

Andreas Witt / Oliver Schonefeld

Informationsinfrastrukturen am Institut für Deutsche Sprache

Abstract

This paper describes the effort of the Institut für Deutsche Sprache (IDS), the central research institution for the German language, connected with Information and Communications Technology (ICT). Use of ICT in a language research institute is twofold. On the one hand, ICT provides basic services for researchers to accomplish their daily work. On the other hand, several national and international institutions have a strong interest in ICT. Therefore, ICT can also be seen as an amplifier for language research. The first part of this paper reports on the activities of the IDS in internal and external ICT-related projects and initiatives. The second part describes a general strategy towards an ICT strategy that could be useful both for the IDS and other national language institutes. We think such a general strategy is necessary to create a strong foundation not only for the ICT-related projects, but as a basis for a modern research institute.

1. Einleitung

Informations- und Kommunikationstechnologie ("Information and Communications Technology (ICT)") ist in den letzten Jahren in den Geisteswissenschaften bzw. der geisteswissenschaftlichen Forschung zu einem festen Bestandteil der Arbeitswelt geworden und wird in vielfältiger Weise eingesetzt. Dabei können grob zwei verschiedene Bereiche unterschieden werden: Auf der einen Seite finden sich IT-Systeme nahezu überall im Alltag eines (Geistes-)Wissenschaftlers. Forschungsinstitute müssen daher heutzutage eine Reihe von IT-Basisdiensten bereitstellen. Am Institut für Deutsche Sprache (IDS) sind daher verschiedene IT-Systeme etabliert und akzeptiert. Auf der anderen Seite ist die IT-Infrastruktur auch für grundlegende wissenschaftliche, am IDS natürlich sprachwissenschaftliche, Forschungen ein Motor. Dies zeigt sich am IDS unter anderem auch daran, dass in den letzten Jahren eine Vielzahl von Infrastrukturforschungsprojekten mit sprachwissenschaftlichem Fokus institutionalisiert worden sind.

Am 31.01.2011 veröffentlichte der Wissenschaftsrat Empfehlungen für Forschungs- und Informationsinfrastrukturen für die Geistes- und Sozialwissenschaften. Das IDS fühlt sich durch diese Veröffentlichungen aus unterschiedlichen Gründen angesprochen: direkt, da die IDS-Aktivitäten im Bereich Forschungsinfrastrukturen mehrfach explizit erwähnt wurden, aber auch indirekt, da die sich das IDS bestätigt fühlt, in den vergangenen Jahren seine Forschungs- und Entwicklungsaktivitäten so intensiv in diesem Bereich vorangetrieben zu haben. Zu diesen Aktivitäten gehören nicht nur genuine IDS-Projekte, die sich im Kontext der neuen Forschungsumgebungen positioniert haben, sondern auch Aktivitäten die im Rahmen nationaler oder internationaler Verbünde durchgeführt werden.

Aus der Sicht des IDS sind die nachfolgend aufgeführten Empfehlungen des Wissenschaftsrats besonders relevant:

- Das Engagement insbesondere für die von Forschungsfragen getriebenen Infrastrukturentwicklungen muss aus den wissenschaftlichen Gemeinschaften kommen, und entsprechende Projektvorschläge müssen sich in einem Ideenwettbewerb auszeichnen. (Wissenschaftsrat 2011a, S. 8)

- [Um] Infrastrukturbedarfe der disziplinären Grundversorgung [...] unabhängig von konkreten Projekten auf der Ebene einer Fachgemeinschaft [zu ermitteln, ist es] eine notwendige Voraussetzung [...], dass die entsprechende Fachgemeinschaft über eine für Infrastrukturfragen sensibilisierte und gegenüber den Zuwendungsgebern artikulationsfähige Organisationsstruktur verfügt. [...] Das Problembewusstsein und die Artikulationsfähigkeit insbesondere der Geisteswissenschaften [...] [sind] im Infrastrukturbereich vergleichsweise gering ausgeprägt. Hier besteht auch angesichts des erwartbaren nationalen Roadmap-Prozesses, der zu einer disziplinübergreifenden Priorisierung von Forschungsinfrastrukturen führen soll, noch Verbesserungsbedarf. (ebd., S. 42f.)
- Den Infrastruktur tragenden Einrichtungen empfiehlt der Wissenschaftsrat, sich bereits in der Konzeptionsphase für neue Infrastrukturprojekte in den Geistes- und Sozialwissenschaften mit anderen nationalen und internationalen Einrichtungen abzustimmen. (ebd., S. 82)
- Öffentlich finanzierte Informationsinfrastrukturen sollten externen Nutzerinnen und Nutzern aus dem In- und Ausland für wissenschaftliche Zwecke grundsätzlich zugänglich sein. (Wissenschaftsrat 2011b, S. 49)
- Informationsinfrastrukturen sollten eine zentrale Rolle bei der Entwicklung von sowohl national als auch international ausgerichteten Kooperations- und Vernetzungsstrategien der verantwortlichen Einrichtung einnehmen. In kooperativen Forschungsprojekten ist darauf zu achten, dass die entsprechenden Einrichtungen nicht ausschließlich oder überwiegend auf Servicefunktionen wie beispielsweise die Bereitstellung von Daten beschränkt bleiben, sondern sich mit eigenen Forschungsbeiträgen beteiligen. (ebd., S. 49)

Die bereits etablierten und in diesem Beitrag skizzierten forschungsinfrastrukturbezogenen Aktivitäten des IDS können im Kontext dieser Empfehlungen betrachtet werden.

Der Artikel ist dergestalt gegliedert, dass in der ersten Hälfte ein Vielzahl von IDS-Aktivitäten mit starkem Bezug zu ICT aufgeführt sind. Diese sind entweder reine Hausprojekte, die zum Teil seit mehreren Jahrzehnten erfolgreich laufen, oder nationale und internationale Verbundprojekte an denen sich das IDS beteiligt. Der zweite Teil des Beitrags skizziert eine Gesamtstrategie zum Umgang mit Informations- und Kommunikationstechnologie im IDS. Eine derartige Strategie ist unabdingbar um eine moderne und zukunftssichere Basis für den Umgang mit ICT am IDS zu schaffen und bildet damit nicht nur Grundlage für die Durchführung der im ersten Teil beschriebenen Projekte, sondern ist vielmehr die Basis für den Betrieb einer geisteswissenschaftlichen Institution, die sich der Erforschung und Dokumentation der Sprache widmet. Wir sehen daher die beschriebenen Aktivitäten des IDS in diesem Kontext auch als mögliches Modell für andere EFNIL-Einrichtungen.

