prämissen zur Vertrauenswürdigkeit

- Ein *digitales Langzeitarchiv* ist **vertrauenswürdig**:
  - wenn es gemäß seinen Zielen und Spezifikationen zum Informationserhalt über lange Zeiträume hinweg operiert
  - seine Nutzer, Produzenten, Betreiber, Partner ihm dieses zutrauen
- Digitale Archive sind derzeit (noch) nicht per se vertrauenswürdig
- Nutzer können nur in Einzelfällen die Vertrauenswürdigkeit überprüfen

Nationale Entwicklung

- 2004/5 Gründung der nestor-AG „Vertrauenswürdige Archive“
- 2006: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Version 1
- 2008: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Version 2

- 2008: Kriterienkatalog wird im NABD 15 des DIN zu einem Normungsprojekt
- 2012: Veröffentlichung der DIN 31644

nestor-Siegel: Verfahren

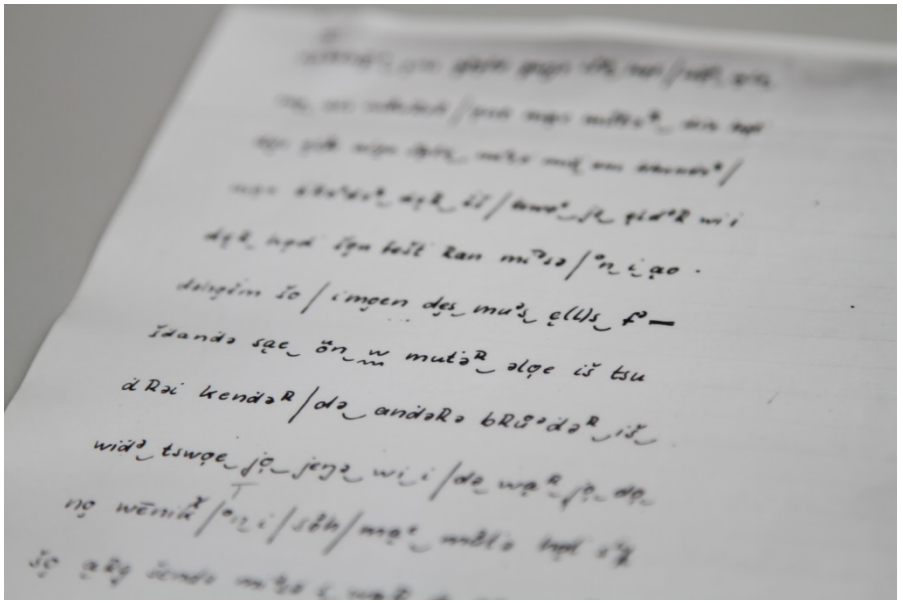
- Orientierung an DIN 31644
- Kein akkreditiertes Zertifizierungsverfahren
- „erweiterte Selbstevaluierung“ durch die nestor-AG Zertifizierung (2 Gutachter)
- Eingebettet in internationales MoU mit Data Seal of Approval und ISO 16363: http://www.trusteddigitalrepository.eu/Trusted%20Digital%20Repository.html

Das Institut für Deutsche Sprache

- Sitz in Mannheim
- Gründung 1964
- Außeruniversitäre Forschungseinrichtung
- Mitglied der Leibniz Gemeinschaft
- Vernetzung
  - Betrieb eines CLARIN-B Zentrums
  - Mitglied in nestor
  - Sitz des Rats für Deutsche Rechtschreibung

Beispiele aus der AGD am IDS

- Transkriptionen auf Papier



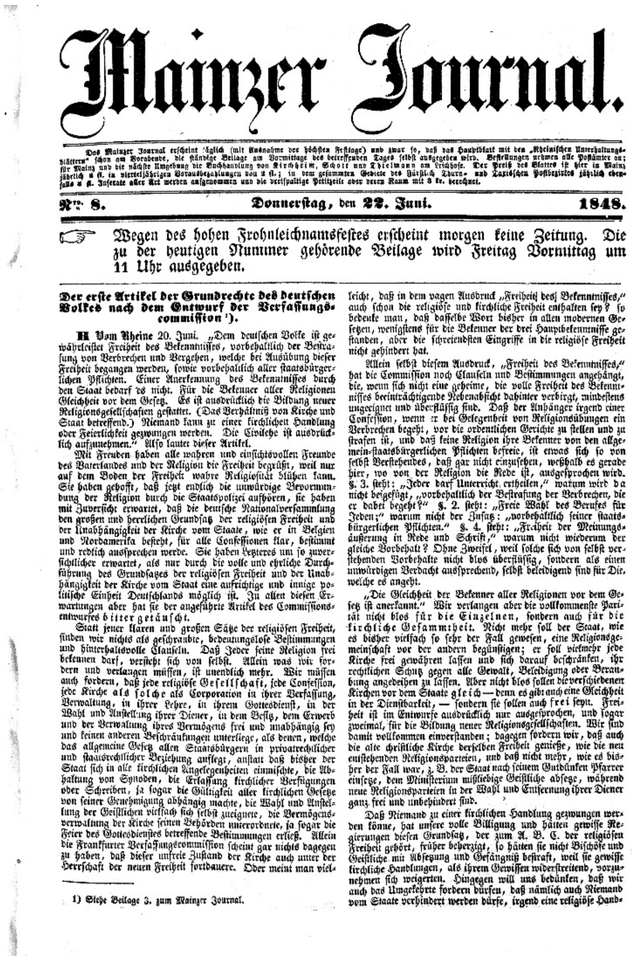
Beispiele aus der AGD am IDS

- Tonbänder und Kassetten

Zwei Beispiele für Kurationsprojekte am IDS

- Mannheimer Korpus Historischer Zeitungen und Zeitschriften (mkhz)
  - Konvertierung von TUSTEP nach TEI
  - COSMAS II
- Monash Corpus of Australien German
  - Daten von Michael Clyne
  - DGD
  - LZA

Mannheimer Korpus Historischer Zeitungen und Zeitschriften



Mainzer Journal.

