

# Zwei grundlegende Fragen der digitalen Nachhaltigkeit: Wie können wir die heterogenen Forschungsfragen und die Community bei der Verfügbarmachung von Forschungsdaten miteinbeziehen?

**Odebrecht, Carolin**

carolin.odebrecht@hu-berlin.de

Humboldt-Universität zu Berlin, Deutschland

**Dreyer, Malte**

malte.dreyer@cms.hu-berlin.de

Humboldt-Universität zu Berlin, Deutschland

**Lüdeling, Anke**

anke.luedeling@hu-berlin.de

Humboldt-Universität zu Berlin, Deutschland

**Krause, Thomas**

krauset@hu-berlin.de

Humboldt-Universität zu Berlin, Deutschland

## Zielstellung

Digitale Nachhaltigkeit verstehen wir als eine komplexe Anforderung, die aus besonders vielen Blickwinkeln betrachtet werden kann und sollte. Wir stellen mit dem LAUDATIO-Projekt eine Möglichkeit vor, digitale Nachhaltigkeit herzustellen, in dem wir einen unabhängigen und freien Zugriff auf historische Korpora zum Zweck der Wiederverwendung für Forschungszwecke, die über diejenigen hinausgehen, für die die Daten ursprünglich gesammelt wurden, ermöglichen. Eine so definierte digitale Nachhaltigkeit realisieren wir, in dem wir eine Vielzahl an verschiedenen, teils sehr unterschiedlichen historischen Korpusdaten inklusive einer umfangreichen aber einheitlichen Dokumentation, die ihre Erschließbarkeit in Bezug auf konkrete Nutzungsszenarien gewährleistet, bereitstellen. Aus dieser Arbeit zeigt sich, dass eine interdisziplinäre und insbesondere lokale institutionelle Zusammenarbeit

mit den Korpuserstellern notwendig ist, die einen engen kommunikativen Austausch von Zielen und Anforderungen diesbezüglich sowie eine Identifikation von möglichen Kooperationen erst ermöglicht. Digitale Nachhaltigkeit definiert sich dementsprechend aus der jeweiligen Perspektive der Daten, der Dokumentation und der Institutionen ein wenig anders. Diese drei Perspektiven und deren Zusammenspiel in Bezug auf die digitale Nachhaltigkeit wollen wir anhand von historischen Korpora (vgl. Claridge 2008; Kytö 2011; Gippert und Gehrke 2015) diskutieren und an einem Best-Practice-Beispiel einer interdisziplinären Zusammenarbeit der Philosophischen Fakultät und dem Rechenzentrum der Humboldt-Universität zu Berlin (HU) erklären.

## Heterogenität von Forschungsdaten

Die vielfältige Datenlandschaft in den Geisteswissenschaften stellt eine große Herausforderung in Bezug auf die digitale Nachhaltigkeit dar, da es unterschiedliche Datenmodelle und -formate und verschiedene Aufbereitungs- und Analyseverfahren gibt, die alle aus immer neuen Nutzungsszenarien und ihren Forschungsfragen, resultieren. In diesem Sinne verstehen wir die Arbeit mit Korpora als eine innovative, fortlaufende, wissenschaftliche Arbeit, die sich nicht ausschließlich auf existierende Standards stützen kann (auch wenn solche Standards natürlich immer beachtet werden müssen). Es werden daher zusätzlich andere, neue Forschungsdatenmodelle und -formate sowie Ressourcentypen entwickelt und auf die unterschiedlichsten Weisen weiter- und wiedergenutzt. Beispielsweise werden in den Projekten, die LAUDATIO unterstützt, eine Vielzahl an allein XML-basierten Formaten (nach z.B. Dipper 2005; Schmidt und Wörner 2009; Romary et al. 2015; TEI Consortium 2015) sowie CSV-basierte Formate (nach z.B. Nivre, Hall und Nilsson 2004; Krause und Zeldes 2016) für die Erstellung von historischen Korpora genutzt. Daneben finden auch und graphbasierten Lösungen (nach z.B. Ide und Suderman 2014) sowie proprietäre Formate, wie bei Ágel und Hennig (2007), und immer häufiger JSON-basierte Formate wie bei Vertan et al. (2016) Anwendung. Zusätzlich dazu können Korpora in mehreren Formaten vorliegen, vor allem wenn Korpora in einer Mehrebenenarchitektur (vgl. Romary und Ide 2004; Lüdeling 2012) entworfen und verschiedene Formate für unterschiedliche Abschnitte des Forschungsdatenzyklus (vgl. Rümpel 2011) genutzt werden.

