PROCEEDINGS OF SCIENCE

Zeit.shift: driving citizens to Tyrolean historical newspapers

Greta H. Franzini,[a,*] Paolo Brasolin,[a] Verena Lyding,[a] Egon W. Stemle,[a] Andrea Abel[a] and Johannes Andresen[b]

[a] *Eurac Research,*
Viale Druso 1, Bolzano, Italy
[b] *Landesbibliothek Dr. Friedrich Teßmann,*
Via Armando Diaz 8, Bolzano, Italy

E-mails: greta.franzini@eurac.edu, paolo.brasolin@eurac.edu, verena.lyding@eurac.edu, egon.stemle@eurac.edu, andrea.abel@eurac.edu, johannes.andresen@tessmann.it

Zeit.shift is an ongoing digital humanities cooperation between Eurac Research (Bolzano, Italy), the Landesbibliothek Dr. Friedrich Teßmann (Bolzano, Italy) and the Universitäts- und Landesbibliothek Tirol (Innsbruck, Austria) for the digital preservation, enrichment and dissemination of the textual heritage of historical Tyrol. Using digitised German-language historical newspapers as a use-case, Zeit.shift is, among other project goals, developing two participatory tasks—image annotation and text transcription—to increase citizen awareness of, and active engagement with, these collections, while exploring the potential of the tasks themselves as data acquisition tools for research. As of July 2022, results show 25 contributors for the annotation activity with 303 (geo)tagged images and 1,430 unique tags, and 288 played games from 241 unique devices totalling 1,910 words transcribed for the web game.

Engaging Citizen Science Conference 2022 (CitSci2022)
25-26 April 2022
Aarhus University, Denmark

*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

PoS(CitSci2022)003

1. Introduction

> There are many reasons that we put digital collections online. The single most important is to make history accessible so that we can invite students, researchers, teachers, and the public to explore and connect with our past. [1, p. 127]

In acknowledging the power of cultural heritage as a means of strengthening social cohesion and cultural diversity, the Faro Convention calls for the involvement of all citizens in its definition and management, and for "an attractive approach to heritage education [...] that embraces new technologies and retains young people's interest" [2, p. 4]. One such approach is Citizen Science (hereafter CS) and yet recent studies have shown that the social sciences and the humanities account for only 11% of CS practice [3]. Indeed, it is only recently that CS scholars have started using the term Citizen Humanities to describe a field primarily "driven by professional (university) researchers and cultural heritage institutions" focussing on engaging "members of the public in humanistic research with the aim of (co-)producing knowledge, or even change" [4, p. 174] with tasks as diverse as the creation of linguistic resources or enhancing historical records for others to use. These projects typically follow a top-down approach, whereby data contributed by citizens is used by researchers to answer research questions and advance academic progress [5, p. 6], and are not particularly concerned with scale (e.g., increasing the quantity of data or the number of volunteers involved) so much as creating meaningful cultural engagement and influencing a small number of stakeholders [4, p. 172].

Here we present Zeit.shift—A digital journey into yesterday's future, an ongoing digital humanities cooperation funded by the Interreg Italy—Austria Programme[1] for the digital preservation, enrichment and dissemination of the textual heritage of historical Tyrol (2020-2023)[2]. The project aims at digitising a large number of historical newspapers held at the Landesbibliothek Dr. Friedrich Teßmann and at the Universitäts- und Landesbibliothek Tirol, and processing them with standard computational linguistic tools, namely Named Entity Recognition and topic classification, to make the documents available and searchable via a newly created digital portal. In addition to creating the technical environment to access the newspapers online, Zeit.shift encourages the public to work with these materials. This is done by, on the one hand, providing online training materials (a MOOC - Massive Open Online Course) and workshops on how to use the newspapers and, on the other, by developing two contributory activities to increase citizen awareness of, and active engagement with, these historical collections (1880-1950), while exploring the potential of the activities themselves as data acquisition tools for research. The paper presents these two CS activities in greater detail.

2. Research questions

Zeit.shift responds to two research questions: "how can we improve the discoverability of Tyrolean textual heritage?" and "how can we increase citizen's awareness of, and engagement with, Tyrolean textual heritage?"

[1] https://www.interreg.net/en/default.asp
[2] https://all4ling.eurac.edu/projects/zeitshift/

PoS(CitSci2022)003

3. Methodology

We address these questions along two avenues. Firstly, by making the historical data available and searchable online as digital collections; the project has, in fact, digitised approximately 420,000 pages of historical newspapers[3]. Secondly, by developing two distinct participatory and distributed intelligence tasks to actively engage citizens in getting to know the material, understanding how it can be used and how everybody can contribute to increasing its value and usefulness. While these activities are primarily geared towards engagement, their design also serves to collect data for research-oriented tasks, namely topic classification and OCR post-correction. Where possible, we leverage existing and open technological solutions to save resources and reap the benefits of tested applications [6, p. 18].