2. Grundständige hausinterne Projekte des IDS und interne Vernetzung

Das IDS führt seine Forschungsarbeiten in sogenannten Programmbereichen durch, die innerhalb einzelner Abteilungen oder im Bereich Zentrale Forschung angesiedelt sind. Nachfolgend werden exemplarisch aus zwei Programmbereichen einige Projekte, die insbesondere auf eine moderne informationstechnologische Infrastruktur angewiesen sind, vorgestellt.

2.1 Korpuslinguistik

Der Programmbereich Korpuslinguistik widmet sich seit mehreren Jahrzehnten u.a. dem Aufbau, der Pflege und dem Zugang zu Korpora geschriebener Sprache. Die empirische

Basis der Forschungsarbeiten der Korpuslinguistik bildet das Deutsche Referenz Korpus (DeReKo), das kontinuierlich weiter ausgebaut wird. Es ist bereits seit vielen Jahren weltweit eines der größten Sammlungen geschriebener deutscher Sprachdaten. Anfang des Jahres 2011 umfasste es etwa 4 Milliarden Wörter. Das Design des Korpus orientiert sich an dem im IDS konzipierten Ansatz der Urstichprobe. Dieser Ansatz steht im Kontrast zu den repräsentativen, ausgewogenen oder auch gewichteten Korpora, deren Design eine finale Ausbaustufe, meist definiert bezüglich Größe und Texttypverteilung, definiert, die dann beim Korpusaufbau angestrebt wird. Eine Urstichprobe dagegen wächst kontinuierlich und umfasst alle Texte, die akquiriert werden können. Die tatsächlichen Arbeitskorpora, die als virtuelle Korpora bezeichnet werden (Kupietz et al. 2010a), sind eine Teilmenge der Urstichprobe und die Auswahl der Texte basiert auf den Metadaten, die den einzelnen Texten zugeordnet sind.

Aufgrund von urheber- und lizenzrechtlichen Restriktionen ist es nicht möglich, die (virtuellen) Korpora frei, z.B. zum Download, zur Verfügung zu stellen. Vielmehr erfolgt der Zugriff mit spezialisierter Software. Derzeit steht hierfür das Korpusrecherche- und -analyse-system COSMAS (Corpus Search, Management and Analysis System) zur Verfügung. In seiner ersten Fassung (COSMAS I; vgl. al-Wadi 1994) wurde es 1991 der wissenschaftlichen Öffentlichkeit zur kostenlosen Benutzung zur Verfügung gestellt. Das Nachfolgesystem COSMAS II (Bodmer 2005) wurde 2003 in Betrieb genommen und hat sich seitdem im Dauerbetrieb bewährt. Es können u.a. über eine WWW-Schnittstelle Suchanfragen zu Wörtern, Wortfolgen und grammatischen Mustern eingegeben werden, nach denen die Korpus Texte durchsucht werden sollen. Die Ergebnisse können gemäß unterschiedlichen Kriterien, u.a. nach Entstehungszeit, Erscheinungsland und Thematik, sortiert werden. Die gefundenen Belege zeigen die gesuchten Wörter im Kontext ihrer textuellen Umgebung gemeinsam mit den bekannten bibliographischen Angaben, wie Verlag, Autor, Entstehungszeit, Seitenangabe. Derzeit hat COSMAS II ca. 17.000 registrierte Benutzer.

Da COSMAS I und II bereits Anfang der 90er Jahre konzipiert wurden und die darauffolgenden softwaretechnischen Umsetzungen auf dem damaligen Stand der Informationstechnologie erfolgte, steigt heutzutage der Arbeitsaufwand diese Software zu pflegen und zu erweitern überproportional an. Zwischenzeitlich haben sich sowohl die technischen als auch die wissenschaftlichen Rahmenbedingungen derart stark verändert, dass die Entwicklung eines neuartigen Analyse-Tools erstrebenswert wurde. Mit dem Projekt "Korpusanalyseplattform der nächsten Generation" (KORAP) gelang es dem IDS in einem kompetitiven Auswahlverfahren der Leibniz Gemeinschaft ein Projekt einzuwerben, in dem eine neuartige Korpusanalyseplattform entwickelt werden wird, die eine Grundlage für den methodisch validen Umgang mit *very large corpora* insbesondere im Bereich der empirischen germanistischen Sprachwissenschaft schafft.

Um die eindeutige und nachhaltige Referenzierbarkeit von Sprachressourcen, z.B. auch der virtuellen Korpora, sicherzustellen, hat sich der IDS-Programmbereich Korpuslinguistik gemeinsam mit dem Programmbereich Forschungsinfrastrukturen auch im Bereich der Standardisierung eines persistenten Identifikationsmechanismus für Sprachdaten engagiert. Die Relevanz dieser Aufgabe ist auch in dem Wissenschaftsratsempfehlungen hervorgehoben: "Für die Langzeitarchivierung von Forschungsprimärdaten empfiehlt

der Wissenschaftsrat den Ausbau von Referenz- und Zitationsmöglichkeiten für Datensätze" (Wissenschaftsrat 2011a, S. 9). Die international abgestimmte Standardisierungsaktivität mündete in dem vor seiner endgültigen Verabschiedung stehenden Standard ISO 24619 "Language resource management – persistent identification and sustainable access (PISA)".

© Cyril Belica: Modelling Semantic Proximity - Self-Organizing Map (version: 0.32, init tau: 0.04, dist: u, iter: 5000)

Sprache

Tschechisch	Italienisch	Fremdsprache	Rechnen
Niederländisch	Spanisch	gelehrt	Grundkenntnis
Ungarisch	Latein	pauken	Kulturtechnik
Schwedisch	Griechisch	büffeln	Einmaleins
Russisch	Französisch	Sprachunterricht	Lautsprache
Finnisch	Slowenisch	Unterrichtsfach	Landeskunde
Dänisch	Unterrichtssprache	Erdkunde	Phikologie
Rumänisch	Sprachkenntnisse	unterrichten	Naturwissenschaft
Portugiesisch	Muttersprache	Gebärdensprache	Grammatik
Amtssprache	Landessprache	Konversation	Grundbegriff
Jiddisch	Hebräisch	rudimentär	Rechtschreibung
Hindi	Polnisch	Muttersprachler	Sprechen
Hochdeutsch	Englisch		Lesen
fließend	Chinesisch		Metier
Schweizerdeutsch	Türkisch		Lesen
akzentfrei	Japanisch		Handwerk
Idiom	Schriftsprache	slawisch	Mentalität
Umgangssprache	Vokabel	Diktion	Bauchtanz
abfassen	übersetzt	linguistisch	Kulturraum
Mundart	übersetzen		Ausdrucksweise
Dialekt	Hochsprache		Ausdrucksmittel
abgefasst	Fachsprache		Kultur
Originalsprache	Blindenschrift		Bildsprache
gesprochen	verfasst		Stilelement
phonetisch	hebräisch	Literatur	Abstammung
Wort	Übersetzung	Dichtung	Herkunft
jiddisch	lateinisch	Lyrik	Staatsbürger
Alltagssprache	Text	Kalligraphie	Nationalität
hochdeutsch	kyrillisch	Ursprung	Folklore
Fachausdruck	Wörter	Sprachraum	Provenienz
plattdeutsch	Schriftzeichen	germanisch	Intellektueller
Redewendung	Alphabet	Einsprengsel	Lebensart