Das Mainzer Journal erscheint täglich (mit Ausnahme der höchsten Festtage) und zwar so, daß das Hauptblatt mit den „Rheinischen Unterhaltungsblättern" schon am Vorabende, die ständige Beilage am Vormittage des betreffenden Tages selbst ausgegeben wird. Bestellungen nehmen alle Postämter an; für Mainz und die nächste Umgebung die Buchhandlung von Kirchheim, Schott und Thielmann am Leichhofe. Der Preis des Blattes ist hier in Mainz jährlich 8 fl. in vierteljährigen Vorausbezahlungen von 2 fl.; in dem gesammten Gebiete des Fürstlich Thurn- und Taxis'schen Postbezirkes jährlich ebenfalls 8 fl. Inserate aller Art werden angenommen und die dreispaltige Petitzeile oder deren Raum mit 3 kr. berechnet.

Nro. 8. Donnerstag, den 22. Juni. 1848.

Wegen des hohen Frohnleichnamsfestes erscheint morgen keine Zeitung. Die zu der heutigen Nummer gehörende Beilage wird Freitag Vormittag um 11 Uhr ausgegeben.

Der erste Artikel der Grundrechte des deutschen Volkes nach dem Entwurf der Verfassungscommission [1]).

Vom Rheine 20. Juni. „Dem deutschen Volke ist gewährleistet Freiheit des Bekenntnisses, vorbehaltlich der Bestrafung von Verbrechen und Vergehen, welche bei Ausübung dieser Freiheit begangen werden, sowie vorbehaltlich aller staatsbürgerlichen Pflichten. Einer Anerkennung des Bekenntnisses durch den Staat bedarf es nicht. Für die Bekenner aller Religionen Gleichheit vor dem Gesetz. Es ist ausdrücklich die Bildung neuer Religionsgesellschaften gestattet. (Das Verhältniß von Kirche und Staat betreffend.) Niemand kann zu einer kirchlichen Handlung oder Feierlichkeit gezwungen werden. Die Civilehe ist ausdrücklich aufzunehmen." Also lautet dieser Artikel.

Mit Freuden haben alle wahren und einsichtsvollen Freunde des Vaterlandes und der Religion die Freiheit begrüßt, weil nur auf dem Boden der Freiheit wahre Religiosität blühen kann. Sie haben gehofft, daß jetzt endlich die unwürdige Bevormundung der Religion durch die Staatspolizei aufhören, sie haben mit Zuversicht erwartet, daß die deutsche Nationalversammlung den großen und herrlichen Grundsatz der religiösen Freiheit und der Unabhängigkeit der Kirche vom Staate, wie er in Belgien und Nordamerika besteht, für alle Confessionen klar, bestimmt und redlich aussprechen werde. Sie haben Letzteres um so zuversichtlicher erwartet, als nur durch die volle und ehrliche Durchführung des Grundsatzes der religiösen Freiheit und der Unabhängigkeit der Kirche vom Staat eine aufrichtige und innige politische Einheit Deutschlands möglich ist. In allen diesen Erwartungen aber hat sie der angeführte Artikel des Commissionsentwurfes bitter getäuscht.

Statt jener klaren und großen Sätze der religiösen Freiheit, finden wir nichts als geschraubte, bedeutungslose Bestimmungen und hinterhaltsvolle Clauseln. Daß Jeder seine Religion frei bekennen darf, versteht sich von selbst. Allein was wir fordern und verlangen müssen, ist unendlich mehr. Wir müssen auch fordern, daß jede religiöse Gesellschaft, jede Confession, jede Kirche als solche als Corporation in ihrer Verfassung, Verwaltung, in ihrer Lehre, in ihrem Gottesdienst, in der Wahl und Anstellung ihrer Diener, in dem Besitz, dem Erwerb und der Verwaltung ihres Vermögens frei und unabhängig sey und keinen anderen Beschränkungen unterliege, als denen, welche das allgemeine Gesetz allen Staatsbürgern in privatrechtlicher und staatsrechtlicher Beziehung auflegt, anstatt daß bisher der Staat sich in alle kirchlichen Angelegenheiten einmischte, die Abhaltung von Synoden, die Erlassung kirchlicher Verfügungen oder Schreiben, ja sogar die Gültigkeit aller kirchlichen Gesetze von seiner Genehmigung abhängig machte, die Wahl und Anstellung der Geistlichen vielfach sich selbst zueignete, die Vermögensverwaltung der Kirche seinen Behörden unterordnete, ja sogar die Feier des Gottesdienstes betreffende Bestimmungen erließ. Allein die Frankfurter Verfassungscommission scheint gar nichts dagegen zu haben, daß dieser unfreie Zustand der Kirche auch unter der Herrschaft der neuen Freiheit fortdauere. Oder meint man vielleicht, daß in dem vagen Ausdruck „Freiheit des Bekenntnisses," auch schon die religiöse und kirchliche Freiheit enthalten sey? so bedenke man, daß dasselbe Wort bisher in allen modernen Gesetzen, wenigstens für die Bekenner der drei Hauptbekenntnisse gestanden, aber die schreiendsten Eingriffe in die religiöse Freiheit nicht gehindert hat.

