

Vanya Dimitrova und Heike Renner-Westermann

Das Linguistik-Portal: Übergang von einer Virtuellen Fachbibliothek zu einem Fachinformationsdienst

Linguistics portal: Transition from virtual library to specialist information service

<http://doi.org/10.1515/bd-2018-0033>

Zusammenfassung: Das Linguistik-Portal bietet seit 2013 eine überregionale Plattform für wissenschaftliche Informationen zu allen Bereichen der Sprachwissenschaft. Ursprünglich konzipiert als Virtuelle Fachbibliothek, wird das Linguistik-Portal gegenwärtig zu einem Fachinformationsdienst ausgebaut. Dieser Artikel stellt die bisherige Entwicklung des Portals chronologisch vor und beschreibt detaillierter die anvisierte Erweiterung der Funktionalität und des Serviceangebots. Der Fokus liegt dabei auf fachspezifischen Ressourcentypen wie Sprachkorpora und der Anwendung innovativer Technologien im Bereich Linked Open Data.

Schlüsselwörter: Linguistik-Portal, Forschungsdaten, Linked Open Data, Virtuelle Fachbibliothek, Fachinformationsdienst Linguistik

Abstract: Since 2013, the linguistics portal offers a national platform for scientific information concerning all fields of linguistics. Originally designed as a virtual library, the linguistics portal is now being expanded into a specialist information service. This article presents the portal's previous development in chronological order and describes more detailed how the features and service offers are planned to be expanded. The focus is on subject-specific types of resources like language corpora and the application of innovative technologies in the field of Linked Open Data.

Keywords: linguistics portal, research data, Linked Open Data, virtual library, specialist information service linguistics

Vanya Dimitrova: v.dimitrova@ub.uni-frankfurt.de

Heike Renner-Westermann: h.renner-westermann@ub.uni-frankfurt.de

Das Linguistik-Portal startete als Virtuelle Fachbibliothek (ViFa) mit dem Ziel, ein Rechercheinstrument für die Suche nach wissenschaftlichen Informationen zur Allgemeinen Linguistik, zur Allgemeinen und Vergleichenden Sprachwissenschaft sowie zu einzelphilologischen Sprachwissenschaften anzubieten. In den nächsten Jahren soll das Portal die Funktion eines Fachinformationsdienstes (FID) übernehmen, indem die Erschließung von Informationsquellen durch neue Methoden weiter ausgebaut und zusätzliche wissenschaftsnahe Services angeboten werden.

1 Linguistik-Portal

Das Linguistik-Portal wurde mit Unterstützung der DFG an der Universitätsbibliothek Frankfurt a. M. erstellt und ging 2013 online unter www.linguistik.de. In der ersten Projektphase 2012–2014 galt es, eine klassische ViFa mit den üblichen Modulen aufzubauen: Verzeichnisse wurden erstellt zu thematischen Links, Forschungsprojekten und Online-Wörterbüchern. Linguistisch relevante, elektronische Zeitschriften und Online-Datenbanken wurden intellektuell aus EZB bzw. DBIS selektiert und hinsichtlich behandelter Sprache und linguistischer Thematik erschlossen. Zur Erweiterung der Aufsatztitel-Datenbank Online Contents Linguistik wurden die Inhaltsverzeichnisse ausgewählter Zeitschriften gescannt und zu rudimentären Titelaufnahmen aufbereitet. Die Verzeichnisse zu thematischen Links, elektronischen Zeitschriften, Datenbanken und Wörterbüchern flossen gemeinsam mit Bibliothekskatalogen, Repositorien und Bibliografien in eine indexbasierte Metasuche ein¹.

