

# Increasing the visibility of Tyrol's cultural heritage through historical newspapers – the triple-open approach of the Zeit.shift project

**Lyding, Verena**

verena.lyding@eurac.edu  
Eurac Research Bolzano/Bozen, Italien

**Franzini, Greta**

greta.franzini@eurac.edu  
Eurac Research Bolzano/Bozen, Italien

## Introduction

Tyrolean tangible cultural heritage is manifested in various artefacts and documents spread across different regions in Austria and northern Italy, spanning the territory of historical Tyrol as it was delineated in the early 20<sup>th</sup> century. The written testimonies are preserved in historical institutions, private collections and public libraries. Efforts to digitize historical Tyrolean documents to safeguard their long-term preservation began about two decades ago with the Austrian Literature Online (ALO) project (Egger and Mühlberger, 2000) and Europeana Newspapers (Pekárek and Willems, 2012). However, the digitization of all relevant historical documents is still far from complete and digitization processes are often pursued by institutions with limited coordination between them.

The fact that many historical documents are only available as paper copies limits their access to expert groups and the public alike.<sup>1</sup> It follows that historical text sources are known and used by only a few people when they could be valuable sources of information for larger parts of society, be those relating to the educational sector, public culture or research.

This is where the Zeit.shift project comes in. Using historical newspapers as a use case, this interdisciplinary and interregional project aims at increasing the visibility and use of historical text sources by bringing together dispersed items in digital form, making them accessible and promoting their value and use through educational and participatory campaigns.

## Project consortium and project goals

Zeit.shift is funded by the European Regional Development Fund and Interreg V-A Italia - Austria 2014-2020. The project started in October 2020 and will run until June 2023. It is a cooperation between two libraries, the Dr. Friedrich Teßmann Library (Bolzano, Italy, project lead) and the University and State Library of Tyrol (Innsbruck, Austria), together with the research Institute for Applied Linguistics at Eurac Research (Bolzano, Italy). The library partners contribute their data and expertise in collecting, cataloguing, digitizing and publishing historical text sources. The research institute contributes its expertise in computational linguistics, Digital Humanities, crowdsourcing and participatory approaches for the (semi-)automatic processing of textual resources with the aim of enhancing their informational content, their searchability and thus their overall use. Seven cultural institutions support the project as associated partners.<sup>2</sup>

The Zeit.shift projects seeks to preserve, develop and communicate the cultural and textual heritage of the historical region of Tyrol by, firstly, digitizing and bringing together Tyrolean historical newspapers from the early 20<sup>th</sup> century (mostly written in German Fraktur); secondly, by upvaluing the data, improving their searchability and making them accessible via a dedicated and free online portal; and, thirdly, by disseminating the data to be used by various stakeholder groups through free educational materials, workshops and participatory activities. The activities are promoted through campaigns, and complemented by free educational workshops and an e-learning course.

## Research and development

The aims of Zeit.shift to extend access, visibility and use of historical text documents call for an open research and development approach. Indeed, the project fosters open procedures and open access in three distinct but interconnected areas of the project implementation: (1) openness in collaboration and knowledge sharing, (2) openness of data and tools, and (3) openness in education and participation.

### Open collaboration and knowledge sharing

Collaboration openness mostly concerns the work of the two library partners, who join forces in digitization efforts which were previously carried out in parallel with little interaction. This includes sharing experiences on good practices, exchanging information on the cataloguing work, agreeing on naming conventions and harmonizing workflows. In addition, both data and digitization costs are shared between the partners and the digitized data will be published through a joint online portal.

The collaboration with the research partner also builds on openness in data and knowledge exchange, but it

is less challenging as in this typical research collaboration the competences of the partners are complementary and thus responsibilities clearly defined.

## Open data and tools

All data that is produced within the project is and will be openly available. The digital copies of the historical newspapers and their metadata are published under Creative Commons Attribution licenses. The computational linguistic processing and enhancement of the newspaper data is based on non-commercial open tools; all scripts, toolchains and other code developed within Zeit.shift are and will be made available via Eurac Research's GitLab repository.<sup>3</sup> Data collected through the participatory activities is and will also be accessible under open licenses.