Allein selbst diesem Ausdruck, „Freiheit des Bekenntnisses," hat die Commission noch Clauseln und Bestimmungen angehängt, die, wenn sich nicht eine geheime, die volle Freiheit des Bekenntnisses beeinträchtigende Nebenabsicht dahinter verbirgt, mindestens ungeeignet und überflüssig sind. Daß der Anhänger irgend einer Confession, wenn er bei Gelegenheit von Religionsübungen ein Verbrechen begeht, vor die ordentlichen Gerichte zu stellen und zu strafen ist, und daß keine Religion ihre Bekenner von den allgemein-staatsbürgerlichen Pflichten befreie, ist etwas sich so von selbst Verstehendes, daß gar nicht einzusehen, weßhalb es gerade hier, wo von der Religion die Rede ist, ausgesprochen wird. §. 3. steht: „Jeder darf Unterricht ertheilen," warum wird da nicht beigefügt, „vorbehaltlich der Bestrafung der Verbrechen, die er dabei begeht?" §. 2. steht: „Freie Wahl des Berufes für Jeden;" warum nicht der Zusatz: „vorbehaltlich seiner staatsbürgerlichen Pflichten." §. 4. steht: „Freiheit der Meinungsäußerung in Rede und Schrift," warum nicht wiederum der gleiche Vorbehalt? Ohne Zweifel, weil solche sich von selbst verstehenden Vorbehalte nicht blos überflüssig, sondern als einen unwürdigen Verdacht aussprechend, selbst beleidigend sind für Die, welche es angeht.

„Die Gleichheit der Bekenner aller Religionen vor dem Gesetz ist anerkannt." Wir verlangen aber die vollkommenste Parität nicht blos für die Einzelnen, sondern auch für die kirchliche Gesammtheit. Nicht mehr soll der Staat, wie es bisher vielfach so sehr der Fall gewesen, eine Religionsgemeinschaft vor der andern begünstigen; er soll vielmehr jede Kirche frei gewähren lassen und sich darauf beschränken, ihr rechtlichen Schutz gegen alle Gewalt, Beleidigung oder Beraubung angedeihen zu lassen. Aber nicht blos sollen die verschiedenen Kirchen vor dem Staate gleich — denn es gibt auch eine Gleichheit in der Dienstbarkeit, — sondern sie sollen auch frei seyn. Freiheit ist im Entwurfe ausdrücklich nur ausgesprochen, und sogar zweimal, für die Bildung neuer Religionsgesellschaften. Wir sind damit vollkommen einverstanden; dagegen fordern wir, daß auch die alte christliche Kirche derselben Freiheit genieße, wie die neu entstehenden Religionsparteien, und daß nicht mehr, wie es bisher der Fall war, z. B. der Staat nach seinem Gutdünken Pfarrer einsetze, dem Ministerium mißliebige Geistliche absetze, während neue Religionsparteien in der Wahl und Entfernung ihrer Diener ganz frei und unbehindert sind.

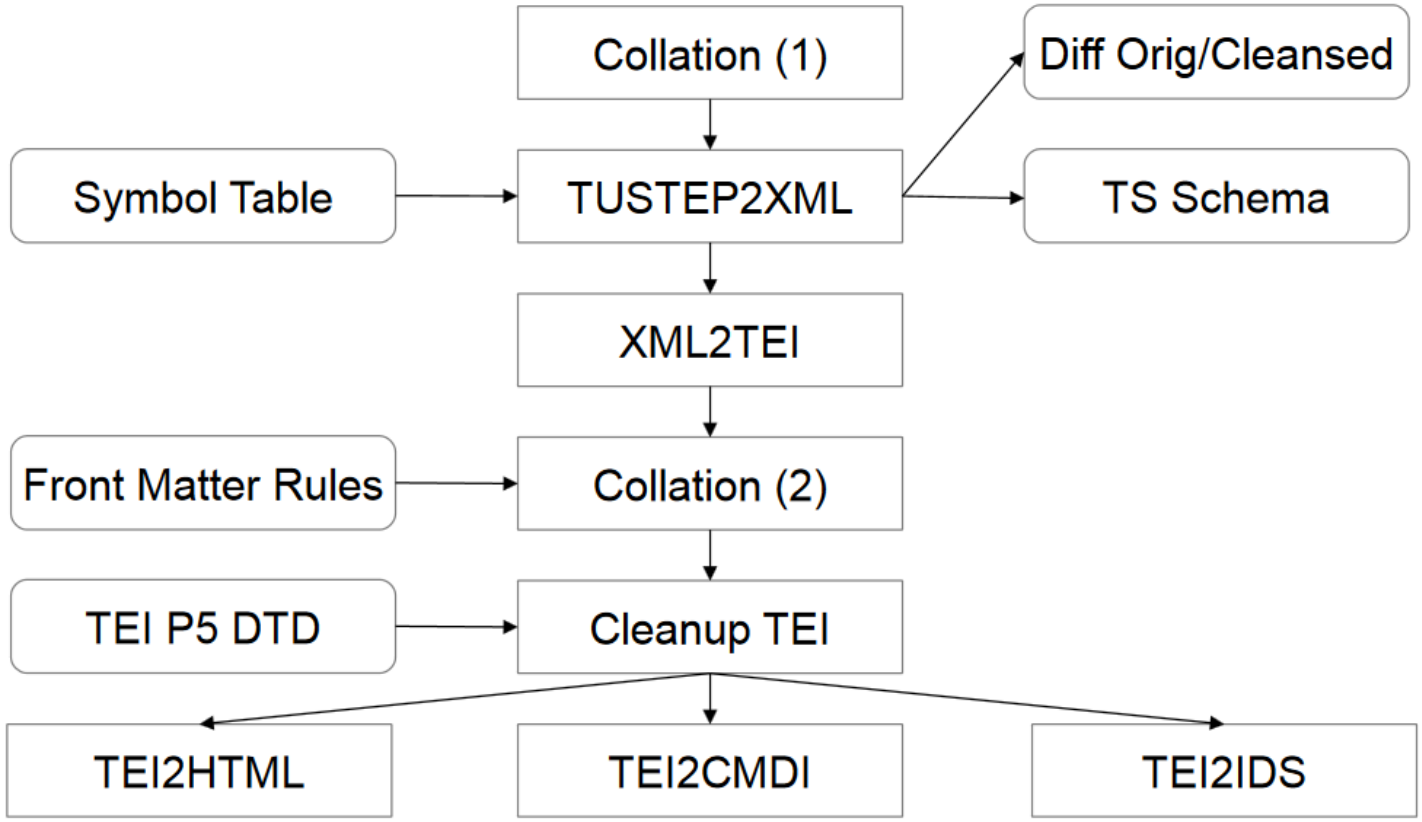
Daß Niemand zu einer kirchlichen Handlung gezwungen werden könne, hat unsere volle Billigung und hätten gewisse Regierungen diesen Grundsatz, der zum A. B. C. der religiösen Freiheit gehört, früher beherzigt, so hätten sie nicht Bischöfe und Geistliche mit Absetzung und Gefängniß bestraft, weil sie gewisse kirchliche Handlungen, als ihrem Gewissen widerstreitend, vorzunehmen sich weigerten. Hingegen will uns bedünken, daß wir auch das Umgekehrte fordern dürfen, daß nämlich auch Niemand vom Staate verhindert werden dürfe, irgend eine religiöse Hand-

