

**exploreAT!**  
**Perspektiven einer Transformation am Beispiel eines lexikographischen  
Jahrhundertprojekts**

Im vorliegenden Paper wird erstmals das interdisziplinäre, internationale Projekt „exploreAT! exploring austria's culture through the language glass“ vorgestellt, das ab 2015 an der Österreichischen Akademie der Wissenschaften im Bereich „Digital Humanities“ für 4 Jahre umgesetzt wird.

Die beteiligten Kolleg:innen arbeiten in den Bereichen Soziologie, Softwareentwicklung, Mensch-Maschine-Interaktion, Lexikographie und Digital Humanities.

Auf Basis des Jahrhundertprojekts „Wörterbuch der bairischen Mundarten in Österreich (WBÖ)“ (1911-) und seines digitalen Schwesterprojekts „Datenbank der bairischen Mundarten in Österreich (DBÖ)“ [1993-; samt Weiterentwicklungen wie: „Datenbank der bairischen Mundarten in Österreich electronically mapped (dbo@ema)“ 2007-] wird ein Beispiel für einen Transformationsprozess eines geisteswissenschaftlichen Projekts diskutiert.

Die Datensammlung ist quantitativ umfassend:

Wir arbeiten mit rund 200,000 ungedruckten, nicht-lexikographisch finalisierten Stichwörtern; geschätzten 4 Mio. Dateneinträgen (Wörter, Kontexte); 5 Wörterbuchbänden, mit über 50,000 Stichwörtern; jeweils von den Anfängen der deutschen Sprache bis heute, unter besonderer Berücksichtigung des bairischen Dialekts in Österreich (sei es die Habsburgermonarchie, wie es für den Großteil des Sammelzeitraums essentiell ist, oder das Staatsgebiet des heutigen Österreich).

Die Zusammenarbeit ist motiviert im Bestreben der Kolleg:innen, folgende Aspekte einem zu einem synergetischen Ganzen zusammenzufügen:

1. Das Vorhandensein einer unikalen Wörtersammlung für die ländliche, österreichische Kultur des letzten Jahrhunderts.
2. Das Vorhandensein darauf aufbauender wissenschaftlicher Arbeiten und Dokumentationen (z.B: WBÖ, DBÖ, dbo@ema).
3. Das Bewusstsein um die identitätsstiftende Verankerung des Dialekts in Österreich (z.T. noch vermehrt durch die neuen Medien und das Web 2.0, z.B. Facebook, SMS, WhatsApp)-

4. Das umfassende, steigende lexikographische Interesse breiter Bevölkerungsteile (Beispiel: Wikipedia, Regionalwörterbücher).
5. Das wissenschaftliche Interesse der Kolleg.innen an (der Weiterentwicklung) zeitgemäßer Lexikographie und ihren Produkten bzw. an der kritischen Analyse des lexikographischen Prozesses und seiner Einbettung in einen aktualisierten realweltlichen Kontext, markiert durch Cyberscience und Web 2.0- / Web 3.0 im Alltag).

Die grundlegenden Aspekte der Zusammenarbeit werden vorgestellt:

1. Infrastruktur:

Weiterentwicklung einer bestehenden Datenbank zu einer web-basierten, kollaborativen Infrastruktur für Archivierung, Edition, Publikation und Analyse multilingualer Non-Standard-Daten, deren lexikographische Output (diverse Wörterbuchprodukte) und deren Wissensquellen, sowohl für Wissenschaftler als auch für Laien, basierend auf Standards und Up-to-date-Technologien.

Weitere Schwerpunkte der technischen Neuentwicklungen stellen dar:

- Semantic Web:  
Nutzung der Vorteile des Semantic Web für die Lexikographie: Modellierung und Datenpublikation (WBÖ+DBÖ) in Linked Open Data; Semantische Erschließung des Datenkorpus durch Ableitung von Basiskonzepten aus dem Fragebogen (20,000 Detailfragen).
- Visual Analytics:  
Entwicklung und Weiterentwicklung von Visual Analytics Werkzeugen für Non-Standard Daten. Es wird darauf abgezielt, Tools zu entwickeln, die für analoge Projekte eingesetzt und zum Vergleich von Datenstrukturen angewendet werden können.
- Serious Games:  
Entwicklung spielerischer Anwendungen für Lehre und Laien. Damit werden Werkzeuge zur Verfügung gestellt, welche die direkte Kommunikation mit neuen Usern gewährleisten sollen.

## 2. Daten Enrichment / Re-Use:

Die Zusammenarbeit verfolgt Open Science Prinzipien.