Abb. 1: Kookkurenzprofil des Wortes "Sprache" visualisiert als Selbstorganisierende Karte (SOM)

Ein weiteres in der Korpuslinguistik entwickeltes Werkzeug bildet die Kookkurenzdatenbank (CCDB; Keibel/Belica 2007). Die CCDB wird beschrieben als eine "korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs" (<http://corpora.ids-mannheim.de/ccdb/>). In Abbildung 1 findet sich beispielhaft die Anzeige der Kookkurenzprofile des Wortes "Sprache". In unterschiedlichen Feldern, die in der Online-Fassung farbig angezeigt werden,

sind verschiedene Lemmata aufgeführt, die sich, basierend auf statistischen Analysen in den Textkorpora, gruppieren lassen. Es kann beobachtet werden, dass die Wörter in den einzelnen Feldern gewisse semantische Eigenschaften teilen, obwohl keine manuelle Bearbeitung der Ergebnisse erfolgte.

Das Beispiel der CCDB zeigt ein weiteres Charakteristikum sowohl des Instituts für Deutsche Sprache als auch der Einsatzgebiete der Informations- und Kommunikationstechnologie in der sprachwissenschaftlichen Forschung: die Schnittstellen zwischen den einzelnen Bereichen und Abteilungen sind nicht immer trennscharf beschreibbar. So basiert die Kooperationsdatenbank zwar auf den Korpora und wurde konsequenterweise im Programmbereich Korpuslinguistik entwickelt, die Ergebnisse dieser Forschungsarbeiten sind aber auch für den Programmbereich Lexik äußerst nützlich und werden in bestimmte Onlineangebote der Lexik integriert.

2.2 Lexik

Im Programmbereich Lexik wird seit mehreren Jahren das Online-Wortschatz-Informationssystem Deutsch (OWID) entwickelt. Dieses Webportal (www.owid.de) ermöglicht den Zugang zu verschiedenen im Bereich Lexik entwickelten Angeboten. Hierbei handelt es sich um "elexiko" (Müller-Spitzer 2008), das digitale "Neologismenwörterbuch", eine digitale Sammlung "Fester Wortverbindungen" und das elektronisch verfügbare "Diskurswörterbuch 1945-1955". Diese digitalen Ressourcen haben zum Teil auch ein analoges Äquivalent, z.B. wurden die Neologismenwörterbücher des IDS auch auf Papier veröffentlicht (Herberg et al. 2004). Des Weiteren hat der Bereich Lexik eine Online-Bibliografie zur elektronischen Lexikografie (OBELEX) erarbeitet, die über das World Wide Web zur Verfügung gestellt wird.

The screenshot shows the OWID website interface. At the top, there is a search bar with the text 'elexiko' and a search button. Below the search bar, there are navigation links: 'Startseite elexiko', 'Wortartikel', 'Suchanfragen', 'Projekt', 'Benutzungshinweise', and 'Erweiterte Suche'. A secondary navigation bar contains letters A-Z for quick access. The main content area is titled 'Sprachpolitik' and 'Lesartenübergreifende Angaben'. Under 'Orthografie', there are links for 'Normgerechte Schreibung' and 'Worttrennung', both pointing to 'Sprachpolitik' and 'Sprachpolitik'. The 'Belege (automatisch ausgewählt)' section contains a text snippet: 'In Bredau waren zum Beispiel um 1890 mehr als drei Viertel der Bürger Juden, und es bestanden nur zwei Schulen: eine deutsche und eine polnische. Die Juden beantragten zwei zusätzliche deutsche Schulen, dies wurde aber vom Landesausschuss und vom Unterrichtsausschuss im Zeichen der Polonisierungspolitik abgelehnt. - In der Zukunft war das Übergewicht einer Nationalität und einer Muttersprache nicht gegeben. Das Gleichgewicht der Volksgruppen Rumänen, Ruthenen, Deutsche, Polen und Juden machte die Sprachpolitik schwierig. 1900 fand die "Jüdische Sprachkonferenz" in Czernowitz statt. (PNS/NAI 17443 Die Presse, [Tageszeitung], 04.08.1998. - Originalressort: Spectrum)'. Below this, there is a 'STICKEL' section with text: 'STICKEL: Das Wort Sprachpflege mag ich nicht, es unterstellt, dass Sprache so etwas sei wie ein Patient. Hin und wieder sprachlich einzugreifen macht mir aber Spaß. Auch wenn es nicht zum offiziellen Auftrag dieses Instituts gehört, habe ich mich als Wissenschaftler und sprachlicher Zeitgenosse eingemischt. Beim Sprachgebrauch im Rechtswesen und der Verwaltung oder sonstwie auch im Bereich der Sprachpolitik, wenn es darum geht, für unsere Sprache weiterhin auch internationale Beteiligungsfelder zu erhalten, ohne in Wissenschaft und Außenpolitik. Ich glaube, dass man Sprachentwicklung

Abb. 2: Das elektronische Wörterbuch "elexiko"

Abbildung 2 zeigt einen Screenshot des elektronischen lexikalischen Informationssystems "elexiko". Da es über das Portal OWID aufgerufen wurde, befinden sich in der Kopfzeile die Logos von OWID, elexiko und dem IDS. Die Basisinformation, die für ca. 300.000 Wörter über elexiko bereitgestellt wird, bildet den Hauptteil der Webseite. Zum Einen werden für die Wörter die korrekte Schreibung oder gegebenenfalls die normgerechten Schreibvarianten sowie die Grenzen für die Silbentrennung aufgeführt. Zum Anderen werden Vorkommen des Wortes in den IDS-Sprachkorpora angezeigt. Dies ist ein weiteres Beispiel für eine programmereichsübergreifende Nutzung digitaler Sprachressourcen. Zu diesen Belegen werden zusätzlich die entsprechenden Metadaten angezeigt. So kann man beispielsweise sehen, dass der erste angezeigte Beleg aus der Tageszeitung "Die Presse" vom 4.5.1996 stammt.