1) Siehe Beilage 3. zum Mainzer Journal.

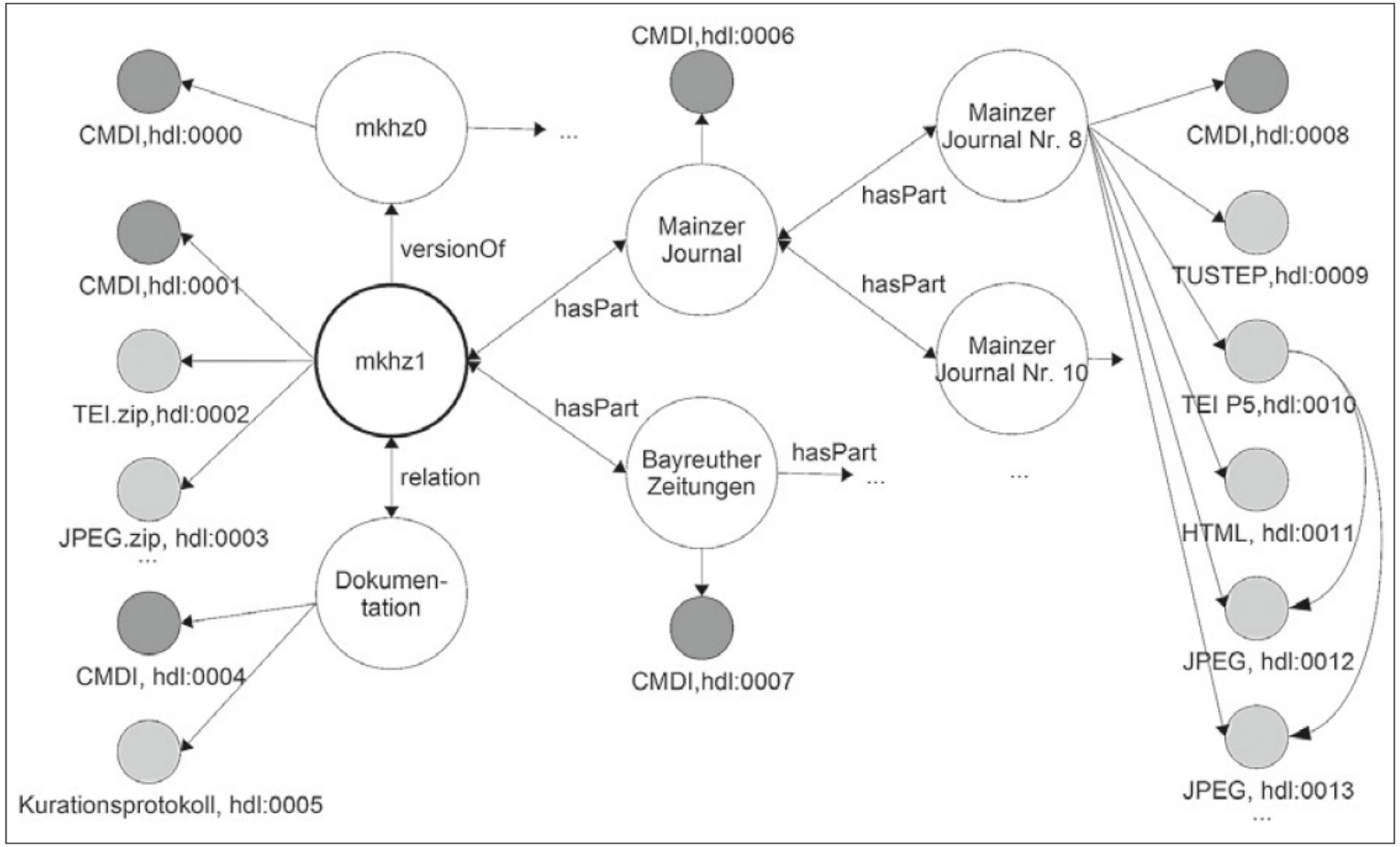
TUSTEP

```
@@0@<filename:MainzerJournal(2)0001.jpg/>
$0 @@1@<SE1>
$0001.01 @z <Z><W+1>Mainzer Journal</W+1>.</Z>
$0001.02 #(IMAGE)
$0001.03 <P>{Numero} 240. <Z><W+1>Dienstag, den 9. October</W+1>.</Z> <R>1849.</R></P>
$0001.04 #(IMAGE)
$0001.05 @z <duolan 2><Z>Die dritte Generalversammlung der katholischen
$0001.06 @Z Vereine in Regensburg.
$0001.07 @Z <A>II</A>.</Z>
$0001.08 <P>__<A>CR</A> Regensburg 2. October. Das feierliche Geläute der
$0001.09 alten, herrlich in gothischem Style erbauten, aber leider nicht
$0001.10 vollendeten Domkirche rief heute früh um 8 Uhr die Mitglieder
$0001.11 des katholischen Vereines Deutschlands zum Gottesdienste. Die
$0001.12 Abgeordneten versammelten sich durch die bereits von Tausenden
$0001.13 gefüllte Kirche auf ihren Ehrenplätzen im Chore, und das Hoch=
$0001.14 amt, welchem der Hochwürdigste Bischof von Regensburg mit
$0001.15 seinem ganzen Domcapitel beiwohnte und bei welchem durch einen
$0001.16 unsichtb{??}aren, gleichsam überirdischen Sängerchor in wahrhaft
$0001.17 künstlerischer Vollendung eine großartige musikalische Messe auf=
$0001.18 geführt wurde, konnte nicht anders, als den tiefsten Eindruck auf
$0001.19 alle Anwesenden machen und die katholischen Männer von nahe
```