Der innovative Charakter der Korpuserstellung zeigt sich neben den verwendeten Formaten auch in den Annotationsrichtlinien: Einige Formate wie TIGER-XML (Romary et al. 2015) und tcf (Heid et al. 2010) legen die Anzahl und die Bedeutung der Annotationen inklusive ihrer Tagsets fest, andere wie TEI-XML (TEI Consortium 2015) lassen Spielraum in der Serialisierung und in der Anwendung, in dem es beispielsweise mehrere

valide Möglichkeiten gibt, Autoren in einem Dokument auszuweisen. Wieder andere Formate wie EXMARaLDA-XML (Schmid und Wörner 2009) oder PAULA-XML (Dipper 2005) geben keinerlei solcher Beschränkungen vor.

Insbesondere bei solchen Formaten mit größtmöglichem Spielraum für Interpretationen in Form von Annotationen zeigt sich die Vielfältigkeit der korpusbasierten Forschung. Ein typisches Annotationsbeispiel für historische Korpora sind Normalisierungen. Viele Korpora besitzen eigene Richtlinien für die Normalisierung, vgl. zum Beispiel für historisches Deutsch Jurish (2010), Bollmann et al. (2012); Odebrecht et al., (eingereicht). Auch für verschiedene Forschungsfragen wesentliche Kategorisierungen wie Wortarten gibt es annähernd für jedes historische Korpus eine eigene Lösung. So werden beispielsweise de-facto-Standards wie das STTS (Schiller et al. 1999) jeweils für ein Korpus angepasst, z.B. Dipper et al. (2013) oder Fürstinnenkorrespondenzkorpus.

Diese Beispiele zeigen, dass sich Korpora desselben Formats und ein Korpus, das in verschiedenen Formaten vorliegt, in Bezug auf ihre Annotationen stark unterscheiden können.

Ausgehend von dieser Datenlage erscheint die Nachvollziehbarkeit des Lebenszyklus der Korpora als ein weiterer wesentlicher Faktor für die digitale Nachhaltigkeit. Welche der bisher genutzten Formate und Annotationskonzepte sich langfristig durchsetzen, welche Formate wie technisch unterstützt und welche neuen Lösungen entwickelt werden, hängt dann im Wesentlichen von der Entwicklung der korpusbasierten Forschung ab. Daher setzen wir eine Archivierung und Dokumentation aller verwendeten Formate und Annotationen eines Korpus in LAUDATIO um.

## Metadaten

Über einheitliche extensive Metadaten dieser heterogenen Korpusdaten und deren Lebenszyklus kann eine umfangreiche Korpusdokumentation sowie eine einheitliche Suche und ein gezielter Zugriff über eine Plattform auf die Forschungsdaten erstellerunabhängig gewährleistet werden (Bird und Simons 2001; Broeder et al. 2010; Burnard 2013; Hedges et al. 2013; Odebrecht et al. 2015). Die relevanten Kriterien für die Dokumentation und die Suche leiten sich aus den Wiederverwendungsszenarien ab (Odebrecht 2014). Um eine überfachliche Suche zu ermöglichen, wird in LAUDATIO eine technisch-abstrakte Modellierung der Metadaten eingesetzt, um die fachspezifischen Konzepte von Korpora überfachlich abzubilden (Odebrecht 2015). Neben den deskriptiven Metadaten sind für eine nachhaltige Vorhaltung von Forschungsdaten auch administrative, strukturelle, technische und Archivierungsmetadaten relevant (vgl. Xie und Matusiak 2016; Solodovnik 2011; NISO 2004), die für eine technische Infrastruktur berücksichtigt werden müssen,