Das Projekt fand in Kooperation mit zwei Projektpartnern statt. Die Arbeitsgruppe LinseLinks der Universität Duisburg-Essen war hauptsächlich an der Erarbeitung des Link-Verzeichnisses und des zugrundeliegenden Sacherschließungskonzepts beteiligt. Letzteres besteht im Großen und Ganzen aus einer Verzahnung der thematischen Klassifikation der LinseLinks mit der Klassifikation der Bibliography of Linguistic Literature (BLL²). Der zweite Projektpartner, das Institut für Deutsche Sprache (IDS) Mannheim, arbeitete an der Erweiterung von

¹ Zu den integrierten Quellen gehören im Einzelnen: Fach-OPAC Linguistik der UB Frankfurt; Katalog der Bibliothek des IDS Mannheim; Online Contents Linguistik; IDS-Publikationsserver; Linguistik-Repository UB Frankfurt; Linguistik-Auswahl aus BASE (Bielefeld Academic Search Engine); Bibliography of Linguistic Literature (BLL); frei verfügbare elektronische Dissertationen zu Sprachthemen aus der Deutschen Nationalbibliografie.

² <http://www.blldb-online.de> [Zugriff: 05.01.2018].

Online Contents Linguistik, dem Link-Verzeichnis und dem Wörterbuch-Verzeichnis mit und war alleinverantwortlich für den Aufbau des IDS-Publikationsservers.

In der zweiten von der DFG bewilligten Förderphase 2015–2017 wurde der Aufbau der klassischen ViFa fortgesetzt: Die unterschiedlichen Verzeichnisse wurden gepflegt und selektiv erweitert. Es wurden Online-Tutorials zur Nutzung des Portals und einzelner Module erstellt. Die Indexsuche wurde um zwölf Kataloge und Bibliografien³ ergänzt sowie für Smartphones optimiert.

Verbunden mit der Gewinnung eines neuen Partners – Prof. Chiarcos aus dem Bereich Angewandte Computerlinguistik am Institut für Informatik der Goethe-Universität Frankfurt a. M. – bekam das Projekt jedoch eine grundsätzlich neue Ausrichtung: Das Linguistik-Portal öffnete sich für Semantic-Web-Technologien.

2 Virtuelle Fachbibliothek Linguistik und Linked Open Data

In der zweiten ViFa-Förderphase wurden die Voraussetzungen geschaffen, um das Linguistik-Portal mit Linked Open Data (LOD) zu vernetzen. Diese Vernetzung generiert zweifachen Nutzen: Auf der einen Seite ermöglicht sie die Integration von linguistisch relevanten LOD-Ressourcen in die Portal-Suche, auf der anderen Seite ermöglicht sie durch die damit einhergehende Veröffentlichung von bibliografischen Metadaten die Nachnutzung dieser Daten durch die LOD-Community. Dank der stark gestiegenen Verbreitung und Akzeptanz von LOD werden immer mehr freie, linguistisch relevante Ressourcen nach Linked-Data-Prinzipien⁴ gemäß der entsprechenden W3C-Standards⁵ aufbereitet. So entstand auch die Linguistic Linked Open Data Cloud (LLOD-Cloud⁶), die auf eine Initiative der Open Linguistics Working Group der Open Knowledge Foundation zurückgeht

3 IDS Bibliografie zur deutschen Grammatik; IDS Bibliografie zur Gesprächsforschung; Bibliografie-Datenbank Konnektoren; Bibliografie-Datenbank Präpositionen; Online-Bibliografie zur elektronischen Lexikografie; Linguistik-Sammlung des MPI für Menschheitsgeschichte; Language Description Heritage; OPAC des MPI für Psycholinguistik; DNB-Sachgruppen 400–499; Bibliothekskatalog des Instituts für empirische Sprachwissenschaft; Bibliographie der deutschen Sprach- und Literaturwissenschaft (BDSL)/„Deutsche Sprachwissenschaft (in Auswahl)“; Frankfurter Retrokatalog.

4 <https://www.w3.org/DesignIssues/LinkedData.html> [Zugriff: 05.01.2018].

5 <https://www.w3.org/standards/semanticweb/data> [Zugriff: 05.01.2018].

6 <http://linguistic-lod.org/lod-cloud> [Zugriff: 05.01.2018].

und Ressourcen wie Wörterbücher, terminologische Repositorien oder Korpora zusammenfasst.