Mannheimer Korpus Historischer Zeitungen und Zeitschriften



Mannheimer Korpus Historischer Zeitungen und Zeitschriften



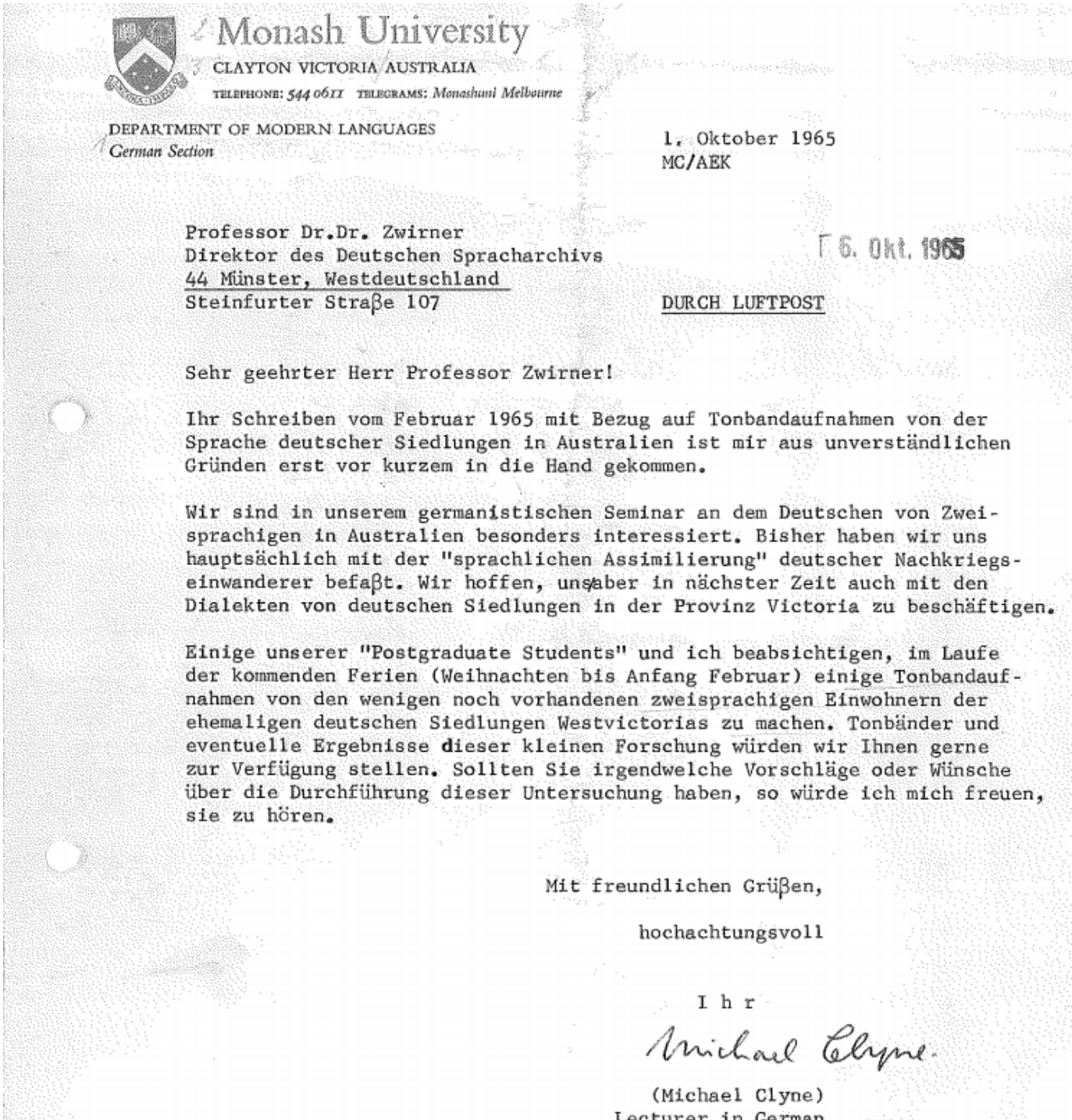
Mannheimer Korpus Historischer Zeitungen und Zeitschriften

- Integration in COSMAS II
- Langzeitarchivierung
- Auffindbar im Virtual Language Observatory CLARIN

- Übergabe der Daten in das Deutsche Text Archiv an der Berlin-Brandenburgischen Akademie

(s. a. Fankhauser, P., Fiedler, N., & Witt, A. (2013). Forschungsdatenmanagement in den Geisteswissenschaften am Beispiel der germanistischen Linguistik. *Zeitschrift für Bibliothekswesen und Bibliographie*, *60*(6), 296-306)

Monash Corpus of Australian German



Monash University
CLAYTON VICTORIA AUSTRALIA
TELEPHONE: 544 0611 TELEGRAMS: Monashuni Melbourne

DEPARTMENT OF MODERN LANGUAGES
German Section

1. Oktober 1965
MC/AEK

Professor Dr.Dr. Zwirner
Direktor des Deutschen Spracharchivs
44 Münster, Westdeutschland
Steinfurter Straße 107

6. Okt. 1965

DURCH LUFTPOST

Sehr geehrter Herr Professor Zwirner!