um die Nachvollziehbarkeit über längere Zeiträume und wechselnde Anwendergruppen hinweg gewährleisten zu können. So ermöglicht das einheitliche Metadatenmodell eine extensive und transparente Informationsarchitektur für die unterschiedlichen Ressourcen, was wiederum ein Baustein für deren digitale Nachhaltigkeit darstellt.

## Institutionelle (Zusammen-)Arbeit

Bedingt durch die unterschiedlichen Sichtweisen und Anforderungen zur digitalen Nachhaltigkeit aus den Fachdisziplinen und durch die technische Infrastruktur sind Lösungen nur im Team und in Zusammenarbeit unterschiedlicher Kompetenzen plan- und erstellbar. Im Beispiel ist für den erstellerunabhängigen Zugang zu historischen Korpora die enge Zusammenarbeit zwischen den FachwissenschaftlerInnen und LAUDATIO erforderlich. Der Computer- und Medienservice (CMS) und die Arbeitsgruppe Korpuslinguistik am Institut für deutsche Sprache und Linguistik der HU setzen mit dem LAUDATIO-Projekt einen Schwerpunkt auf eine enge institutionelle Zusammenarbeit mit den verschiedenen Arbeitsgruppen der philosophischen Fakultät, in dem es die Anforderungen für die digitale Nachhaltigkeit von Forschungsdaten der Fakultät mit den ForscherInnen erarbeitet und umsetzt und in Bezug auf die Hochschule als Ganzes in einen Entwicklungsrahmen einbettet (Dreyer und Vollmer 2016). Der Betrieb des LAUDATIO-Repositoriums wird nach Ende des Projektes, in dem Entwicklungen und Anpassungen am System vorgenommen werden, weiterhin durch das CMS sowie durch die Arbeitsgruppe Korpuslinguistik am Institut für deutsche Sprache und Linguistik gewährleistet. Um die heterogene Datenlandschaft zu verstehen, die umfassende Dokumentation zu erstellen und die neuen Entwicklungen aufzunehmen, erweist sich eine lokale Zusammenarbeit über die Fakultäten hinweg als sehr vorteilhaft. So bestehen enge Kooperationen unter anderem mit Projekten

- der Sprachgeschichte: DDD-AD, HIPKON, DDB, Fürstinnenkorrespondenzkorpus
- der Germanistik: Märchenkorpora und RIDGES
- der Slawistik: "Korpuslinguistik und diachrone Syntax: Subjektkasus, Finalität und Kongruenz in slavischen Sprachen"

die jeweils sehr unterschiedliche Arbeitsweisen und Zielrichtungen haben. Durch diese Synergiebildung können Projekte ohne gesonderte Finanzierung oder Ressourcen nach ihrem Projektende durch eine genaue Kenntnis der Forschungsdatenlandschaft in einer Institution identifiziert und unterstützt werden. Die durch die Förderer (z.B. Deutsche Forschungsgemeinschaft 2015) oder Universitäten (z.B. Humboldt-Universität zu Berlin 2014) vorgegeben Richtlinien zur Veröffentlichung und Archivierung von Projektergebnissen können so ebenfalls berücksichtigt werden.