Von besonderem Interesse für die linguistische Forschung und dementsprechend für das Linguistik-Portal sind die Korpora. Dabei verstehen wir unter Korpora (Sprachkorpora, Textkorpora) im Volltext vorliegende Textsammlungen, die linguistisch annotiert sind, d. h. die enthaltenen Wörter oder anderen Sprachbestandteile sind mit Begriffen zur Beschreibung von sprachlichen Phänomenen markiert – z. B. „Kardinalzahl“, „Personalpronomen“. Als Markierung werden sogenannte Tags verwendet, und die Liste aller Tags stellt das Annotationschema bzw. das Tag-Set eines Korpus dar.

Die Heterogenität⁷ der linguistischen Annotationen erschwert oft die Nachnutzung der linguistischen Ressourcen. Um die Interoperabilität möglichst vieler linguistischer Daten und Metadaten zu ermöglichen, wurden die Ontologies of Linguistic Annotations (OLiA⁸) entwickelt, die einen zentralen Knoten in der LLOD-Cloud darstellen. OLiA ist ein modular aufgebautes Metadaten-Repository, das aus verschiedenen Komponenten besteht⁹:

- Das OLiA Reference Model definiert die Terminologie (z. B. „Nomen“, „Plural“, „Akkusativ“), die als Referenz für linguistische Annotationen dient.
- Die OLiA Annotation Models sind formalisierte Versionen der integrierten Tag-Sets.
- Die OLiA Linking Models definieren jeweils die Beziehungen zwischen einem gegebenen Annotation Model und dem OLiA Reference Model.

Da das OLiA Reference Model als Mediator zwischen den verschiedenen Annotationsschemata dient, spielt es bei der Vernetzung des Linguistik-Portals mit LOD eine besondere Rolle: Es wurde als Anker in der LLOD-Cloud ausgewählt. Als Anknüpfungspunkt auf Seiten des Linguistik-Portals fungiert der Schlagwort-Thesaurus der BLL, der seit über 40 Jahren an der UB Frankfurt gepflegt und weiterentwickelt wird. Der BLL-Thesaurus ist für diese neue Funktion prädestiniert, denn er liefert die Grundlage sowohl für die thematische Klassifikation des Linguistik-Portals als auch für die normierten Schlagwörter, die der intellek-

⁷ Es existieren zahlreiche Tag-Sets, und die linguistischen Annotationen können sehr heterogen sein – konzeptuell, aber auch rein formal. So sind sogar bedeutungsgleiche Konzepte in unterschiedlichen Tag-Sets häufig verschieden kodiert; z. B. ist der Tag für „Kardinalzahl“ im Stuttgart-Tübingen-TagSet „CARD“ und im Brown Corpus Tagset „CD“; „Personalpronomen“ wird entsprechend durch „PPER“ bzw. „PRP“ gekennzeichnet.

⁸ <http://acoli.cs.uni-frankfurt.de/resources/olia/> [Zugriff: 05.01.2018].

⁹ Vgl. Chiarcos, Christian; Sukhareva, Maria: OLiA – Ontologies of Linguistic Annotation. In: *Semantic Web Journal* 6 (2015) 4, S. 379–386.

tuellen Erschließung in den Modulen dienen. In der zweiten ViFa-Förderphase wurde der BLL-Thesaurus nach LOD-Prinzipien aufbereitet und mit dem OLiA Reference Model verknüpft. Der Fokus lag dabei auf der Bearbeitung der Thesaurus-Begriffe aus den Bereichen Syntax, Morphologie und Morphosyntax, die auch das Kernvokabular des OLiA Reference Model darstellen.

Die konkrete Umsetzung erfolgte, indem der vollständige Thesaurus zuerst in SKOS konvertiert wurde. SKOS¹⁰ steht dabei für „Simple Knowledge Organisation System“, was das standardmäßig genutzte Vokabular für die Linked-Data-konforme Repräsentation von Thesauri und anderen Dokumentationssprachen ist. Ein Terminologie-Modell kann allerdings nur dann reibungslos in die modulare Architektur von OLiA integriert werden, wenn es die formalen Kriterien einer Ontologie erfüllt. Die Beziehungen zwischen den Begriffen innerhalb des BLL-Thesaurus genügen jedoch nur bedingt diesen Kriterien¹¹. Deshalb war eine inhaltliche und formale Aufarbeitung vonnöten: Der Thesaurus wurde basierend auf der SKOS-Version und unter Verwendung der Web Ontology Language (OWL) um eine komplexe, ontologische Struktur ergänzt.