## Open education and participation

Zeit.shift pursues an active and open approach towards citizen participation and dissemination of project results. One of its goals is to engage the public and raise interest for the historical data that is digitized within the project. This is done by developing initiatives that invite the local population to interact with the data and contribute to their processing and annotation (see #4.2.). These initiatives are promoted through specific campaigns, including stands at conferences and transfer-oriented events such as the *Long Night of Research*, as well as radio and TV broadcasts, announcements in local print media, flyers, stickers and a prize competition (ibid.).

Additionally, Zeit.shift is investing in the formation of multipliers for different stakeholder groups, ranging from professional archivists and library staff to hobby historians ('Chronisten' in German), teachers and teacher trainers. The course materials are jointly elaborated by the partners and are openly shared with the community via multiple institutional repositories. They introduce users to both the basics of digital and automated procedures in text processing and historical studies, and to the participatory activities offered by the project. In addition, the project has published a free Massive Open Online Course (MOOC) to show students and interested lay citizens how historical daily newspapers from North, East and South Tyrol can be used as online sources for research, to illustrate the potential of Digital Humanities and to provide insights into some computational linguistic techniques that facilitate work with historical newspapers.<sup>4</sup>

## Project results

Two years into the project, we can report the following results.

### Digitized historical newspapers

Some 47.631 issues (amounting to 423.782 pages) from 41 newspapers published between 1880 to 1950 in the historical region Tyrol have been digitized and are cur-

rently being made available through the online portals of the Teßmann<sup>5</sup> and Innsbruck University Libraries<sup>6</sup>. The digitized data includes OCR'd text in ALTO-XML format, aligned at the page level with scanned 400 dpi resolution color images.

The data is automatically annotated for Named Entities using the spaCy named entity tagger (Honnibal and Montani, 2017) augmented with a trimmed list of local toponyms ('Flurnamen' in German<sup>7</sup>). Each newspaper page is automatically tagged with one or more topic labels distinguishing between local news, global news, announcements, advertisements, and serial stories. The final, joint newspaper portal, which is still under development, will make use of these annotations to enhance search functionality.

## Participatory activities

Two participatory activities have been developed and are available online. Both are open to all participants alike but were designed for specific user groups.

The first activity (*macro-task*), targets German-speaking specialists (librarians, chroniclers, historians, etc.) and invites them to geo- and semantically tag historical advertisements automatically extracted from the Zeit.shift newspaper collection using the Historypin platform (Fig. 1).<sup>8</sup> Currently, the Zeit.shift collection in Historypin contains some 7,000 adverts from ten different newspapers and, as of December 2022, 27 people have (geo)tagged 351 adverts with a total of 1,668 unique tags. As a by-product of the project, the Zeit.shift team produced the German translation of the Historypin platform to the benefit of the German-speaking community.<sup>9</sup>



Figure 1. An advert from the *Schwazer Lokalanzeiger* about cough sweets in Historypin. The pink and black pin on the map marks the pharmacy's location. Source: <https://tinyurl.com/mzkzzhwz>

The second activity is *Ötzi!*, a custom web game inviting participants between 11-14 years of age to speed-type words extracted from the newspapers, which is designed to help them learn this script whilst contributing to OCR post-correction efforts (*micro-task*).<sup>10</sup> In the game, alpine animals walk in the direction of Ötzi the Iceman<sup>11</sup> looking to harm him while Fraktur words appear on the screen; players must type the words correctly as fast as possible to fend off the animals and thus preserve Ötzi's health (Fig. 2). Knowledge of German is not necessary to play the game but certainly an advantage.

Ötzi! is released under an MIT license and can be played on both desktop and mobile devices. It is made up of two software components.<sup>12</sup> The *frontend* implements user interaction, that is, the gameplay; it is written in Javascript using Phaser<sup>13</sup> and is distributed via Itch.io<sup>14</sup>, the *de facto* standard platform for indie games. The *backend* implements all data flow (providing words to the frontend and receiving game data via an API) and data analysis; it is written in Javascript using Express<sup>15</sup> and is hosted on Eurac Research infrastructure. Ötzi! sources assets (audio files and graphics) published under open licenses only.

