

# Das Archiv für Sozialwissenschaft und Sozialpolitik (AfSS, 1904–1933) als integriertes Textkorpus.

Zur geplanten Digitalisierung des AfSS im CLARIN-Servicezentrum der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW)

Das CLARIN-Servicezentrum der BBAW erstellt und integriert Textressourcen für deren disziplinenübergreifende Beforschung mit computer- und korpuslinguistischen Werkzeugen sowie zur vielseitigen Nachnutzung im BMBF-geförderten Infrastrukturprojekt CLARIN-D [1]. Ein Schwerpunkt liegt dabei auf historischen Sprachdaten, die in Form von einschlägigen Monographien und Einzeldokumenten sowie Zeitungen und Periodika über das DFG-geförderte *Deutsche Textarchiv (DTA)* [2] bzw. die Korpora des Akademienvorhaben *Digitales Wörterbuch der deutschen Sprache (DWDS)* [3] bereitgestellt werden. Die Textressourcen werden in der Regel manuell erfasst, was deren herausragende Qualität auf der Ebene der Textgenauigkeit garantiert, und einheitlich gemäß den Vorgaben der *Text Encoding Initiative (TEI)* [4] strukturiert, was deren Interoperabilität und nachhaltiger, plattformunabhängiger Nutzbarkeit dient. Das AfSS soll, als Pionierunternehmen und herausragende Zeitschrift der internationalen historischen Sozialforschung in der skizzierten Weise in die an der BBAW aufgebaute Korpusinfrastruktur integriert werden. Der Vortrag gibt einen Ausblick auf die geplante Digitalisierung des AfSS und demonstriert anhand bereits integrierter Korpusressourcen computer- und korpuslinguistische Analysen im Kontext einer geplanten Kooperation der Europa-Universität Viadrina Frankfurt (Oder), der Bayerischen Akademie der Wissenschaften und der BBAW.

- [1] <http://www.clarin-d.net/de/>; CLARIN-D ist der vom BMBF geförderte deutsche Partner der EU-weiten, web- und zentrenbasierten Forschungsinfrastruktur CLARIN (Common Language Resources and Technology Infrastructure). CLARIN stellt Textressourcen und Werkzeuge zu deren Analyse für die Geistes- und Sozialwissenschaften zur Verfügung, wobei deren Standardisierung, Interoperabilität und nachhaltige Nutzbarkeit innerhalb des Verbundes sichergestellt werden.
- [2] Deutsches Textarchiv (DTA), <http://www.deutschestextarchiv.de/>.
- [3] Digitales Wörterbuch der deutschen Sprache (DWDS). Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart, <http://www.dwds.de/>.
- [4] TEI: Text Encoding Initiative, <http://www.tei-c.org/>; TEI: P5 Guidelines, <http://www.tei-c.org/Guidelines/>. Auf diesen Empfehlungen zur Strukturierung von XML-Dokumenten basiert das "Basisformat" des *Deutschen Textarchivs*, <http://www.deutschestextarchiv.de/doku/basisformat>, das mittlerweile auch von der DFG empfohlen wird: [http://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/informationen\\_fachwissenschaften/geisteswissenschaften/standards\\_sprachkorpora.pdf](http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/standards_sprachkorpora.pdf), S. 24. Das DTA-Basisformat beinhaltet Anpassungen zur Auszeichnung spezifischer Textsorten, im Falle des AfSS werden die Regeln zur Auszeichnung von Zeitungstexten zugrunde gelegt: [http://www.deutschestextarchiv.de/doku/basisformat\\_zeitungen](http://www.deutschestextarchiv.de/doku/basisformat_zeitungen). Siehe auch S. Haaf, M. Schulz: Historical Newspapers & Journals for the DTA. In: Proceedings of the LRT4HDA workshop, held at the 9th LREC, Reykjavik (Iceland) 2014, <http://www.lrec-conf.org/proceedings/lrec2014/workshops/LREC2014Workshop-LRT4HDA%20Proceedings.pdf#page=57>.