Zusätzlich zu diesen Informationen werden für ausgewählte Lemmata detaillierte Wortartikel in elexiko angezeigt. Derzeit sind bereits ca. 1.600 Wortartikel verfasst worden. Für das in der Abbildung angezeigte Lemma steht ein derartiger Wortartikel bisher nicht zur Verfügung, so dass in dem Screenshot nur die Basisinformationen zu sehen sind.

Die starke Integration unterschiedlicher elektronischer Angebote ist in dem Portal auch an der Einbettung eines Verweises auf die Online-Bibliografie zur elektronischen Lexikografie sichtbar. Die Integrationsmöglichkeit für verschiedene elektronische Angebote stellt einen besonderen Mehrwert gegenüber traditionellen Arbeitsweisen dar. Bei der Aufbereitung von Informationen für das WWW haben sich in den vergangenen zwei Jahrzehnten hierbei sehr elaborierte Techniken entwickelt, die häufig genutzt werden. Eine neue Entwicklung besteht darin, die Integration und Interoperabilität von digitalen Sprachressourcen auch für über Web-Präsentationen hinausgehende Anwendungen zu ermöglichen. Dies ist ein Ziel von verschiedenen Verbundprojekten an denen auch das IDS beteiligt ist.

3. Externe Vernetzung

Das IDS ist mit dem Programmbereich Forschungsinfrastrukturen an einer Reihe von nationalen und internationalen Kooperationsprojekten beteiligt. Weiterhin wirkt das IDS in Standardisierungsgremien mit und beteiligt sich als aktiver Partner im Kompetenznetzwerk für Langzeitarchivierung (nestor). Diese Aktivitäten sollen in den nachfolgenden Abschnitten kurz vorgestellt werden.

3.1 CLARIN / D-SPIN

Das europäische Verbundprojekt CLARIN (Common Language Resource Infrastructure) hat das Ziel, ein europäisches Netzwerk von Infrastrukturzentren für den Zugang und die kollaborative Nutzung von digitaler Sprachressourcen aufzubauen. In verschiedenen an CLARIN beteiligten Ländern gibt es zudem nationale Verbünde, die die Netzwerke in den einzelnen Staaten auf- bzw. ausbauen. In Deutschland arbeitet an dieser Aufgabe das vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Verbundprojekt D-SPIN (Deutschen Sprachressourcen-Infrastruktur), das seit 2011 als CLARIN-D weitergeführt wird. D-SPIN bzw. CLARIN-D hat das Ziel in Deutschland diese Infrastruktur und entsprechende spezialisierte Software interoperabel und nachhaltig verfügbar

bereitzustellen. Um dies zu erreichen werden dedizierte Zentren aufgebaut, in denen Primärdaten, wie Textkorpora oder Sprachaufnahmen, und Metadaten zu derartigen Ressourcen dergestalt bereitgehalten werden, dass sie von Forschungseinrichtungen oder auch von Einzelpersonen insbesondere für die sprachwissenschaftliche Forschung genutzt werden können. Das IDS hat das Ziel im Rahmen von D-SPIN ein Sprachressourcenzentrum aufzubauen und in diesem Verbundprojekt den Standort Mannheim als wichtigen Partner zu etablieren.

Des Weiteren arbeitet das IDS in diesem Rahmen auch an Rechtheften beim Umgang mit Sprachressourcen, wobei insbesondere Urheber-, Verwertungs- und Persönlichkeitsrechte von Bedeutung sind. Konkret erarbeitet das IDS u.a. Modelle für Lizenzvereinbarungen und für den Umgang mit Sprachdaten, die datenschutzrechtlichen Auflagen unterliegen. Außerdem werden weitergehende Fragestellungen betrachtet. Genannt sei hier exemplarisch die Problematik des Konflikts der im Grundgesetz verankerten Grundrechte zum Schutz des Eigentums und zur Freiheit der Forschung und Wissenschaft, da die freie Verfügbarkeit von Sprachressourcen den Rechten der Besitzer der Verwertungsrechte, z.B. der Verlage, entgegenstehen kann.

Ebenfalls im Kontext von D-SPIN und CLARIN-D wird in Mannheim an der Implementierung von Katalogen und Registraturen für virtuelle Kollektionen von Datenressourcen, wie z.B. für die bereits oben erwähnten virtuellen Korpora, gearbeitet, deren Bestandteile auch über verschiedenen Standorte verteilt sein können sollen. Um hierbei keine Insellösungen zu produzieren, war das IDS auch der Konzeption des bereits erwähnten ISO-Standards für die persistente Referenzierbarkeit von Sprachressourcen (ISO/DIS 24619) beteiligt.

3.2 TextGrid

Die nationalen, ebenfalls vom BMBF geförderten Forschungsverbünde TextGrid und WissGrid, an denen das IDS beteiligt ist, widmen sich der Grid-Technologie, einem informationstechnologischen Ansatz, der auf einer – auch räumlich – verteilten Rechnerarchitektur basiert. Die Computer können sowohl als distribuierte Datenspeicher als auch als Ansammlung einer großen Menge von Rechenprozessoren genutzt werden. Letzteres hat in den Naturwissenschaften zu Erfolgen geführt, da z.B. die Verwendung einer Vielzahl von Arbeitsplatzrechnern zu Problemen führte, deren Lösungsalgorithmen parallelisiert werden können und teure Supercomputer ersetzen können. In den digitalen Geisteswissenschaften, in denen sich das Projekt TextGrid verorten lässt, ist der Bedarf an sehr großen Rechenkapazitäten derzeit noch nicht so stark ausgeprägt. Der Einsatz der Grid-Technologie ist dennoch auch für TextGrid sehr sinnvoll, da die damit verbundene Rechnerarchitektur für die verteilte Datenhaltung genutzt werden kann, wodurch die Lastverteilung bei hohen Nutzerzahlen ermöglicht wird. Des Weiteren wird die Ausfallsicherheit erheblich verbessert, da z.B. bei einem Defekt eines Rechners die Nutzer auf den anderen Rechnern weiterarbeiten können.

