

Exil:Trans - a blueprint for research data reuse

Kremmel, Stefanie

stefanie.kremmel@univie.ac.at
University of Vienna

Steiner, Christian

christian.steiner@dhcraft.org
Digital Humanities Craft

Pollin, Christopher

christopher.pollin@dhcraft.org
Digital Humanities Craft

The project Exil:Trans (2019-2022, <https://exiltrans.univie.ac.at/> and <https://gams.uni-graz.at/exil>) investigates the impact of exile due to National Socialist persecution on translators and their translation activities from a translation studies perspective. The overall aim of this international project is to contribute to a more comprehensive understanding of the cultural effects resulting from the forced displacement of persons by the Nazis (*Tashinskiy et al. 2021*). By reconstructing biographies and tracing translators' networks and involved institutions, trans-individual patterns concerning the role of translators in exile are to be extracted. To achieve its aims, the project team joined forces with the newly founded University of Graz Spin-Off "Digital Humanities Craft" (<https://dhcraft.org/>) to set up a biographical database that follows the paradigm of Semantic Web (*Hitzler 2021*) and Linked Open Data (*Bizer, Vidal, and Skaf-Molli 2018*).

After identifying the names of translators in exile, biographical data and information about their translatorial work was gathered through archival research. The broad data sets, complemented by extensive case studies, have already shown that the topic of exile due to National Socialist persecution is not "over-researched" with regard to translators and translation processes and that many biographies, institutions and therefore archival collections have not been studied in this regard. Exil:Trans shows that the structure of the translational fields and the products of translation in exile are very varied.

Since the sources on individual translators in exile are scarce and scattered, making all collected data easily available and reusable was one of the core ambitions of Exil:Trans. The project does not strive for exhaustiveness, but wants to formulate new questions and contribute a new perspective through carefully collected and curated data on individual translators. The collaboration with a company specialised in data modelling for the creation and long-term storage of FAIR (*Wilkinson et al. 2016*) data provides a blueprint for research data reuse and possible points of contact with further or other research projects in related domains.

The blueprint involves 5 mandatory simple steps to make research data reusable in a way that other research projects can directly integrate it into their own implementations:

1. **Record** the data in a structured way with an easily accessible API (e.g. Excel, Google spreadsheets, XML, JSON ...)
2. **Create** an ontology that holds all the rules and definitions of the data model (RDFS or OWL). The reusability of this model is ensured by aligning it with top-level ontologies such as

CIDOC-CRM (*Velios 2021; Doerr, Light, and Hiebel 2020*) or Basic Formal Ontology (*Otte, Beverley, and Ruttenberg 2022*) for specific concepts such as the event to go into exile. For more general concepts, this is ensured directly by using schema.org (*Patel-Schneider 2014*). At the entity level, in turn, controlled vocabulary, authority files or Wikidata are used.

3. **Transform** your structured data from step 1 to RDF that follows all the rules from the ontology you created in step 2 using the API available (e.g. via a Python script using *RDFlib* (*Carl 2018*)). Of course, there are mapping scenarios from tabular data to RDF. However, the peculiarity arises from the fact that, with regard to the use of modelling patterns (*Hitzler et al. 2016*), a flat structure is not sufficient. Taking into account the possibilities for integrating data cleaning and semantic enrichment (reconciliation) is another strong point in favour of choosing a Python workflow.
4. **Expose** all your data online in a way that is easily accessible and understandable (e.g. create an API/Downloads/Export section for all the data and provide a human-readable description of the data model)
5. Use formats, standards and infrastructures that are suitable for **long-term archiving**. This is made possible with GAMS (*Stigler and Steiner 2018*).
6. Optional but highly useful: Provide a **sample reuse scenario** in your documentation (<https://gams.uni-graz.at/o:exil.tutorial>).

Data recorded and stored in this way helped with further in-depth investigations into translators' lives. Maps were created with a combination of SPARQL queries on the backend and JavaScript on the frontend. These help to identify the translators' networks, and various filters help to analyse the intersection of the various perspectives involved in the translation process. This will extrapolate more general, transindividual patterns on the basis of the established case studies, to identify regularities of translation in exile.

Bottom line, data exposed as Linked Open Data provides us with stable, reusable and interconnectable research and generating such data can be very easy following a simple blueprint.

Bibliography

Bizer, Christian, Maria-Esther Vidal, and Hala Skaf-Molli (2018): 'Linked Open Data'. In *Encyclopedia of Database Systems*, edited by Ling Liu and M.T. Özsu, 2096–2101. New York: Springer. https://doi.org/10.1007/978-1-4614-8265-9_80603.

Carl, Boettiger (2018): 'Rdfli: A High Level Wrapper around the Redland Package for Common Rdf Applications'. <https://doi.org/10.5281/zenodo.1098478>.

Doerr, Martin / Richard Light / Gerald Hiebel (2020): 'Implementing the CIDOC Conceptual Reference Model in RDF'. Version 1.1. <https://www.cidoc-crm.org/sites/default/files/issue%20443%20-%20Implementing%20CIDOC%20CRM%20in%20RDF%20v1.1.pdf>.

Hitzler, Pascal (2021): 'A Review of the Semantic Web Field'. *Communications of the ACM* 64 (2): 76–83. <https://doi.org/10.1145/3397512>.

Hitzler, Pascal / Aldo Gangemi / Krzysztof Janowicz / Krisnadhi Adila / Presutti Valentina (2016): *Ontology Engineering with Ontology Design Patterns: Foundations and Applications*. IOS Press.

Otte, J Neil / John Beverley / Alan Ruttenberg (2022): 'BFO: Basic Formal Ontology'. *Applied Ontology*, 17–43. <https://doi.org/10.3233/AO-220262>.

Patel-Schneider, Peter F. ((2014): 'Analyzing Schema.Org'. In *The Semantic Web – ISWC 2014*, edited by Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig Knoblock, Denny Vrandečić, Paul Groth, Natasha Noy, Krzysztof Janowicz, and Carole Goble, 261–76. Cham: Springer International Publishing.

Stigler, Johannes / Elisabeth Steiner (2018): 'GAMS–Eine Infrastruktur Zur Langzeitarchivierung Und Publikation Geisteswissenschaftlicher Forschungsdaten'. *Mitteilungen Der Vereinigung Österreichischer Bibliothekarinnen Und Bibliothekare* 71 (1): 207–16.

Tashinskiy, Aleksey / Tomasz Rozmysłowicz / Julija Boguna Hrsg. (2022): *Translation und Exil (1933–1945): Namen und Orte. Recherchen zur Geschichte des Übersetzens*. Berlin: Frank & Timme.

Velios, Athanasios (2021): 'Definition of the CIDOC Conceptual Reference Model'. Edited by Chryssoula Bekiari, George Bruseker, Martin Doerr, Christian-Emil Ore, Stephen Stead, and Athanasios Velios. CIDOC CRM Special Interest Group. https://ualresearchonline.arts.ac.uk/id/eprint/18425/1/cidoc_crm_v.7.1.1_0.pdf.

Wilkinson, Mark D / Michel Dumontier / IJsbrand Jan Aalbersberg / Gabrielle Appleton / Myles Axton / Arie Baak / Niklas Blomberg, et al. (2016): 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3 (1): 1–9.