Ihr Schreiben vom Februar 1965 mit Bezug auf Tonbandaufnahmen von der Sprache deutscher Siedlungen in Australien ist mir aus unverständlichen Gründen erst vor kurzem in die Hand gekommen.

Wir sind in unserem germanistischen Seminar an dem Deutschen von Zweisprachigen in Australien besonders interessiert. Bisher haben wir uns hauptsächlich mit der "sprachlichen Assimilierung" deutscher Nachkriegseinwanderer befaßt. Wir hoffen, uns aber in nächster Zeit auch mit den Dialekten von deutschen Siedlungen in der Provinz Victoria zu beschäftigen.

Einige unserer "Postgraduate Students" und ich beabsichtigen, im Laufe der kommenden Ferien (Weihnachten bis Anfang Februar) einige Tonbandaufnahmen von den wenigen noch vorhandenen zweisprachigen Einwohnern der ehemaligen deutschen Siedlungen Westvictorias zu machen. Tonbänder und eventuelle Ergebnisse dieser kleinen Forschung würden wir Ihnen gerne zur Verfügung stellen. Sollten Sie irgendwelche Vorschläge oder Wünsche über die Durchführung dieser Untersuchung haben, so würde ich mich freuen, sie zu hören.

Mit freundlichen Grüßen,

hochachtungsvoll

I h r

Michael Clyne

(Michael Clyne)
Lecturer in German

Monash Corpus of Australian German

- 1966/67 rund 200 Aufnahmen aus drei Distrikten (Barossa Valley, Wimmera, Western District) und mit „Prewar“ Sprechern
- Die Daten wurden zur Forschung von u.a. Michael Clyne, Sandra Kipp, Brigitte Lambert, Doris Schüpach, Claudia Riehl, … benutzt.
- 2009 – Michael Clyne und der Direktor des IDS einigen sich auf die Aufnahme in die AGD

Michael Clyne

- *1939 †2010
- Australischer Linguist mit deutschen Wurzeln
- Professor für Deutsche Linguistik an der Monash University, Melbourne
- Expert für Soziolinguistik, Zwei- und Mehrsprachigkeit, interkulturelle Kommunikation
- Gastwissenschaftler am SFB 538 in Hamburg
- Zusammenarbeit mit dem IDS seit den 1960s

Monash Corpus of Australian German

- 2012 – Eine Kiste mit CDs, ohne Transkriptionen und ohne Metadaten wird ans IDS geliefert

- Claudia Riehl kann mit Namen und Emailadressen von drei Doktoranden von Michael Clyne in Melbourne aushelfen
- Alle haben (Word) Transkriptionen (und ein paar Metadaten) von verschiedenen Teilen des Korpus
- Die Daten auf den CDs scheinen nicht komplett zu sein.
- Sandra Kipp findet (Kopien?) der originalen Tonbänder und schickt sie ans IDS
- Sandra Kipp stellt Kontakte mit der Barossa Valley historical society und den Lutheran Archives in Adelaide her

Monash Corpus of Australian German

- Kuration der Audiodaten
  - Digitalisierung der Bänder
  - Vergleich mit CDs
  - Editierung der besseren Version (meistens von den Bändern)
    - Normalisierung, Rauschunterdrückung
    - Schneiden, Zusammenfügen der Aufgezeichneten Interaktionen

Monash Corpus of Australian German

- Metadaten Dokumentation
  - Einige Hinweise auf den Covern, Bändern, CDs
  - Einige Hinweise in Publikationen
  - Einige Hinweise in Transkriptionen
  - Biographische Information en in den Aufnahmen selber
  - Immer noch große Lücken

Monash Corpus of Australian German

- Transkriptionen
  - Konvertierung DOC > TXT, Vorverarbeitung mit Skripten
  - Import in EXMARaLDA mittels Simple EXMARaLDA
  - Manuales Alignment via Easy Alignmentfunktion
  - Manuale Vervollständigung (Interviewer meistens nicht transkribiert)
  - Pseudonymisierung der Transkripte / Anonymisierung im Audiosignal
  - Automatische Konvertierung, Orthographische Normalisierung, Lemmatisierung, Part-Of-Speech Tagging

Monash Corpus of Australian German

- Etwa neun Stunden Aufnahmen des ersten Teilkorpus benötigten etwa 250 Personenstunden für die Kuration
- Übernahme in die DGD
- Langzeitarchivierung
- Auffindbar im Virtual Language Observatory CLARIN

- Gorisch, Jan/ Schmidt, Thomas/ Stift, Ulf-Michael (i.V.): Data of German Speech Minorities in the Archive for Spoken German. Eingereicht für: Boas, Hans C. (Hrsg.): Comparative language island research: Data, methods, and goals. Amsterdam: Benjamins.

Data Center for the Humanities

- Forschungsdatenaktivitäten an der UzK
- Aktivitäten des DCH
  - Fallbeispiele
  - Beratungsworkflow
- Die Rolle des DCH im bei der Implementierung eines Forschungsdatenmanagements an der Universität zu Köln

Forschungsdatenaktivitäten an der Philosophischen Fakultät der UzK

- Mitwirkung in der Pilotphase von DARIAH-DE
- Beteiligung am Verbund CLARIN
  - Koordination der F-AG zu Sprachtypologie
  - Etablierung von K-Zentrum auf der Ebene der CLARIN ERIC
  - Innerhalb des nationalen Verbunds CLARIN-D soll noch 2017 ein assoziiertes B-Zentrum zertifiziert werden
- KA^3 - BMBF-Zentrumsausschreibung
- Die UzK ist ein zentraler Partner der AWK-NRW und Akademienunion

Fallbeispiele

Beratungsangebot Projektanträge

Das DCH hilft Ihnen für Ihren Projektantrag ein Konzept für Datenhaltung, Archivierung und Bereitstellung Ihrer Forschungsdaten zu entwickeln und umzusetzen.