Mit diesem Ansatz können gleichzeitig zwei Ziele erreicht werden: Die Projekte müssen sich keinen umfangreichen semantischen Anforderungen unterwerfen und können sich frei ausdrücken. Gleichzeitig können sie ihre fachspezifischen Anforderungen direkt an die Arbeitsgruppe Korpuslinguistik bzw. an den CMS richten. Andersherum können die Korpusprojekte in Umfragen zur gewünschten Anforderungen und Softwarelösungen direkt befragt und in die Entwicklung mit einbezogen werden. Auf diese Art findet ein Community-Aufbau statt, der sich nicht nur über gemeinsame technische Plattformen definiert, sondern rein über die inhaltliche Gemeinsamkeit der Arbeit mit historischen Korpora und somit auch über disziplinäre Grenzen hinweg.

## Einordnung und Schlussfolgerungen

Der hier vorgestellte Weg, digitale Nachhaltigkeit von Forschungsdaten zu ermöglichen, stützt sich auf eine Spezialisierung auf einen bestimmten Typ von Forschungsdaten - historisches Korpus - und grenzt sich so von Ansätzen wie Zenodo und dem Virtual Language Observatory (Van Uytvanck 2012) ab, die keine deutliche Eingrenzung hinsichtlich der Daten und deren Nutzungsszenarien machen.

Weiterhin zielt unsere Strategie auf die Unterstützung der Diversität der genutzten Formate und Konzepte, die sich von den TEI spezialisierten Ansätzen wie dem Deutschen Textarchiv (Geyken 2013) und Textgrid (Hedges et al. 2013) unterscheiden.

Die Korpusersteller selbst nutzen LAUDATIO auch mehr und mehr, um ihre eigenen neuen Versionen der historischen Korpora und damit ihren wissenschaftlichen Fortschritt zu veröffentlichen. Dass unser Ansatz, sich auf die erstellerunabhängige Wiederverwendung von Korpora als eine Strategie für digitale Nachhaltigkeit zu fokussieren, auch außerhalb der Institution funktioniert, zeigt Dumont (2016).

Um die unterschiedlichen Entwicklungen und Innovationen bei der Korpuserstellung zu identifizieren, kennenzulernen und zu dokumentieren, ist eine enge, auf eine rein fachliche Ebene bezogene Zusammenarbeit mit der Community notwendig, die wir angefangen haben, im Rahmen der Philosophischen Fakultät der HU aufzubauen.

## Fußnoten

1. LAUDATIO steht für Long-term Access and Usage of Deeply Annotated Information. [www.laudatio-repository.org](http://www.laudatio-repository.org). Zugriff am 16.08.2016.
2. Ágel, Vilmos; Hennig, Mathilde; KAJUK (Version 1.1), Justus-Liebig-Universität Gießen. <http://www.uni-giessen.de/kajuk/index.htm>.
3. Wie es zum Beispiel mit den Element <author> möglich ist, vgl. <http://www.tei-c.org/release/doc/tei-p5-doc/de/html/ref-author.html> Zugriff am 19.08.2016.

4. Donhauser, Karin; Gippert, Jost; Lühr, Rosemarie; ddd-ad (Version 0.1), Humboldt-Universität zu Berlin. <https://referenzkorpusaltdeutsch.wordpress.com/>. <http://hdl.handle.net/11022/0000-0000-7FC2-7>
5. Lühr, Rosemarie; Faßhauer, Vera; Prutscher, Daniela; Seidel, Henry; Fuerstinnenkorrespondenz (Version 1.1), Universität Jena, DFG. <http://www.indogermanistik.uni-jena.de/Web/Projekte/Fuerstinnenkorrr.htm>.
6. <http://www.deutschiachrondigital.de/> Zugriff am 16.08.2016.
7. <https://www.linguistik.hu-berlin.de/de/institut/professuren/sprachgeschichte/forschung/sfb632-informationsstruktur> Zugriff am 16.08.2016.
8. <http://korpling.german.hu-berlin.de/ddb-doku/index.htm> Zugriff am 16.08.2016.
9. [http://dwee.eu/Rosemarie\\_Luehr/?Projekte\\_\\_DFG-Projekte\\_\\_Fruehneuzeitliche\\_Fuerstinnenkorrespondenz\\_im\\_mitteldeutschen\\_Raum](http://dwee.eu/Rosemarie_Luehr/?Projekte__DFG-Projekte__Fruehneuzeitliche_Fuerstinnenkorrespondenz_im_mitteldeutschen_Raum) Zugriff am 16.08.2016.
10. Die Erstellung des Korpus erfolgte im Rahmen von universitärer Lehre <http://www.textbewegung.de/lehre.html> Zugriff am 16.08.2016.
11. <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/ridges-projekt/> Zugriff am 16.08.2016.
12. <https://www.slawistik.hu-berlin.de/de/member/meyerrol/subjekte/corpora> Zugriff am 16.08.2016.
13. Zusätzlich helfen uns auch Evaluationen mit Kooperationspartner durch Dritte, die angebotenen Lösungen zu verbessern (z.B. Stiller et al. 2016).
14. <http://zenodo.org> Zugriff am 19.08.2016.