Im Zuge dessen musste jedes Schlagwort hinsichtlich seiner Beschaffenheit intellektuell geprüft und ggf. überarbeitet werden. Einige Schlagwörter wurden umbenannt, andere wurden neu hierarchisiert. „Pronominalisierung“ ist im Thesaurus bspw. eine Subkategorie von „Pronominalsyntax“ und wurde in der Ontologie als Subklasse von „Syntaktischer Prozess“ eingeordnet. Mehrdeutige Begriffe wurden als komplexe Klassen modelliert. So wurde z. B. das Thesaurus-Schlagwort „Komposition“, das sowohl den morphologischen Prozess als auch das Ergebnis dieses Prozesses abdeckt, als die Disjunktion zweier neu definierter Klassen – „Kompositum“ und „Komposition“ – dargestellt. Die so entstandene BLL-Ontologie eignet sich für maschinelle Reasoning-Verfahren und kann wiederum mit anderen Ontologien verknüpft werden.

Anschließend wurde ein Linking Model (BLL-OLiA-Linking) erstellt, das die Zuordnung von BLL-Schlagwörtern zu den einzelnen Begriffen aus dem OLiA Reference Model beinhaltet. Damit war die Grundlage für die Realisierung eines Anwendungsszenarios geschaffen: die Implementierung einer LOD-basierten Suchfunktion im Linguistik-Portal.

Um LOD-Ressourcen in das Linguistik-Portal einbinden zu können, wurde ein Webcrawler entwickelt. Ausgehend vom Metadaten-Portal Linghub¹² durchsucht

¹⁰ <https://www.w3.org/2004/02/skos/> [Zugriff: 05.01.2018].

¹¹ So ist bspw. „Adjektiv“ im BLL-Thesaurus eine Subkategorie von „Adjektivsyntax“, was keine zulässige Ober-/Unterbegriff-Relation darstellt.

¹² <http://linghub.org/> [Zugriff: 05.01.2018].

der Crawler zunächst die LLOD-Cloud nach relevanten Datensätzen. Indiziert werden alle LLOD-Ressourcen, die eine Verknüpfung mit dem OLiA Reference Model aufweisen. Das sind vor allem Korpora, deren Tag-Set bereits in die modulare Architektur von OLiA integriert wurde, aber auch Wörterbücher, die Begriffe aus dem OLiA Reference Model als linguistische Referenzterminologie einsetzen. Anschließend werden die indizierten Daten unter Verwendung des BLL-OLiA-Linking nach den entsprechenden BLL-Begriffen durchsucht. Die Ergebnisliste wird um die Linghub-Metadaten ergänzt und in das Linguistik-Portal exportiert. Durch die nahtlose Integration in die bestehende Katalogsuche wird ein niederschwelliger Zugriff auf LLOD-Ressourcen ermöglicht.

Zu Projektende erfolgte unter <http://data.linguistik.de/bll> die Veröffentlichung der LLOD-Edition des BLL-Thesaurus. Dazu gehören nicht nur die SKOS-Version und die BLL-Ontologie, sondern auch eine Datei, die die Zuordnung frei zugänglicher BLL-Titeldaten zu BLL-Schlagwörtern beinhaltet (BLL-Index). Die Veröffentlichung dieser Datei dient dabei der Realisierung eines zweiten Anwendungsszenarios: Der BLL-Index ermöglicht es, ausgehend von linguistischen Ressourcen in der LLOD-Cloud auf dem Wege über die Verknüpfung von OLiA zu BLL-Schlagwörtern auf thematisch qualifizierte Publikationen im Linguistik-Portal zu verlinken.