Das Ziel von TextGrid besteht darin, eine *virtuelle Forschungsumgebung* (engl. Virtual Research Environment, VRE) für die Geisteswissenschaften aufzubauen. Eine VRE wird vom britischen Joint Information Systems Committee (JISC; www.jisc.ac.uk) u.a. dadurch

charakterisiert, dass sie den Forscher/innen und ihren Forschungsprozessen die größtmögliche Unterstützung während der zunehmend komplexer werdenden Arbeitsabläufe bietet und hierfür Werkzeuge und Technologien bereitstellt, die benötigt werden, um kollaborativ, disziplinübergreifend, international und institutionell unabhängig interagieren zu können. Das Projekt TextGrid implementiert für diese Zwecke eine Software, die Module für verschiedene geisteswissenschaftliche Disziplinen unter einer gemeinsamen Benutzeroberfläche zusammenbringt. Diese Module wurden entweder neu entwickelt oder es erfolgte eine Integration existierender Dienste in das TextGrid-System. Zu den beteiligten Fachwissenschaften, für die die Software insbesondere entwickelt wurde, gehören neben der Sprachwissenschaft auch die Musikwissenschaft, die Editionsphilologie und die Kunstgeschichte.

Das Projekt WissGrid hat die Aufgabe, den Einsatz der Grid-Technologie in verschiedene Wissenschaften zu unterstützen und nachhaltig zu etablieren. Beteiligt sind neben der Sprachwissenschaft, die durch das IDS vertreten ist, u.a. auch Klimaforschung und medizinische Informatik. Zu den Zielen von WissGrid gehört die Erstellung von Konzepten für die Langzeitarchivierung im Grid, die Erarbeitung langfristiger Betriebsmodelle, die es ermöglichen die auch kostenaufwändige Infrastruktur langfristig zu betreiben, und das Verfassen von Leitfäden für Fachdisziplinen, die die Grid-Technologien nutzen möchten.

3.3 Standardisierung

In den eingangs erwähnten Empfehlungen des Wissenschaftsrates wird auch die Wichtigkeit von Standards hervorgehoben:

Mit Blick auf die Bereitstellung und externe Nutzung bereits erhobener Daten und fertig gestellter Digitalisate ist ebenfalls eine verstärkte Koordination unter den dezentral verteilten Anbietern anzumahnen. Der Wissenschaftsrat empfiehlt hier nachdrücklich den Trägern von Infrastrukturen, sich bei disziplinär und thematisch verwandten Daten auf gemeinsame Standards der Aufbereitung und auf ein gemeinsames – von einer federführenden Einrichtung zentral bereitgestelltes – Portal zu verständigen, das den interessierten Nutzerinnen und Nutzern eine "one-stop-shopping"-Option ermöglicht. Das heißt, dass dezentral gesammelte, aufbereitete und angebotene Daten den Forscherinnen und Forschern über eine gemeinsame Anlaufstation zur Verfügung stehen müssten, auf die sie mit der ihnen gewohnten Terminologie zugreifen könnten. (Wissenschaftsrat 2011a, S. 83)

Das IDS ist sich der Bedeutung von Standards seit längerer Zeit bewusst. Dies zeigt sich nicht nur daran, dass die zusammengetragenen Sprachressourcen, wenn immer möglich, gemäß existierender internationaler Standards aufbereitet werden, sondern auch an der aktiven Mitwirkung von IDS-Mitarbeitern in den Standardisierungsorganisationen DIN und ISO sowie in der Text Encoding Initiative (TEI).

3.4 Langzeitarchivierung

Die Thematik Langzeitarchivierung digitaler Daten, der sich auch das Projekt WissGrid widmet, ist seit einigen Jahren für viele Institutionen hoch relevant. Von 2003 bis 2008 wurde daher vom BMBF das Netzwerk nestor gefördert, in dem Bibliotheken, Archive, Museen und führende Experten Strategien und Lösungen zu dieser Thematik arbeiten.

Nach dem Auslaufen der finanziellen Projektförderung durch das BMBF wurde dieser Verbund nicht nur von den meisten Projektpartnern weitergeführt, sondern wurde sogar erweitert. Das IDS stieß 2009 zu diesem Netzwerk dazu, da die dauerhafte Aufbewahrung digitaler Sprachdaten für das Institut von höchster Wichtigkeit ist, um seinen Auftrag der Dokumentation der deutschen Gegenwartssprache auch weiterhin so erfolgreich wie bisher erfüllen zu können.

4. Grundüberlegungen für eine IT-Gesamtstrategie am Institut für Deutsche Sprache

Die Anforderungen an Umfang und Qualität der IT-Dienste sind in den letzten Jahren stark gestiegen und diese Entwicklung wird sich auch in der Zukunft fortsetzen. Weiterhin sind eine Reihe neuer Anforderungen durch die notwendige externe Vernetzung in verteilten Infrastrukturen hinzugekommen (CLARIN, TextGrid, föderierte Authentifikations- und Autorisierungs-Infrastrukturen, wie z.B. DFN-AAI).

Um diesen zusätzlichen, bereits bestehenden oder antizipierbaren Anforderungen optimal gerecht zu werden und dabei eine Ausuferung von Insellösungen zu vermeiden, ist es sinnvoll, für das gesamte Institut Ziele für eine IT-Infrastruktur im IDS zu definieren und eine entsprechende IT-Gesamtstrategie zu entwerfen, die sich an diesen Zielen ausrichtet und die vorhandenen Ressourcen berücksichtigt. Eine umfassende Strategie zum Aufbau von IT-Infrastrukturen setzt sich hierbei aus unterschiedlichen Komponenten zusammen.

Benutzer kommen hauptsächlich mit der Komponente der "Dienste" der IT-Infrastruktur in Kontakt, d.h. Nutzung von E-Mail, Internet-Zugang, VPN etc. Diese Dienste bilden das Dach der IT-Infrastruktur, das auf den Säulen "Identitätsmanagement", "Sicherheit" und "Wartung und Betrieb" ruht (vgl. Abbildung 3).

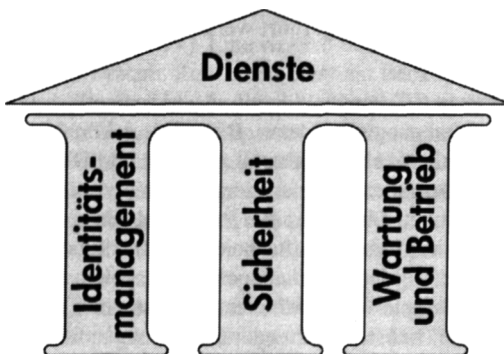


Abb. 3: Säulen der IT-Infrastruktur

4.1 Dienste

Die wichtigste Aufgabe innerhalb der IT-Infrastruktur ist das Bereitstellen von Diensten für den Benutzer. Diese umfassen beispielsweise den Zugriff auf das Internet zu Recherchezwecken, E-Mail, zentrale Datensicherung, zentrale Netzwerkfreigaben für Projekte.