Transcriptions anonymously typed by players are collected to test the efficacy of the game as an OCR manual post-correction tool. As of December 2022, the game has been played by 1,754 unique devices and out of the 6,909 words typed thus far, 890 confirm the OCR output, 442 provide corrections and 5,577 are pending automatic evaluation.<sup>16</sup>

In September 2022, the project launched a prize competition for middle schools in South Tyrol to advertise the game to the younger population.<sup>17</sup> The competition and advertising campaign, which ran between 12<sup>th</sup> September and 31<sup>st</sup> October 2022, was very successful in recruiting players from both the target group and the general public, with 1,820 games initiated, 729 games completed, 88.72 hours of total game time (an average 3 minutes per initiated game and 7 minutes per completed game) distributed across 914 unique devices.

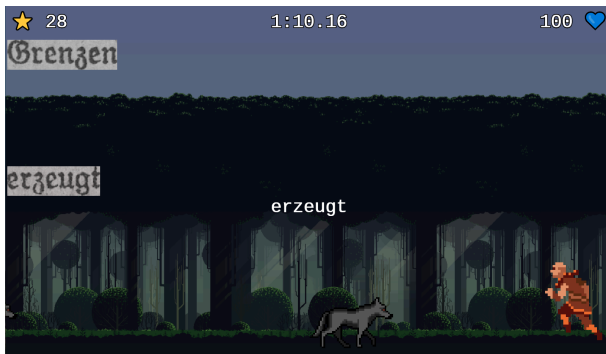


Figure 2. Ötzi! screenshot. The player types Fraktur words as they appear on the screen. At the top of the screen, points accumulated are shown in the left corner, elapsed game time in the center and health status points in the right corner.

## Dissemination and training

Multiple training and dissemination activities have been carried out in both South Tyrol and North Tyrol. To date, the project has conducted nine training workshops for the Historypin activity for a rough total of 100 participants, and another 200 people have been informed about the project activities through presentations at local stakeholder institutions, including Tiroler Landesmuseum Ferdinandeum and Standtarchiv Bozen. The MOOC contains six modules and is followed by 63 participants. Flyers and posters have been distributed by all partners in their vicinity. Several press releases have been publis-

hed<sup>18</sup>, and scientific publications have been presented at three events: a demo session at the Engaging Citizen Science Conference 2022 (CitSci2022)<sup>19</sup> in Aarhus (Denmark) (Franzini et al., forthcoming), and two poster presentations at the annual conference of the *Associazione per l'Informatica Umanistica e la Cultura Digitale* (AIUCD)<sup>20</sup> in Lecce (Italy) (Franzini et al., 2022) and at the 7<sup>th</sup> Austrian Citizen Science Conference<sup>21</sup> in Dornbirn (Austria).

## Joint newspaper portal

The design phase of the Zeit.shift newspaper portal addressed the requirements and technical needs of the three project partners. The portal is currently being implemented by an ICT service provider and will go live in the spring of 2023.

## Conclusions and future work

In this article we describe how an open approach is helping to increase the visibility and support the use of historical resources within the Zeit.shift project. Indeed, we believe that the wider promotion and use of cultural heritage data should move beyond the mere provision of open data to foster the engagement of citizens through education and participatory initiatives, as well as the collaboration among cultural institutions in relation to data, procedures and networks.

At the same time, historical sources are often difficult to process automatically, which makes manual processing by researchers and wider stakeholder groups highly necessary. Follow-up work on the project might focus on developing additional participatory activities to enhance the historical data sources in a playful and educational fashion. This could also include further research on how any data collected from a non-expert audience can be processed and aggregated to improve the original sources in a reliable way. These activities would serve a twofold purpose: improve the quality of digital copies of fragile historical items for long-term preservation, and increase their visibility and use for educational purposes and public cultural awareness.

## Acknowledgements

Zeit.shift has received funding from the European Regional Development Fund under the Interreg Italia - Österreich 2014-2020 Programme (Project no. ITAT3030). We thank the conference reviewers for their helpful suggestions and are especially grateful to our contributing citizens.