3 Fachinformationsdienst Linguistik und Linked Open Data

Das Linguistik-Portal wird seit 2017 als FID Linguistik an der UB Frankfurt in Kooperation mit dem Institut für Informatik weiter ausgebaut (Projektlaufzeit 2017–2019). Da LOD-Technologien über ein großes Potential verfügen, beim Aufbau eines flexiblen und zukunftsfähigen Systems der Informationsversorgung einen wichtigen Beitrag zu leisten¹³, werden im FID Linguistik die Arbeiten hinsichtlich LOD konsequent fortgesetzt. Zum einen wird die LOD-Funktionalität erweitert, indem die BLL-Sprachbezeichner formalisiert und verlinkt werden. Zum anderen werden Semantic-Web-Technologien angewendet, um Korpora-Metadaten anzureichern und dadurch den Suchhorizont über die LLOD-Cloud hinaus auszudehnen.

¹³ Vgl. Christoph, Pascal; Pohl, Adrian: Dezentral, offen, vernetzt – Überlegungen zum Aufbau eines LOD-basierten FID-Fachinformationssystems. In: Bibliothek, Forschung und Praxis 38 (2014) 1, S. 114–123.

3.1 Vernetzung mit Glottolog und Lexvo

Basierend auf dem Verfahren, das in der zweiten ViFa-Förderphase für die Bearbeitung der Thesaurus-Begriffe aus dem Bereich linguistischer Metasprache entwickelt wurde, werden nun die BLL-Sprachbezeichner in die BLL-Ontologie integriert und mit LLOD-Repositorien verknüpft.

In der Linguistik werden Sprachen in der Regel gemäß ihren Verwandtschaftsbeziehungen hierarchisiert. Ein prototypisches Beispiel ist die Indoeuropäische Sprachfamilie mit ihren Unterfamilien (Germanisch, Keltisch usw.) und den dazugehörigen Einzelsprachen (Isländisch, Walisisch usw.). In vielen Regionen der Welt wie z. B. Papua-Neuguinea oder Australien stößt das Stammbaum-Modell allerdings an seine Grenzen, sodass die Systematisierung oft durch eine Einteilung in geografische Zonen erfolgt.

Die Anordnung der Normdatensätze im BLL-Thesaurus folgt in erster Linie bibliografiepraktischen Erfordernissen. Bei der OWL-Modellierung wird jedoch darauf geachtet, dass die angewandten Definitionen auch den wissenschaftlichen Erwartungen gerecht werden. So finden hypothetische, im Thesaurus vorhandene, aber in der Linguistik umstrittene Gruppierungen wie die Balkanindoeuropäischen Sprachen keine Repräsentation in der BLL-Ontologie. In anders gelagerten Fällen wird die Ontologie hingegen sogar um eine weitere, im BLL-Thesaurus nicht gegebene Ebene ergänzt, um allgemein akzeptierte Untergliederungen bestimmter Sprachgruppen darzustellen (z. B. die Verzweigung der Slawischen Sprachen in Ost-, Süd- und Westslawisch).

Als Anknüpfungspunkt innerhalb der LLOD-Cloud wurden die beiden Metadaten-Repositorien Glottolog¹⁴ und Lexvo¹⁵ ausgewählt, die jeweils unterschiedliche Ansätze hinsichtlich der Modellierung von Sprachbezeichnern haben. Glottolog¹⁶ verfolgt das Ziel, basierend auf beschreibendem Quellenmaterial (Langdoc-Bibliografie) eine möglichst detaillierte, genealogische Klassifikation von Sprachvarietäten und Varietätengruppen zu erstellen. Eine Sprachvarietät wird dabei ausschließlich extensional über die Menge der damit verbundenen Literatur definiert.

¹⁴ <http://glottolog.org/> [Zugriff: 05.01.2018].

¹⁵ <http://lexvo.org/> [Zugriff: 05.01.2018].

¹⁶ Vgl. Nordhoff, Sebastian; Hammarström, Harald: Glottolog/Langdoc: Defining dialects, languages, and language families as collections of resources. In: Proceedings of the First International Workshop on Linked Science 2011 (LISC2011), Bonn, Germany, October 24, 2011, S. 53–58. <http://www.mpi.nl/publications/escidoc-1752673> [Zugriff: 05.01.2018].