## Bibliographie

- Ágel, Vilmos / Hennig, Mathilde (eds.) (2007): *Zugänge zur Grammatik der gesprochenen Sprache*. Germanistische Linguistik 269. Tübingen: Niemeyer.
- Bird, Steven / Simons, Gary (2001): „The OLAC Metadata Set and Controlled Vocabularies“, in: *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics* 7–18 arXiv:cs/0105030v1.
- Bollmann, Marcel / Dipper, Stefanie / Krasselt, Julia / Petran, Florian (2012): „Manual and semi-automatic normalization of historical spelling - case studies from Early New High German“, in: *Proceedings of KONVENS 2012* 342–350 [http://www.oegai.at/konvens2012/proceedings/51\\_bollmann12w/](http://www.oegai.at/konvens2012/proceedings/51_bollmann12w/) [letzter Zugriff 22 August 2016].
- Broeder, Daan / Kemps-Snijders, Marc / Van Uytvanck, Dieter / Windhouwer, Menzo / Withers, Peter / Wittenburg, Peter / Zinn, Claus (2010): „A data category registry- and component-based metadata framework“, in: *Proceedings of LREC 2010* 43–47.
- Burnard, Lou (2013): „The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure“, in: *Journal of the Text Encoding Initiative* 5: 1–13 10.4000/jtei.811.