Bereitstellung des Content-Management-Systems (CMS) für den Webauftritt des Instituts sowie Projekt- und personal-bezogenen Webseiten und das Intranet. Hinzu kommen auch noch Dienste der Bibliothek wie der Online Public Access Catalogue (OPAC) oder die Publikationsliste.

Insbesondere durch die zunehmende Mobilität der Mitarbeiter (Konferenzen etc.) oder Heimarbeit ist es weiterhin notwendig, dass die Mitarbeiter bequem auf notwendige Dienste zugreifen können. Für den E-Mail-Verkehr sollte daher die Möglichkeit bestehen, beispielsweise über eine Web-Mail-Schnittstelle E-Mails abrufen und versenden zu können. Der Zugang zu Systemen innerhalb des Instituts oder internen Informationen wie dem Intranet sollte weiterhin über ein Virtual-Private-Network (VPN) gewährleistet sein, wie er bereits erfolgreich im Testbetrieb angewandt wird. Weiterhin sollte über eine Teilnahme am Dienst eduROAM, der unterschiedliche teilnehmende Institutionen zu einer Föderation zusammenschließt, nachgedacht werden. Mitgliedern dieser Föderation wird es ermöglicht, an allen teilnehmenden Institutionen (das sind z.B. nahezu alle Universitäten in Deutschland) mit einer eigenen Benutzerkennung sich am W-LAN anzumelden und so das Internet zu nutzen. Besonders für Besucher des Instituts im Rahmen von Konferenzen oder Workshops bietet dies eine bequeme und einfache Möglichkeit um das Internet zu nutzen. Weiterhin profitieren auch Institutsmitarbeiter von der Teilnahme an eduROAM, da sie, wenn sie bei anderen Institutionen, die an eduROAM teilnehmen, zu Gast sind, ebenso einfach Zugriff auf das Internet haben.

4.2 Identitätsmanagement

Verschiedene Dienste, wie z.B. E-Mail oder VPN, sind auf personenbezogenen Daten angewiesen, um ordnungsgemäß zu funktionieren. Jeder Dienst verwaltet jedoch standardmäßig seine eigenen Stammdatensätze. Wenn sich daher beispielsweise durch eine Heirat der Namen eines Mitarbeiters ändern sollte, müssten diese Änderungen auf allen beteiligten Systemen einzeln durchgeführt werden. Dies führt schnell zu Fehlern und Inkonsistenzen.

Ein Identitätsmanagement (IdM, auch "Identity-Management") ermöglicht eine zentrale Verwaltung von personenbezogenen Daten. Diese Daten können von den zuständigen Stellen, z.B. der Personalverwaltung, aktuell gehalten und Diensten für ihren Betrieb notwendigen Daten aus dem Identitätsmanagement bereitgestellt werden. Über entsprechende Schnittstellen können Daten aus den Personaldatenbanken mit dem Identitätsmanagement synchronisiert werden. Mitarbeiter des Instituts können über eine Web-Schnittstelle die über sie im Identitätsmanagement gespeicherten Daten einsehen und gegebenenfalls ändern. Beispielsweise könnten sie so an einer zentralen Stelle ihr Passwort für die verschiedenen Dienste ändern oder Einstellungen für Dienste, wie Abwesenheitsnotizen bei E-Mails, einrichten bzw. anpassen.

Das Identitätsmanagement dient so für verschiedene Dienste als Quelle für personenbezogene Daten, z.B. E-Mail oder die Telefonliste auf der Instituts-Webseite. In Zukunft sind zudem weitere Verknüpfungen mit dem Identitätsmanagement notwendig. Beispielsweise durch den Beitritt des Instituts zu den AAI-Föderationen des DFN bzw. CLARIN ist das IDS die Verpflichtungen eingegangen, Benutzerdaten innerhalb eng angelegter

Zeitraumen auf dem neuesten Stand zu halten und ausgeschiedene Mitarbeiter zeitnah zu sperren. Weiterhin ist auch eine zentrale Verwaltung von Unix-Accounts, z.B. für Pool-Arbeitsplätze von Hilfskräften oder Zugriff von Mitarbeitern auf verschiedene interne Server, wünschenswert.

4.3 Betrieb und Wartung

Der Betrieb einer modernen IT-Infrastruktur ist eine anspruchsvolle Aufgabe und erfordert sorgfältige Planung, um einen robusten und störungsfreien Betrieb zu gewährleisten. Die Komponenten der Infrastruktur lassen sich grob in grundlegende Netzwerkinfrastruktur (Internet-Verbindung, lokales Netzwerk, zugehörige Dienste wie DHCP, DNS und WINS), Server (intern, sowie extern), allgemeine Infrastruktur (Drucker etc.), Arbeitsplatzrechner, Telefonanlage und weitere Anlagen (wie z.B. Video-Konferenz-Systeme) einteilen.

Die Komponenten der Infrastruktur sollten entsprechend ihrer Wichtigkeit nach bewertet und Konzepte für Ausfälle entworfen werden. Besonders kritische Komponenten (wie z.B. Internet-Verbindung, oder zentrale Netzwerkdienste) sollten durch redundante Systeme gesichert werden (z.B. Hot-Standby).