3.1 Macro-task: Topic classification

A macro-task [7] invites specialists, hobbyists and the general public to (geo)tag newspaper advertisements uploaded to the Historypin[4] platform to help recreate the economic landscape of Tyrol from roughly 100 years ago. In this task, participants are asked to, where possible, locate ("pin") an advert on a map, either roughly or precisely, and to add descriptive keywords ("tags") to identify its content. We believe historical advertisements lend themselves well to public engagement not only because of their concise, visual and often amusing essence, but for the captivating sense of nostalgia they evoke in reminding us of products, traditions, people or businesses from years passed (Fig. 1). As such, the task relies on participants' affective motivation [8], expecting—as is often the case in CS projects—a small number of particularly active or returning participants.

The CS initiative that most closely resembles this macro-task is Altes-Leipzig, a reconstruction of historical Leipzig made possible thanks to digitised archival documents and genealogical information provided by citizens[5].

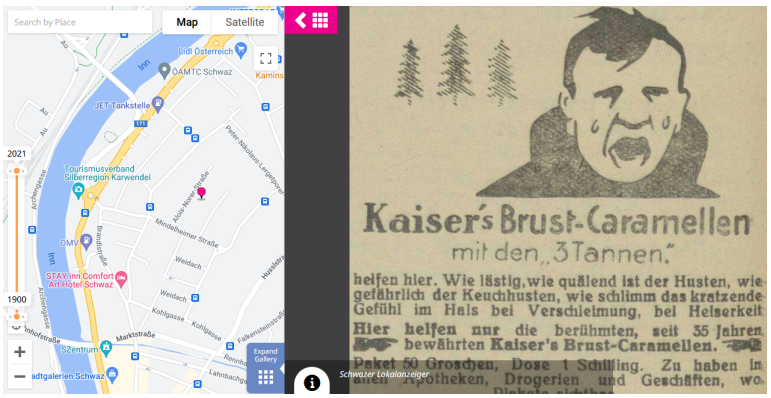

Figure 1: An advert about cough sweets in Historypin. The chemist's exact location on the map is marked by a pink pin. Advert licensed under a CC BY 4.0 licence. Map data ©2022 Google.

[3]Corresponding to 41 newspapers, some 47,000 issues, 61 million sentences, 840 million tokens and 13 terabytes of data.

[4]https://www.historypin.org/en/zeit-shift/

[5]https://www.altes-leipzig.de/

PoS(CitSci2022)003

The Historypin platform was chosen as the best-fitting technological solution for a number of reasons, namely its popularity with GLAM institutions, its long-term data storage strategy, its open licensing, its established community of active users, as well as its range of functionality, which for users include, among others, the option of pinning images on a map, adding comments, creating virtual tours, sharing content and saving copies of the images in personal collections; for content providers, the platform supports bulk-upload/download of data and further automation through its API. Historypin is free to use but premium packages are also available offering data providers more storage space, additional functionality and long-term support[6].

To select and upload material to the platform, we firstly use an adapted version of a Europeana Newspapers script to automatically extract the adverts from the newspaper pages[7]; next, we manually sort the extractions removing false positives and any duplicates; we then use a custom script to pair the adverts with their respective metadata (i.e., licence, attribution, source, identifier); finally, we use the platform's bulk uploader to upload all adverts to the platform. For a balanced representation of all Tyrolean communities, we select and upload newspapers from as many cities as possible. To date, the platform hosts over 6,500 Zeit.shift adverts from ten different newspapers, with more to come.

Zeit.shift images and metadata are published in Historypin under CC-BY licences, whereas citizen-contributed data is released by Historypin under a CC0 1.0 licence. All (geo)tagged data can be downloaded from Historypin and the semantic tags used for the improvement of text classification models for German.

The recruitment of participants is mostly done through the organisation of workshops for German-speaking specialists (e.g., librarians, chroniclers, historians) on this specific activity. To date, the project has run eleven workshops and more are planned for the coming year. Twitter and Instagram posts are published in English on a weekly basis to entice a larger variety of online users but, to the best of our knowledge, these efforts have yet to translate into actual contributions.

3.2 Micro-task: OCR post-correction

The micro-task [7] is a custom web game targeting casual gamers of all ages and middle school children in particular (ages 11-14). Inspired by the DigitalKoot [9][8] and TypeAttack [10] games, both developed to correct the noisy OCR of historical documents but no longer available, our web game, Ötzit!, takes its name after Ötzi the Iceman[9] and its purpose is to teach young generations Fraktur so that they may access and explore historical materials in this script. In the game, alpine animals walk in the direction of Ötzi looking to harm him while Fraktur words drop from the sky; players must type the words correctly as fast as possible to fend off the animals and thus preserve Ötzi's health (Fig. 2)[10]. Transcriptions typed by players are collected to test the efficacy of the game as an OCR manual post-correction tool.

Ötzit! relies on a very simplistic formula to attract the attention of casual gamers. For game uptake, we have put in place and are planning a variety of extrinsic motivators. In-game scores and leader-boards[11] are implemented according to basic gamification principles to stir player

[6] See https://about.historypin.org/content/uploads/2019/02/Historypin-Service-List.pdf
[7] https://github.com/cneud/alto-tools
[8] https://scistarter.org/digitalkoot
[9] https://www.iceman.it/en/the-iceman/
[10] https://eurac.itch.io/oetzit
[11] No user data aside from the scores is tracked, and no registration is required to further lower entry barriers.