Projektentwicklung

Das DCH führt in Kooperation mit Fachwissenschaftlern und unseren lokalen Partnern Entwicklungs- und Kurationsprojekte durch.

Digitale Ressourcen erhalten

Laufend entstehen digitale Ressourcen, Datenbanken, Webseiten und Forschungsdatenbestände im Rahmen von zeitlich befristeten Projekten. Was geschieht mit diesen, nach dem Ende des Projekts?

[1] Beispiel Anfrage

„Wir schreiben […] gerade einen Antrag im Rahmen des Programms H2020 Reflective Societies, den wir in der kommenden Woche einreichen wollen.“

Dimension: Wann im Projektverlauf?

vor Antragstellung – laufendes Projekt (anlaufend/auslaufend) – nach Projektende – legacy Data

[2] Beispiel Anfrage

„Wir starten in Kürze ein DFG-Projekt […] und sind an der Einrichtung eines Coworking Spaces und einer Plattform für den Datenaustausch interessiert.“

Dimension: Welche Bedarfe gibt es?

[3a] Beispiel Anfrage

„Im DFG-Projekt […] nähert sich nach drei Jahren das Ende der Förderung (Februar 2018) und wir überlegen gerade, wie wir die angefallenen Forschungsdaten am besten aufbewahren können. Im Antrag hatten wir die Formulierung "Archivierung der Daten und Unterlagen" aufgenommen, wobei es sich in diesem Projekt um typische geisteswissenschaftliche qualitative Daten handelt.“

[3b] Beispiel Anfrage

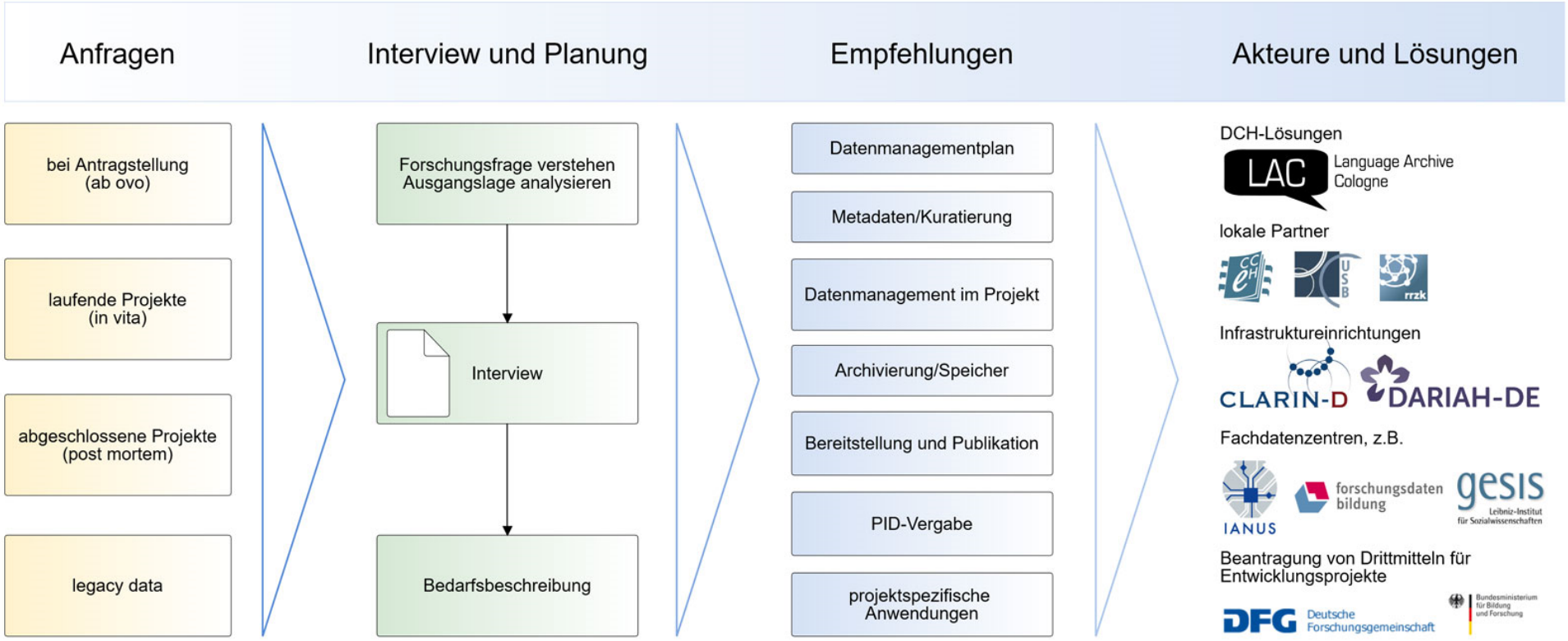
„Leider haben wir zu Beginn des Projekts versäumt, Standards festzulegen, deshalb habe ich drei Fragen: […]

- Liegt nach dem Projektende das Urheberrecht - zum Beispiel für Präsentationen - beim Projekt oder bei mir als individueller Wissenschaftlerin?
- Wie sieht es mit dem Nutzungsrecht von digitalen Fotos aus, die ich in Archiven nach Registrierung eines persönlichen Accounts für den eigenen wissenschaftlichen Gebrauch gemacht habe, sowie bei Digitalisaten, die ich als Wissenschaftlerin für bestimmte Publikationen mit Projektgeldern erworben habe?
- Wie lässt sich nach dem Ende des Projekts die betreffende LitLink-Datenbank am besten archivieren?“