Besonders wichtig für einen robusten Betrieb sind Monitoring und Backup. Das Monitoring dient dazu, verschiedene Dienste und Systeme zu überwachen und bei Problemen entsprechende Warnmeldungen zu generieren. So kann z.B. der ordnungsgemäße Betrieb der Internet-Anbindung oder zentrale Server und Dienste (wie der Web-Auftritt des Instituts) kontrolliert werden. Des Weiteren kann ein Monitoring-System Warnungen versenden, wenn Störungen auftreten, die jedoch nicht unmittelbar zum Ausfall eines Systems führen und so oft, auch über einen längeren Zeitraum, übersehen werden könnten, wie beispielsweise der Ausfall einer Festplatte in einem RAID-Verbund eines Storage-Subsystems oder die Notwendigkeit des Austauschs von Batterien in einer Unterbrechungsfreien Stromversorgung (USV). Eine weitere Schlüsselkomponente für den sicheren Betrieb ist ein Backup-System. Zum einen dient ein Backup dazu, um im Falle eines schwerwiegenden Hardware-Defekts die Daten des betroffenen Systems wieder herstellen zu können. Zum anderen ist ein Backup unerlässlich, um bei fehlerhafter Bedienung eines Systems, z.B. versehentliches Löschen wichtiger Dateien oder Beschädigung von Daten durch eine Fehlfunktion eines Programms, einen Datenverlust zu vermeiden. Wichtige zentrale Systeme bzw. Daten sollten daher täglich gesichert werden. Ein Backup aller Arbeitsplatzrechner ist aus verschiedenen Gründen problematisch, da nicht alle Rechner nachts, wenn normalerweise das Backup durchgeführt wird, in Betrieb sind. Daher sollte beispielsweise für Mitarbeiter jeweils ein Netzlaufwerk angeboten werden, auf dem sie eigenverantwortlich Daten ablegen können. Der Server, auf dem die Daten der Netzlaufwerke vorgehalten werden, wird in das reguläre Backup eingebunden und so die Daten täglich gesichert.

Um den administrativen Aufwand zu minimieren, ist die Einführung einer an die Bedürfnisse des Instituts angepassten, standardisierten Softwareplattform für Serversysteme und Arbeitsplatzrechner unerlässlich. Besonders für Serversysteme ist dies relevant, da auf diese Weise ein "Wildwuchs" von verschiedenen, nicht-konsistenten

Software-Installationen vermieden und der Aufwand für die Pflege der Systeme reduziert wird. Eine zentrale Verteilung von Software-Updates für das Betriebssystem als auch Anwendungssoftware ist auch für die Arbeitsplatzrechner ermöglicht es, beispielsweise besonders kritische Software, wie z.B. Internet-Browser, zeitnah zu aktualisieren. Wünschenswert wäre auch die Unterstützung einer modernen Linux-Distribution als alternatives Betriebssystem für Arbeitsplatzrechner wie etwa für Pool-Arbeitsplätze von Hilfskräften.

Weitere Aufgaben bestehen in der zentralen Hardware-Beschaffung und der Planung des Lebenszyklus von Systemen. Hierzu gehören sowohl die Einführung, der Betrieb und die Wartung, als auch die Außerbetriebnahme inklusive der eventuellen Migration von Diensten auf die neuen Systeme. Hierdurch kann die aufwändige Wartung und Pflege von Altsystemen und der Betrieb paralleler Infrastrukturen vermieden werden. Des Weiteren ermöglicht eine langfristige Planung der Hardwarezyklen eine größere Planungssicherheit bei der Beschaffung und ermöglicht die Aushandlung von längerfristigen Rahmenverträgen mit Lieferanten. Aus der Sicht der Verwaltung ist insbesondere der letztgenannte Punkt essentiell, um einerseits bessere finanzielle Konditionen mit den Händlern aushandeln und andererseits die vom Geldgeber geforderte Transparenz auf unkomplizierte Weise sicherstellen zu können.

In den letzten Jahren sind die Energiekosten, insbesondere die Kosten für Strom, signifikant gestiegen. Da der Anteil der IT-Infrastruktur am Gesamtstromverbrauch des Instituts nicht unerheblich ist, können Einsparungen dort auch erhebliche finanzielle Einsparungen nach sich ziehen. Im Rahmen der Klimaschutzdebatte ist unter anderem der Begriff *Green IT* geprägt worden. Green IT gliedert sich grob in zwei Themenbereiche: zum einen die Verwendung von energieeffizienten Geräten und Arbeitsmitteln und zum anderen die Einsparung von Energie durch die Nutzung von IT, z.B. durch das Durchführen von Videokonferenzen zur Vermeidung von Dienstreisen. Bei der Neuanschaffung von IT-Systemen sollte zukünftig vermehrt auf deren Energieeffizienz geachtet und die bestehende Infrastruktur auf mögliches Einsparungspotential untersucht werden. Beispielsweise durch Konsolidierung von Systemen als virtuelle Server kann die Anzahl der physikalischen Rechner verringert und dadurch unter anderem auch die Kühlung des Rechenzentrums optimiert werden. Im Rahmen einer IT-Gesamtstrategie sollte daher auch die IT-Infrastruktur im Hinblick auf Green IT untersucht und entsprechenden Maßnahmen ergriffen werden, denn der positiven Außenwirkung in Hinblick auf Nachhaltigkeit gibt es finanzieller Seite ein signifikantes Einsparpotential bei den Energiekosten für das Institut.

4.4 Sicherheit

Die Aufrechterhaltung eines angemessenen Sicherheitsniveaus wird beim Betrieb von IT-Infrastruktur leider häufig vernachlässigt. Dies kann vielfältige Gründe haben, wie z.B. fehlende Ressourcen, knappe Budgets, fehlendes bzw. unzureichendes Sicherheitsbewusstsein und nicht zuletzt die steigende Komplexität der Systeme. Auch wenn ein geisteswissenschaftliches Institut auf den ersten Blick ein eher "unattraktives" Ziel für einen Angriff zu sein scheint, sollte man sich nicht in falscher Sicherheit wiegen. Bot-

net-Betreibern geht es hauptsächlich um Quantität, d.h. möglichst viele Rechner zu infizieren, um beispielsweise mehr Spam versenden zu können. Aber auch gezielte Angriffe sind nicht auszuschließen. Weiterhin besitzt das Institut eine Reihe von Ressourcen, die aus verschiedensten Gründen (z.B. Datenschutz, Urheberrechte Dritter, ...) insbesondere vor unberechtigtem Kopieren oder Vandalismus zu schützen sind.

Durch die in den letzten Jahren gewachsene Infrastruktur und auch aufgrund Anforderungen externer Projekte, wie CLARIN oder TextGrid, ist es notwendig, die bestehenden Maßnahmen zur IT-Sicherheit am IDS zu überprüfen und gegebenenfalls zu überarbeiten. Dazu könnte nach den Empfehlungen zum IT-Grundschutz des Bundesamts für Sicherheit in der Informationstechnik (BSI) vorgegangen werden. Entsprechend an die Gegebenheiten im IDS angepasst, können nach einer initialen Gefahrenabschätzung identifizierte Probleme zusammen mit den betroffenen Stellen gezielt angegangen und Maßnahmen schrittweise umgesetzt werden. Weiterhin sollte über die Schaffung der Funktion eines IT-Sicherheitsbeauftragten nachgedacht werden, die, analog zum Datenschutzbeauftragten, die Evaluation und Umsetzung von Sicherheitsmaßnahmen begleitet. Des Weiteren könnte das Institut auch eine Zertifizierung für die Umsetzung des IT-Grundschutzes anstreben. Dies könnte einerseits das Institut bei der sachgerechten Umsetzung des Grundschutzes unterstützen und sich andererseits auch positiv auf die Außenwahrnehmung des Instituts aus der Sicht von Kooperationspartnern oder Drittmittelgebern auswirken.