Lexvo¹⁷ besitzt das Alleinstellungsmerkmal, dass es zu den Sprachcodes ISO 639-3 maschinell adressierbare URIs generiert – ohne den Anspruch, dabei explizite Beziehungen zwischen Sprachen zu postulieren. Darüber hinaus bietet Lexvo auch wertvolle Informationen in Form von Sprachbezeichnern in unterschiedlichen Sprachen. Ein weiterer entscheidender Vorteil von Lexvo ist, dass es bereits mit zahlreichen anderen LOD-Quellen verlinkt ist.

Glottolog und Lexvo unterscheiden sich dabei nicht nur in ihrem konzeptionellen Ansatz, sondern auch in ihrer Abdeckung: Für die meisten Subvarietäten (z. B. Hessisch, Molisano) und Sprachfamilien (z. B. Kartwelisch, Tai-Kadai-Sprachen) bei Glottolog existiert keine Lexvo-Entsprechung¹⁸. Umgekehrt finden sich bei Lexvo historische oder ausgestorbene Sprachvarietäten (z. B. Burgundisch), für die es bei Glottolog aufgrund fehlenden Quellenmaterials keinen Eintrag gibt.

Als Ausgangsbasis für die Verlinkung von BLL zu Glottolog und Lexvo dient ein maschinell erzeugtes Mapping. Dieses vorläufige Modell enthält Verknüpfungskandidaten, die durch einen automatisierten Abgleich der offiziellen ISO-Listen mit den Labels der BLL-Sprachbezeichner generiert wurden. Die Erstellung der Links erfolgt manuell nach der intellektuellen Prüfung, ob oder inwieweit Sprachbezeichner aus allen drei Systemen deckungsgleich sind.

Durch die Verknüpfung von BLL mit den beiden anderen Online-Terminologien erreichen wir eine substantielle Erweiterung der LOD-Suchfunktionalität im Linguistik-Portal. Alle Ressourcen, die Sprachbezeichner von Glottolog oder Lexvo verwenden, werden somit sichtbar und recherchierbar. Im Zuge dessen wird auch die Langdoc-Bibliografie in das Portal eingebunden. Und durch die Verlinkung werden weitere zukünftige Anwendungsszenarien denkbar, bspw. eine geografische Suche, die auf geo-spatiale Informationen aus Glottolog rekurriert, oder die Einblendung enzyklopädischer Informationen, die über Lexvo aus DBpedia referenziert werden können.

3.2 Metadaten-Repository Annohub

Als weitere Maßnahme, um die Recherchierbarkeit von Sprachressourcen zu verbessern, soll im FID Linguistik eine abgewandelte Form automatisierter Sacherschließung zum Einsatz kommen. Der Fokus liegt dabei auf Sprachressourcen,

¹⁷ Vgl. de Melo, Gerard: Lexvo.org: Language-Related Information for the Linguistic Linked Data Cloud. In: Semantic Web Journal 6 (2015) 4, S. 393–400. <http://semantic-web-journal.net/system/files/swj521.pdf> [Zugriff: 05.01.2018].

¹⁸ Dialekte haben per Definition keinen ISO-Code.

die sich außerhalb der LLOD-Cloud befinden. Für einen Großteil frei verfügbarer, linguistisch relevanter Ressourcen, deren Daten bislang nur in konventionellen¹⁹ Formaten vorliegen, werden bereits formale Metadaten LOD-konform bereitgestellt²⁰. Diese basalen Metadaten (Autor, Titel usw.) sollen um Informationen zum Inhalt angereichert werden.

Der Kerngedanke des eingesetzten Verfahrens ist, dass die zur Annotation eines Korpus verwendeten Begriffe sich als valide Suchterme qualifizieren. Folglich kann ein Korpus um inhaltliche Metadaten – i. S. v. Sacherschließung – angereichert werden, indem das verwendete Tag-Set ermittelt wird.

Anhand einer Datenprobe werden die Korpora automatisch dahingehend analysiert, welche Tags eingesetzt wurden und ob für die vorhandenen Annotationen bereits ein OLiA Annotation Model existiert (z. B. Stuttgart-Tübingen-Tag-Set, Brown Corpus Tagset). Falls dies der Fall ist, wird das analysierte Korpus explizit diesem OLiA Annotation Model zugeordnet, d. h. sämtliche im Modell vorgesehenen Annotationen („Kardinalzahl“, „Personalpronomen“ usw.) sind grundsätzlich für die Beschreibung dieser Ressource geeignet und können als Basis für die Indizierung dienen. Für Korpora mit Annotationen – jedoch ohne ermittelbares OLiA Annotation Model – soll später experimentell eine Technologie erprobt werden, bei der existierende Annotationen mithilfe maschinellen Lernens automatisch auf OLiA-Konzepte abgebildet werden.