Bedarfsgruppen

- Beratung/Support:
  - Datenmanagement im Projekt, wie?
  - Modellierung, Tools, Standards, Programme, Formate im Projekt, was/wie?
  - Verwertungsrecht/Wahrung von Persönlichkeitsrechten, was/wie?
- Daten
  - Workspace
  - Archivierung
  - Publikation
- Erhaltung von Systemen
  - Datenpräsentation (online-Portale, Repositories)
  - Begleitende Publikationen (begleitende Websites)
  - …

Beratungsworkflow



DCH generischer FDM-Aktivitäten

- Schnittstelle zwischen generische Services der USB/RRZK und Forschungsprojekten/Forschenden aus den Disziplinen
- Blaupause für Implementierung FDM-Konzepts in den Fakultäten
  - Beratungsworkflow
  - Datenkuratierung
  - Datenmanagementpläne
  - Juristische Aspekte (von Verwertungs- bis Persönlichkeitsrechten)
  - Metadaten (von DublinCore über TEI bis CMDI)
  - PID-Vergabe (von ISO:24619 bis DOI)
  - OAIS konforme Architekturen
  - Repositorysysteme
  - Kodifizierte Übernahmeprozesse für legacy Data
  - dynamischen Forschungsdatensysteme

DCH im Kontext von FDM-Aktivitäten

- Einbindung des auf Landesebene zentralen GW-Datenzentrums in die Aktivitäten der Universität
- Vertretung des Akteurs UzK auf nationale Ebene im Bereich gw-Forschungsdaten
  - Vertretung der AWK in der Akademienunion
  - CLARIN ERIC K-Centre (ZAS Berlin, SOAS London, HZSK Hamburg)
  - CLARIN-D assoziiertes B-Zentrum
  - Verbund der Datenzentren im Verband DHd (AG-Datenzentren)

Infrastruktureinrichtungen

- Langfristige, sichere Aufbewahrung der Forschungsdaten bei einer vertrauenswürdigen Institution
  - Bereitstellen von Speicherplatz
  - Backups
    - Sicherung gegen Hardwareausfälle, menschliche Fehler, kleinere (Überspannung, Brände, Überflutung) und größere Katastrophen
      - Verschiedene Räume/ Gebäudeteile
      - Verschiedene Gebäude
      - Verschiedene Städte
    - Institut für Deutsche Sprache: Sicherung in verschieden Gebäuden in Mannheim

Infrastruktureinrichtungen

- Zertifizierung
  - Data Seal of Approval (DSA)
    - Initiiert von DANS (Data Archiving and Networked Services)
    - Übergang in Core Trust Seal
  - nestor (Zertifizierung gemäß DIN-Norm 31644 „Kriterien für vertrauenswürdige digitale Langzeitarchive“
  - ISO 16363
- Aufnahme in Infrastrukturverbünden
  - Verbund CLARIN (centre assessment)
  - Verbund DARIAH
  - …

Infrastruktureinrichtungen

- Zugangskontrolle
  - Wer hat Zugang?
    - Jeder, bestimmte Gruppen
    - Individuelle Regelungen
  - Wie werden Berechtigungen kontrolliert?
    - Shibboleth für CLARIN Anwendung am IDS
  - ➢ Herausforderungen:
    - ➢ Datensicherheit
    - ➢ Individuelle Regelungen

Infrastruktureinrichtungen

- Zitierfähigkeit

  Persistente Identifikatoren (PID)

  - DOI
  - Archival Resource Key
  - Handle

Infrastruktureinrichtungen

- Auffindbarkeit
- Metadaten
  - DC
  - CMDI
  - ➢ Herausforderung: Optimale Beschreibung der Daten mittels Metadaten
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
  - Beispiel: CLARIN Virtual Language Observatory (VLO)
  - Derzeit 37285 Metadatensätze vom IDS

Infrastruktureinrichtungen

- Einfache Nutzbarkeit der Daten
  - Gute Beschreibung der Daten in den Metadaten
  - Sinnvolle Datenformate
  - Konsistenz der Daten und Datenformate

Lösungsansätze für die Kuratierung von Daten aus den Digital Humanities

- Problemanalyse durch Personen mit einem Hintergrund in ‚data science' und der jeweiligen Fachdisziplin, z.B. Archäologie, Linguistik, Lusitanistik, Theologie, Zeitgeschichte
- Erstellung der Datenmanagementpläne
- Juristische Analyse (Möglichkeiten der Datenfreigabe)
- Daten- und Metadatenformatfestlegungen (Standards)
- Untersuchung von disziplinspezifischen und generischen Services

Lösungsansätze für die Kuratierung von Daten aus den Digital Humanities

Archivierung und Bereitstellung der Forschungsdaten über zertifizierte, lokale Datenzentren, da

- Nähe zu den Datenerzeuger/innen
  - Akzeptanz
  - Interaktion
- Einbindung in lokale Infrastrukturen, wie Bibliotheken, Rechenzentren
  - Nutzung generischer Dienste
  - Interdisziplinäre Nutzung der Ressourcen in der Heimateinrichtung
- Einbindung in Verbünde, z.B. in CLARIN und DARIAH
  - Authentifizierungs- und Autorisierungsinfrastruktur
  - Nachweis in Katalogen

Perspektive: Einbindung in eine NFDI

Vielen Dank an:

- Sabrina Kistner (nestor, DNB)
- Thomas Schmidt (IDS)
- Jan Gorisch (IDS)
- Peter Fankhauser (IDS)
- Denis Arnold (IDS)
- Jonathan Blumtritt (DCH an der UzK)
- Felix Rau (UzK)
- Patrick Sahle (UzK)
- Brigitte Mathiak (UzK, Gesis)

nestor

CLARIN
Common Language Resources and Technology Infrastructure

Forschungsdaten in den Digital Humanities - Der Verbund nestor und Datenzentren am Institut für Deutsche Sprache und an der Universität zu Köln

Andreas Witt

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

IDS INSTITUT FÜR DEUTSCHE SPRACHE
Mitglied der Leibniz-Gemeinschaft