## Fußnoten

1. Access to these typically fragile and valuable materials is often regulated by strict protocols, such as assisted visits by appointment.

2. Euregio Tirol-Südtirol-Trentino, Abteilung Tiroler Landesarchiv der Tiroler Landesregierung, Tiroler Landesmuseen, Südtiroler Kulturinstitut, Tiroler Bildungsforum, Südtiroler Landesarchiv der Autonomen Provinz Bozen-Südtirol, Bibliotheksverband Südtirol.
3. <https://github.com/commul>
4. <https://imoox.at/course/zeitshift2022>
5. <https://digital.tessmann.it/>
6. <https://diglib.uibk.ac.at/obvuiibz>
7. <https://www.natura.museum/de/magazine/flurnamen-suedtirols/>
8. <https://www.historypin.org/en/zeit-shift>
9. <https://about.historypin.org/2021/10/13/historypin-in-german/>
10. <https://eurac.itich.io/oetzit>
11. <https://www.iceman.it/en/the-iceman/>
12. The code is available at <https://gitlab.inf.unibz.it/commul/oetzit>
13. <https://phaser.io/>
14. <https://itch.io/>
15. <https://express.js.com/>
16. In our current experimental setup, an OCR'd word is confirmed or corrected if transcribed by at least three different players.
17. Press releases of the 2022 prize competition are available at <https://news.provinz.bz.it/de/news/otzit-startschuss-fur-online-gewinnspiel-fur-kinder-und-jugendliche>; <https://www.suedtirolnews.it/unterhaltung/kultur/online-gewinnspiel-fuer-kinder-und-jugendliche-startet-am-12-september>; <https://www.lavocedibolzano.it/la-biblioteca-provinciale-tessmann-organizza-sino-al-30-ottobre-il-gioco-a-premi-online-otzit/>
18. <https://all4ling.eurac.edu/projects/zeitshift/press/>
19. <https://conferences.au.dk/citsci2022>
20. <http://conference.unisalento.it/ocs/index.php/aiucd2022/aiucd2022>
21. <https://www.oesterreichische-citizen-science-konferenz-2022.com/>

**Honnibal, Matthew and Ines Montani.** 2017. "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing."

**Pekárek, Aleš and Marieke Willems.** 2012. "The European Newspapers – A Gateway to European Newspapers Online." In *Progress in Cultural Heritage Preservation*, edited by Ioannides, Marinos, Dieter Fritsch, Johanna Leissner, Rob Davies, Fabio Remondino, and Rossella Caffo, 654-659. EuroMed 2012. Lecture Notes in Computer Science, vol 7616. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-34234-9\\_68](https://doi.org/10.1007/978-3-642-34234-9_68)

## Bibliographie

**Egger, Alexander and Günther Mühlberger.** 2000. "Ausland. ALO oder die virtuelle Bibliothek der österreichischen Literatur." *Bibliotheksdienst* 34(6): 958-967. <https://doi.org/10.1515/bd.2000.34.6.958>

**Franzini, Greta, Egon W. Stemle, Verena Lyding, Andrea Abel, Johannes Andresen, Karin Pircher, Silvia Gstrein, Barbara Laner, Johanna Walcher, Marritta Horwath, Christian Koessler.** 2022. "Citizen Humanities in Tyrol: A case study on historical newspapers." In *Proceedings of the Eleventh AIUCD Annual Conference (AIUCD 2022)*, edited by Ciraci, Fabio, Giulia Miglietta and Carola Gatto, 236-238. Lecce, 1-3 June. ISBN: 978-88-9425-356-6. DOI: <http://doi.org/10.6092/unibo/amsacta/6848>

**Franzini, Greta, Paolo Brasolin, Verena Lyding, Egon Stemle, Andrea Abel, Johannes Andresen.** Forthcoming. "Zeit.shift: Driving Citizens to Tyrolean Historical Newspapers." In *Proceedings of the Engaging Citizen Science Conference 2022 (CitSci2022)* edited by Nielsen, Kristian, Gitte Kragh, and Lori Nash, Proceedings of Science.