- Claridge, Claudia** (2008): „Historical Corpora“, in: Lüdeling, Anke / Kytö, Merja (eds.): *Corpus Linguistics. An International Handbook* 1. Berlin: De Gruyter 242–259.
- Dipper, Stefanie** (2005): „XML-based Stand-off Representation and Exploitation of Multi-Level Linguistic Annotation“, in: *Proceedings of Berliner XML Tage* 39–50.
- Dipper, Stefanie / Donhauser, Karin / Klein, Thomas / Linde, Sonja / Müller, Stefan / Wegera, Klaus-Peter** (2013): „HiTS. Ein Tagset für historische Sprachstufen des Deutschen“, in: Zinsmeister, Heike / Heid, Ulrich / Beck, Kathrin (eds.): *Das Stuttgart-Tübingen Wortarten-Tagset: Stand und Perspektiven*. Journal for Language Technology and Computational Linguistics 28(1) 85–137.
- Deutsche Forschungsgemeinschaft** (2015): *Leitlinien zum Umgang mit Forschungsdaten*. Bonn: DFG. [http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\\_forschungsdaten.pdf](http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf) [letzter Zugriff 22. August 2016].
- Dreyer, Malte / Vollmer, Andreas** (2016): „An Integral Approach to Support Research Data Management at the Humboldt-Universität zu Berlin“, in: *Proceedings of the European University Information Systems organisation* 320–327.
- Dumont, Stefan** (2016): „Fürstinnenkorrespondenzen. Experiment einer Nachnutzung“, in: *Entwicklung und Nutzung interdisziplinärer Repositorien für historische textbasierte Korpora. DHd 2016 Workshop* <http://www.laudatio-repository.org/laudatio/workshop-dhd2016/> [letzter Zugriff 22. August 2016].
- Gippert, Jost / Gehrke, Ralf** (eds.) (2015): *Historical Corpora*. Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 5. Tübingen: Narr.
- Geyken, Alexander** (2013): „Wege zu einem historischen Referenzkorpus des Deutschen. das Projekt Deutsches Textarchiv“, in: Hafemann, Ingelore (ed.): *Perspektiven einer corpusbasierten historischen Linguistik und Philologie*. Internationale Tagung des Akademienvorhabens "Altägyptisches Wörterbuch" an der BBAW. Thesaurus Linguae Aegyptiae, 4: 221–234.
- Hedges, Mark / Neuroth, Heike / Smith, Kathleen M. / Blanke, Thomas / Romary, Laurent / Küster, Marc / Illingworth, Malcom** (2013): „TextGrid, TEXTvire, and DARIAH. Sustainability of Infrastructure for Textual Scholarship“, in: *Journal of the Text Encoding Initiative* 5: 1–13.
- Heid, Ulrich / Schmid, Helmut / Eckart, Kerstin / Hinrichs, Erhard W.** (2010): „A Corpus Representation Format for Linguistic Web Services. The D-SPIN Text Corpus Format and its Relationship with ISO Standards“, in: *Proceedings of the Seventh International Conference on Language Resources and Evaluation* 494–499.
- Humboldt-Universität zu Berlin** (2014): *Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin*. Unter Mitarbeit von Elena Simukovic <https://www.cms.hu-berlin.de/de/ueberblick/projekte/dataman/hu-fdt-policy/view> [letzter Zugriff 6. Juni 2016].
- Ide, Nancy / Sudermann, Keith** (2014): „The Linguistic Annotation Framework. a standard for annotation interchange and merging“, in: *Language Resources and Evaluation* 48 (3): 395–418.
- Jurish, Bryan** (2010): „More than Words: Using Token Context to Improve Canonicalization of Historical German“, in: *Journal for Language Technology and Computational Linguistics* 25 (1): 23–40.
- Krause, Thomas / Zeldes, Amir** (2016): „ANNIS3. A new architecture for generic corpus query and visualization“, in: *Digital Scholarship in the Humanities* 31 (1): 118–139 [10.1093/llc/fqu057](https://doi.org/10.1093/llc/fqu057).
- Kytö, Merja** (2011): „Corpora and historical linguistics“, in: *Revista Brasileira de Linguística Aplicada* 11: 417–457 [10.1590/S1984-63982011000200007](https://doi.org/10.1590/S1984-63982011000200007).
- Lüdeling, Anke** (2012): „A corpus-linguistics perspective on language documentation, data, and the challenge of small corpora“, in: Seifart, Frank / Haig, Geoffrey / Himmelmann, Nikolaus P. / Jung, Dagmar / Margetts Anna / Trilsbeek, Paul (eds.): *Potentials of Language Documentation. Methods, Analyses, and Utilization* 4. Language Documentation & Conservation Special Publication 3. Hawaii: University of Hawai‘i Press 32–38.
- NISO** (2004): *Understanding Metadata*. Bethesda: NISO Press <http://www.niso.org/publications/press/UnderstandingMetadata.pdf> [letzter Zugriff 13. Februar 2015].
- Nivre, Joakim / Hall, Johan / Nilsson, Jens** (2004): „Memory-Based Dependency Parsing“, in: *Proceedings of the Eighth Conference on Computational Natural Language Learning* 49–56.
- Odebrecht, Carolin / Belz, Malte / Zeldes, Amir / Lüdeling, Anke / Krause, Thomas** (eingereicht): *RIDGES Herbology - Designing a Diachronic Multi-Layer Corpus*. Vorversion [https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/ridges-projekt/download-files/pubs/odebrechtaleingereicht\\_ridgesherbology.pdf/at\\_download/file](https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/ridges-projekt/download-files/pubs/odebrechtaleingereicht_ridgesherbology.pdf/at_download/file) [letzter Zugriff 22. August 2016].
- Odebrecht, Carolin** (2014): „Modeling Linguistic Research Data for a Repository for Historical Corpora“, in: *DH2016: Book of Abstracts* 284–285.
- Odebrecht, Carolin** (2015): „Interdisziplinäre Nutzung von Forschungsdaten mithilfe einer technisch-abstrakten Modellierung“, in: *DHd 2015: Von Daten zu Erkenntnissen* [https://dh2014.files.wordpress.com/2014/07/dh2014\\_abstracts\\_proceedings\\_07-11.pdf](https://dh2014.files.wordpress.com/2014/07/dh2014_abstracts_proceedings_07-11.pdf) [letzter Zugriff 22. August 2016].
- Odebrecht, Carolin / Krause, Thomas / Lüdeling, Anke** (2015): „Austausch von historischen Texten verschiedener Sprachen über das LAUDATIO-Repository“, in: *DGfS-CL Poster Session. 37. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft* <http://asvdoku.informatik.uni-leipzig.de/dgfs2015cl-ps/index.html> [letzter Zugriff 22. August 2016].