Falls die Sprache des Korpus kein Bestandteil der vorliegenden Metadaten ist, wird sie im Vorfeld mit gängiger Spracherkennungssoftware (z. B. Apache Tika) ermittelt (language guessing). Für die Speicherung der gewonnenen inhaltlichen Metadaten zu Sprache und OLiA Annotation Model soll ein Repositorium „Annohub“ eingerichtet werden, das durch die Bibliothek persistent gehostet wird. Die erzeugten Metadaten werden – unter einer offenen Lizenz – maschinenlesbar zur Verfügung gestellt und den Betreibern existierender Metadatenportale explizit zur Nachnutzung angeboten. Und last but not least dienen die Annohub-Daten dazu, weitere Ressourcen – insbesondere außerhalb der LLOD-Cloud – für die Suche im Linguistik-Portal zu erschließen.

¹⁹ „Konventionell“ bedeutet in diesem Zusammenhang CoNLL, TIGER-XML oder Penn-Listenannotation.

²⁰ Dies geschieht bspw. über Portale wie Datahub, Linghub, Meta-Share, CLARIN Virtual Language Observatory oder Open Language Archives Community.

4 Weitere Dienstleistungen

Wenn auf diese oder andere Weise gefundene oder selbst erstellte Forschungsdaten schlussendlich in wissenschaftlichen Arbeiten analysiert und die Resultate in Publikationen eingeflossen sind, dann gilt es, nicht nur die Daten und die Veröffentlichung, sondern auch ihre Relation zueinander sichtbar zu machen. Zu diesem Zweck wird im Linguistik-Portal ein neues Modul eingerichtet werden: Hier wird es möglich sein, ausgehend von einem gewählten Korpus zu den dazugehörigen Publikationen zu navigieren oder umgekehrt von einer Publikation auf das behandelte Korpus – im Idealfall auf dessen Webpräsenz – zu verlinken. Zudem ist es geplant, die Korpora hinsichtlich ihrer Sprache zu erschließen und entsprechend aufbereitet in der Modul-Navigation anzuzeigen.

Als Ausgangsbasis dienen Einträge aus dem Titelbestand der BLL. Hier finden sich in reservierten Freitextfeldern über 1.000 verschiedene Korpora (z. B. Austrian Media Corpus, The York Corpus of French Child Language), die für die Beschlagwortung von über 3.000 Publikationen verwendet wurden. Da die zu verlinkenden Entitäten jeweils durch eine URL adressierbar sein müssen, wird vorab ermittelt, ob die verschlagworteten Korpora über eine eigene Internetpräsenz verfügen. Für die Suche nach der URL konsultieren wir vor anderen einschlägigen Portalen zuerst den facettierten Browser von CLARIN²¹. Denn der Nachweis im CLARIN-Browser hat den entscheidenden Vorteil, dass die in zertifizierten CLARIN-Zentren gehaltenen Ressourcen bereits über einen Persistent Identifier wie URN, DOI oder Handle verfügen. Wo dies nicht der Fall ist, kann teilweise auf eine International Standard Language Resource Number (ISLRN²²) zurückgegriffen werden. Um die Verlinkung von Publikation zu Korpus technisch zu realisieren, bedarf es – als letztes Glied in der Verknüpfungskette – nur noch der Umwandlung der bisher als Freitext vorliegenden Korpus-Schlagwörter zu einem vollwertigen BLL-Normdatensatz, der auf die oben ermittelte Internetadresse des jeweiligen Korpus referiert.