Die Anreicherung der eigenen Daten durch externe Daten (z.B. mittels Linked Open Data) stellt einen Kernpunkt der Zusammenarbeit dar; ebenso wird die inter- und transdisziplinäre Wiederverwendung der Daten in unterschiedlichen Kontexten gefördert.

Die Zusammenarbeit mit Firmen bzw. das Einbringen lexikographischen Knowhows in den Business Kontext wird erprobt.

## 3. Gesellschaft:

Basierend auf der Rolle und Funktion, die Non-Standard Sprache in Österreich einnimmt, werden Ansätze zur gewinnbringenden Zusammenarbeit und Einbindung der Gesellschaft in den Wissenschaftsprozess diskutiert. Citizen Science Modelle werden aktiv erprobt und umgesetzt.

## 4. Metalexikographie:

Ein Schwerpunkt der Zusammenarbeit widmet sich der methodischen Neuorientierung und somit der Diskussion der Fragestellung, welche Rolle Lexikographie als Teilbereich von „Digital Humanities“ derzeit einnimmt bzw. wo lexikographisches Wissen in anderen Bereichen festgemacht werden kann / könnte.

Welche methodischen Veränderungen bedingen Linked Open Data für die klassische Lexikographie bzw. umgekehrt.

Die Folgen und Auswirkungen lexikographischer Arbeit im Open Science Paradigma wird reflektiert und thematisiert.

Die Zusammenarbeit ist eingebettet in enge Zusammenarbeit mit Stakeholdern in einzelnen Bereichen, wie DARIAH-EU, COST IS 1305 European Network of electronic Lexicography, EUROPEANA, WIKIMEDIA.AT , OPEN KNOWLEDGE FOUNDATION, oder beispielsweise den Projekten LIDER, opendataportal.at, SOCIENTIZE.

Der Beitrag versteht sich als zusammenfassender Überblick über Perspektiven eines Transformationsprozesses für ein lexikographisches Großprojekt. Aufgrund der Ausgangslage – im frühen 20. Jahrhundert sind nach dem Beispiel des Schweizerischen Idiotikons viele analog gestaltete Projekte konzipiert und umgesetzt worden – erscheinen

Analogieentwicklungen und Anwendung diverser Tools auf andere bestehende Projekte unterschiedlichen Digitalisierungsgrads möglich und wünschenswert.

Im Rahmen der COST Aktion IS 1305 werden entsprechende Übertragungen beispielhaft umgesetzt.

Der Beitrag fokussiert neben dem beispielhaften Umreißen einzelner Möglichkeiten vor dem Hintergrund der aktuellen Zusammenarbeit, den Mehrwert für ein lexikographisches Projekt durch einen Transformationsprozess Richtung Digital Humanities.

## Referenzen:

Agosti, M. et al. (Eds.): Digital Libraries and Archives. 2013: 195-206.

Agosti, M. et al. "An Evaluation of the Involvement of General Users in a Cultural Heritage Collection." In: *Digital Humanities* (2013): 75-77.

Agosti, M. et al. "Digital Libraries and Archives. 8th Italian Research Conference, IRCDL, 2012, Bari, Italy, February 9-10, 2012, Revised Selected Papers". *CCIS* Vol. 354.

"[Arbeitsplan](#) und Geschäftsordnung für das bayerisch-österreichische Wörterbuch". Wien. 1913.

Bailey, E. et al: "CULTURA: Supporting Professional Humanities Researchers." In: *Digital Humanities* (2013): 99-101.

Burigat, S., Chittaro, L. "Interactive visual analysis of geographic data on mobile devices based on dynamic queries." *Journal of Visual Languages & Computing* 19.1 (2008): 99-122.

Catarci, T. et al. "Evaluating Cultural Heritage Information Access Systems. Bridging Between Cultural Heritage Institutions." Berlin Heidelberg. 2014: 7-16.

Datenbank der bairischen Mundarten in Österreich electronically mapped ([dbo@ema](#)). Ed. by Wandl-Vogt, E. Wien. 2010.

Dear, M., Ketchum, J., Luria, S. "GeoHumanities: Art, History, Text at the Edge of Place." London, New York. 2014.

De Gasperis, G, Florio, N. "OpenSource Gamification of a Computer Science Lecture to Humanities Students." *Methodologies and Intelligent Systems for Technology Enhanced Learning*. 2014. 119-126.

Declerck, T., et al. "Collaborative Tools: From Wiktionary to LMF, for Synchronic and Diachronic Language Data." Francopoulo, G. (Ed.) LMF Lexical Markup Framework. London. 3/2013.