**Romary, Laurent / Ide, Nancy** (2004): „International standard for a linguistic annotation framework“, in: *Natural Language Engineering* 10 (3-4): 211–225.

**Romary, Laurent / Zeldes, Amir / Zipser, Florian** (2015): „<tiger2/>: serialising the ISO SynAF syntactic object model“, in: *Language Resources and Evaluation* 49 (1): 1–18.

**Rümpel, Stefanie** (2011): „Der Lebenszyklus von Forschungsdaten“, in: Büttner, Stephan / Hobohm, Hans-Christoph / Müller, Lars (eds.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen 25–34.

**Schmidt, Thomas / Wörner, Kai** (2009): „EXMARaLDA. Creating, analysing and sharing spoken language corpora for pragmatic research“, in: *Pragmatics* 19 (4): 565–582.

**Schiller, Anne / Teufel, Simone / Stöckert, Christine / Thielen, Christine** (1999): „Guidelines für das Tagging deutscher Textkorpora mit STTS“, in: Universität Tübingen (ed.): Seminar für Sprachwissenschaft. Technischer Report. <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf> [letzter Zugriff 22. August 2016].

**Solodovnik, Iryna** (2011): „Metadata issues in Digital Libraries. key concepts and perspectives“, in: *Italian Journal of Library, Archives and Information Science* 2 (2): 4663-1–4663-27. 10.4403/jlis.it-4663.

**Stiller, Juliane / Thoden, Klaus / Zielke, Dennis** (2016): „Usability in den Digital Humanities am Beispiel des LAUDATIO-Repositoriums“, in: *DHd 2016: Modellierung - Vernetzung - Visualisierung* 244–247.

**TEI Consortium** (2015): *TEI P5: Guidelines for Electronic Text Encoding and Interchange* (2.9.1). <http://www.tei-c.org/Guidelines/P5/> [letzter Zugriff 15. November 2015].

**Van Uytvanck, Dieter / Stehouwer, Herman / Lempen, Lari** (2012): „Semantic metadata mapping in practice: the Virtual Language Observatory“, in: *Proceedings of LREC 2012* 1029–1033.

**Vertan, Cristina / Ellwardt, Andreas / Hummerl, Susanne** (2016): „Ein Mehrebenen-Tagging-Modell für die Annotation altäthiopischer Texte“, in: *DHd 2016: Modellierung - Vernetzung - Visualisierung* 258–261.

**Xie, Iris / Matusiak, Krystyna** (2016): *Discover Digital Libraries. Theory and Practice*. Oxford: Elsevier.