5. Referenzen

Wissenschaftsrat

Wissenschaftsrat (2011a): *Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften*. Drs. 10465-11. Berlin: Wissenschaftsrat.

Wissenschaftsrat (2011b): *Übergreifende Empfehlungen zu Informationsinfrastrukturen*. Drs. 10466-11. Berlin: Wissenschaftsrat.

DeReKo

Belica, C. et al. (2011): The morphosyntactic annotation of DEREKO: Interpretation, opportunities and pitfalls. In: Konopka, M./Kubczak, J./Mair, C./Šticha, F./Wassner, U. (eds.): *Grammatik und Korpora 2009. Dritte Internationale Konferenz. Mannheim, 22.-24.9.2009*. (= Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 1). Tübingen. Gunter Narr. 451-469.

Kupietz, M. et al. (2010a): The German Reference Corpus DeReKo: A primordial sample for linguistic research. In: Calzolari, Nicoletta et al. (eds.): *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*, 1848-1854. Internet: www.lrec-conf.org/proceedings/lrec2010/pdf/414_Paper.pdf.

Kupietz, M. et al. (2010b): The German Reference Corpus: New developments building on almost 50 years of experience. In: *Proceedings of the LREC 2010 Workshop "Language Resources: From Storyboard to Sustainability and LR Lifecycle Management"*. Valetta, Malta, 39-43.

CCDB

Belica, C. (2007): *Koorkkurrenzdatenbank CCDB – V3. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs*. Mannheim. Institut für Deutsche Sprache. Internet: <http://corpora.ids-mannheim.de/ccdb/>.

Keibel, H./Belica, C. (2007): CCDB: A Corpus-Linguistic Research and Development Workbench. In: *Proceedings of Corpus Linguistics 2007, Birmingham*. Internet: http://corpus.bham.ac.uk/corplingproceedings07/paper/134_Paper.pdf.

COSMAS

al-Wadi, D. (1994): *COSMAS – Ein Computersystem für den Zugriff auf Textkorpora. Version R.1.3-1. Benutzerhandbuch*. Mit einem Geleitwort von Prof. Dr. Gerhard Stickel. Mannheim: Institut für deutsche Sprache.

Bodmer, F. (2005): COSMAS II. Recherchieren in den Korpora des IDS. In: *Sprachreport 3/2005*, 2-5.

D-SPIN

Bankhardt, C. (2009): D-SPIN – Eine Infrastruktur für deutsche Sprachressourcen. In: *Sprachreport 1/2009*, 30-31.

Geyken, A./Klein, W. (2009): Deutsche Sprachressourcen-Infrastruktur. In: Berlin-Brandenburgische Akademie der Wissenschaften (Hg.): *Berlin-Brandenburgische Akademie der Wissenschaften – Jahrbuch 2008*. Berlin: Akademie Verlag, 336-337.

Lexik

Herberg, D./Kinne, M./Steffens, D. (2004): *Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen*. Unter Mitarbeit von Elke Tellenbach und Doris al-Wadi. (= Schriften des Instituts für Deutsche Sprache 11). Berlin/New York: de Gruyter.

Müller-Spitzer, C. (2008): The lexicographic portal of the IDS. Connecting heterogeneous lexicographic resources by a consistent concept of data modelling. In: *Proceedings of the 13th EURALEX International Congress. Euralex 2008*. Barcelona: Universitat Pompeu Fabra, 457-461.

Müller-Spitzer, C. (2010): OWID – A dictionary net for corpus-based lexicography of contemporary German. In: Dykstra, A./Schoonheim, T. (Hg.): *Proceedings of the XIV Euralex International Congress. Leeuwarden, 6-10 July 2010*. Leeuwarden: Fryske Akademy, 445-452.

Standards

ISO/FDIS 24619-2011: *Language resource management – persistent identification and sustainable access (PISA)*. Genf: International Organization for Standardization (ISO).

TextGrid

Kerzel, M./Mittelbach, J./Vitt, T. (2009): TextGrid – Virtuelle Arbeitsumgebung für die Geisteswissenschaften. In: *Künstliche Intelligenz 4/November 2009* (Themenheft "Kulturerbe und Künstliche Intelligenz"), 36-39.

Zielinski, A./Pempe, W./Gietz, P./Haase, M./Funk, S./Simon, C. (2009): TEI Documents in the grid. In: *Literary and Linguistic Computing* 24 (3), 267-279.

WissGrid

Enke, H. (2010): Grid: user perspectives. In: *Proceedings of SwissGrid Day 2010, November 30, 2010*. Internet: www.wissgrid.de/publikationen/presentations/SwissGrid-Day2010-WissGrid.pdf.

Ludwig, J. (2009): Long-term preservation of digital research data. In: *Proceedings of DESY Computing Seminar (DVSEM), November 23, 2009, Hamburg*. Internet: www.desy.de/dvsem/WS0910/ludwig_talk.pdf.

Ludwig, J. (2010): Diversity and interoperability of repositories in a grid curation Environment. In: *Proceedings of Open Repositories Conference, July 6, 2010, Madrid*. Internet: www.wissgrid.de/publikationen/presentations/2010-07-06--or2010-DiversityandInteroperabilityofRepositoriesinGridCurationEnvironment.pdf.

nestor

Neuroth, H./Oßwald, A./Scheffel, R./Strathmann, S./Jehn, M. (Hg.): *Nestor-Handbuch: eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 2.0, Juni 2009. Boizenburg: Hülsbusch.

BSI Grundschatz

Bundesamt für Sicherheit in der Informationstechnik (2008a): *BSI-Standard 100-1: Managementsysteme für Informationssicherheit (ISMS)*.

Bundesamt für Sicherheit in der Informationstechnik (2008b): *BSI-Standard 100-2: IT-Grundschatz-Vorgehensweise*.

Bundesamt für Sicherheit in der Informationstechnik (2008c): *BSI-Standard 100-3: Risikoanalyse auf der Basis von IT-Grundschatz*.

Bundesamt für Sicherheit in der Informationstechnik (2008d): *BSI-Standard 100-4: Notfallmanagement*.