Declerck, T., Wandl-Vogt, E., Mörth, K. "A [SKOS-based Schema](#) for TEI encoded Dictionaries at ICLTT." In: Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2014), ed. Nicoletta Calzolari u. a., Reykjavik, Iceland, ELRA, Paris, 5/2014.

Declerck, T., Wandl-Vogt, E. "How to [semantically relate](#) dialectal Dictionaries in the Linked Data Framework." In: Kalliopi Zervanou u. a. (Ed.), Proceedings of the 8th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTech 2014), Gothenburg, Sweden, ACL, 4/2014.

Deutscher, G. "Through the Language Glass. How Words Colour Your World." London: 2010.

Esteban, A., Therón, R. "Corpusexplorer: supporting a deeper understanding of linguistic corpora." *Smart Graphics*. Vol. 6815 (2011): 126-129.

Europäische Kommission: "[Digital Science](#) in Horizon 2020." 2013.

Ferro N. et al. "Fostering Interaction with Cultural Heritage Material via Annotations: The FAST-CAT-Way." In: T. Catarci et al. (Eds.): Bridging between Cultural Heritage Institutions, Proceedings of the 9th Italian Research on Digital Libraries (IRCIDL 2013), CCIS Vol. 385, Berlin, Heidelberg. 2014. 41-52.

Finke, P. "Citizen Science: Das unterschätzte Wissen der Laien." München. 2014.

Heath, T., Bizer, Ch. "Linked data: Evolving the web into a global data space." *Synthesis lectures on the semantic web: theory and technology* 1.1 (2011): 1-136.

- van Hooland, S., Verborgh, R. "Linked Data for Libraries, Archives and Museums. How to clean, link and publish your metadata." London: 2014.
- Hoque, F., Bear, D. "Everything Connects. How to TRANSFORM and LEAD in the Age of Creativity, Innovation and Change." New York, Chicago, San Francisco, Athens, London, Madrid, Mexico City, Milan, New Delhi, Singapore, Sydney, Toronto. 2014.
- Jagoda, P. "Gaming the Humanities." *differences* 25.1 (2014): 189-215.
- Keim, D.A., Oelke, D. "Literature fingerprinting: A new method for visual literary analysis." *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*. IEEE, 2007.
- McCrae, J., Declerck, T. et al. "Interchanging lexical resources on the Semantic Web." In: Language Resources and Evaluation. Vol. 46, Issue 4, 2012:701-719.
- Nentwich M., König R. "Cyberscience 2.0: Research in the Age of Digital Social Networks." Frankfurt, New York: 2012.
- Novotny, H., Scott, P., Gibbons, M. "Mode 2 Revisited: The New Production of [Knowledge](#)." *Minerva* 41 (2003): 179-194.
- Pink, S. "Interdisciplinary agendas in visual research: re-situating visual anthropology." *Visual studies* 18.2 (2003): 179-192.
- Raddick, M. J., Bracey, G., Carney, K., Gyuk, G., Borne, K., Wallin, J., Jacoby, S., et al. (2009). [Citizen Science](#): Status and Research Directions for the Coming Decade. AGB Stars and Related Phenomena Astro2010 The Astronomy and Astrophysics Decadal Survey, 2010, 46P.
- Reder, C. (Ed.). "Kartographisches Denken." New York 2012.
- "[Straffungskonzept](#) für das Wörterbuch der bairischen Mundarten in Österreich (WBÖ)". Wien. 1998.
- The Royal Society. "Science as an [Open Enterprise](#)". 2012.
- Therón, R., Fontanillo L. "[Diachronic-information visualization](#) in historical dictionaries." (2014).
- Therón, R. et al. "[Visual analytics](#): A novel Approach in corpus linguistics and the Nuevo Diccionario Histórico del Español. *Proc. of III Congreso Internacional de Lingüística de Corpus*. 2011.
- Therón, R., Wandl-Vogt, E. "The Fun of Exploration: How to [Access](#) a Non-Standard Language Corpus Visually". LREC-Proceedings 2014.
- Wandl-Vogt, E., Declerck, T. "Mapping a Traditional Dialectal Dictionary with [Linked Open Data](#)." In: Kosem, I. et al. (Eds.): Electronic lexicography in the 21<sup>st</sup> century: thinking outside the paper. Proceedings. Ljubljana / Tallinn. 2013: 460-471.
- Wiggins, A., Crowston K. "From Conservation to [Crowdsourcing](#): A Typology of Citizen Science." Proceedings of the Forty-fourth Hawai'i International Conference on System Science (HICSS-44). 2011.
- Wörterbuch der bairischen Mundarten in Österreich (WBÖ). Wien. 1963-. [online](#): 2012-.