Ergänzend zu dieser bibliografischen Aufbereitung ist auch die Verfügbarkeit von (kommerziellen) Korpora ein Thema im FID. Im Zuge der Einrichtung von FIDs entstand das Konzept überregionaler FID-Lizenzen für fachlich definierte Nutzergruppen (Virtueller Nutzerkreis²³). Der FID Linguistik plant hier, in einem

²¹ CLARIN Virtual Language Observatory <https://vlo.clarin.eu> [Zugriff: 05.01.2018].

²² <http://www.islrn.org/> [Zugriff: 05.01.2018].

²³ Vgl. Kompetenzzentrum für Lizenzierung (KfL): Nutzerkreismodelle für FID-Lizenzen 1.11.2016 <http://www.fid-lizenzen.de/dateien/nutzerkreismodelle-fuer-fid-lizenzen> [Zugriff: 05.01.2018].

Pilotvorhaben FID-Lizenzen auch für neue Medientypen wie Sprachkorpora anzubieten.

Klassische Medientypen werden weiterhin – wenn auch in beschränktem Umfang – erworben. Um hierbei dem Aspekt der stärkeren Nutzerbeteiligung und Ausrichtung an den konkreten Bedürfnissen der Wissenschaftler Rechnung zu tragen, wurde im Linguistik-Portal ein Online-Formular für überregionale Kaufvorschläge installiert.

Jenseits lizenzpflichtiger bzw. allgemein kostenpflichtiger Erwerbung soll im FID die Einbindung von Open-Access-Dokumenten weiter ausgebaut werden, die in der ViFa mit der Indexierung verschiedener Open-Access-Repositoryen begann. Deshalb bieten wir im laufenden Projekt interessierten Wissenschaftlern oder kleineren Verlagen eine technische Plattform für das Publizieren von frei verfügbaren Zeitschriften.

5 Schlussbemerkungen

Das Linguistik-Portal ist einerseits die letzte ViFa, die von der DFG im auslaufenden Förderprogramm „Überregionale Literaturversorgung“ unterstützt wurde. Andererseits überlappt sich die zweite Förderphase nicht nur zeitlich mit dem Start des neuen DFG-Förderprogramms „Fachinformationsdienste für die Wissenschaft“, sondern weist auch inhaltlich eine Neuausrichtung auf. Bibliothekarisches Arbeiten wird nicht mehr nur als Nachweis und Erschließung gegebener Literatur verstanden, sondern es werden neue Formen der Informationsgewinnung unterstützt. So hat sich hinsichtlich Zeitaufwand und Intensität das Verhältnis von Arbeiten mit aggregierendem Charakter (Indexsuche, Neuerwerbungsliste) über Inhaltserschließung (thematische Verzeichnisse) bis hin zu inhaltsgenerierenden Maßnahmen (Ontologie-Erstellung und -Verlinkung) kontinuierlich verschoben. Und diese Tendenz setzt sich im laufenden FID-Projekt fort (Forschungsdaten-Analyse).

Gleichzeitig ist auch im Bereich der eingesetzten Methoden eine Trendwende zu erkennen: Neben traditionellen Erschließungsmethoden (bibliografisch, intellektuell) wurden zunehmend Semantic-Web-Technologien eingesetzt (Veröffentlichung, Vernetzung, Verarbeitung von LOD). Diese innovativen Ansätze bieten eine hohe Flexibilität bei der Gestaltung der Informationsdienstleistungen und ermöglichen weitere Anwendungen in der Zukunft. Sowohl förderungstechnisch als auch konzeptionell steht das Linguistik-Portal somit für einen fließenden Übergang von Virtueller Fachbibliothek zu Fachinformationsdienst.

**Vanya Dimitrova, M. A.**

Fachinformationsdienst Linguistik
Universitätsbibliothek Johann Christian Senckenberg
Bockenheimer Landstr. 134–138
60325 Frankfurt am Main
Deutschland
E-Mail: v.dimitrova@ub.uni-frankfurt.de

**Heike Renner-Westermann, M. A.**

Fachreferentin für Linguistik
Leitung Bibliography of Linguistic Literature
Fachinformationsdienst Linguistik
Universitätsbibliothek Johann Christian Senckenberg
Bockenheimer Landstr. 134–138
60325 Frankfurt am Main
Deutschland
E-Mail: h.renner-westermann@ub.uni-frankfurt.de