PoS(CitSci2022)003

Figure 2: Ötzit! screenshot. As Fraktur words drop from the sky, the player's corresponding transcriptions appear in the centre of the screen. At the top of the screen, points accumulated are shown in the left corner, elapsed game time in the centre and health status points in the right corner. Image licensed under a CC BY 4.0 licence.

competitiveness either against oneself or others. Real life incentives will be provided too, in the form of free tickets, coupons and gadgets gifted to top players[12]; this is arranged in collaboration with local institutions, cinemas and shops, and the recruitment campaign is due to launch in autumn 2022 at the start of the school year. The campaign will also consist of school workshops conducted in English, during which pupils will be introduced to the Zeit.shift project and the game.

Ötzit! is released under an MIT license and is made up of two software components[13]. The frontend implements user interaction, i.e. the gameplay; it is written in Javascript using Phaser[14] and is distributed via Itch.io[15], 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[16] and is hosted on Eurac Research infrastructure. The game can be played on both mobile and desktop devices, provided these are connected to the Internet.

4. Results and Discussion

Besides the digitisation of over 420,000 pages worth of historical Tyrolean newspapers, we summarise the progress, expected results and main challenges of the project in relation to the three fundamental elements of CS [11].

Inclusion. Zeit.shift has developed two participatory activities for two different audiences with differing modes of interaction. The macro-task relies on intrinsic motivation and since its launch

[12]Since no registration is required, players are invited to provide their email exclusively to receive the prizes.
[13]The code is available at https://gitlab.inf.unibz.it/commul/oetzit
[14]https://phaser.io/
[15]https://itch.io/
[16]https://expressjs.com/

PoS(CitSci2022)003

in October 2021 has attracted 25 contributors (some returning), who have (geo)tagged 303 adverts with a total of 1,430 unique tags. The micro-task, on the other hand, which was pre-launched in April 2022 at the CitSci2022[17] conference for a first round of feedback, depends on both intrinsic and extrinsic motivation and is expected to attract more, albeit casual, users. As of July 2022, the game has been played 288 times from 241 unique devices for a total of 1,910 words transcribed (805 transcribed by at least three users where the transcription matches the OCR; 59 words transcribed by at least three users where the transcription differs from the OCR; and 1,046 words pending transcription agreement). These preliminary numbers are encouraging, seeing as the project has set out to achieve 10,000 actions between the two activities.

Contribution. Zeit.shift's contributions for science and scientists include (meta)data enrichment and further –but yet to be verified– examples of dual-purpose and mutually beneficial activities (engagement and research). Project contributions for the public include centralised free access to a large digitised collection of historical text, an e-learning course[18], as well as a German translation of the Historypin platform[19].

Reciprocity. While well advertised in local newspapers, magazines, radio and TV interviews, cultural networks and CS platforms[20], Zeit.shift's main weakness lies in community building and management. Owing to the absence of a dedicated communication officer, this crucial task of motivating and disseminating scientific information to the public is shared between the project partners and conducted in an irregular fashion to the detriment of participation retention.

5. Conclusion

This paper outlines two participatory CS tasks in the context of the Zeit.shift project, an ongoing Austro-Italian digital humanities cooperation for the digital preservation, enrichment and dissemination of the textual heritage of historical Tyrol: an image annotation activity hosted on the third-party platform Historypin and a custom transcription web game. Our CS activities serve a twofold purpose: on the one hand, they are designed to attract citizens' attention and engage them with the data through playful interaction. On the other, they provide researchers with indicators on the reliability and potential of the activities and of harvested data to enrich these historical resources and/or improve the computational models trained to analyse it.

The success of the CS tasks with respect to the project's research questions will be measured on the number of citizens involved, the contributions or actions completed, and on the quantity of additional traffic these tasks will direct to libraries where the historical newspaper collections are held.

The final outcome of Zeit.shift will consist of a digital library of Tyrolean historical newspapers (over 420,000 pages published between 1880-1950) integrating project outputs and activities, as well as data contributed by participants. In publishing their data, the project seeks to empower Zeit.shift users to act as catalysts to help raise awareness of the existence of these historical resources and of their cultural value.

[17] https://conferences.au.dk/citsci2022/
[18] https://imoox.at/course/zeitshift2022
[19] https://about.historypin.org/2021/10/13/historypin-in-german/
[20] See https://all4ling.eurac.edu/projects/zeitshift/#featured-in for featured locations.

PoS(CitSci2022)003

Acknowledgements

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

References

[1] T. Owens, Digital cultural heritage and the crowd, *Curator: The Museum Journal* **56** (2013) 121.

[2] *The Faro Convention's role in a changing society*, Council of Europe Publishing, Strasbourg (2021).

[3] S. Hecker, L. Garbe and A. Bonn, The European citizen science landscape—a snapshot, in: *Citizen science. Innovation in open science, society and policy*, eds. S. Hecker et al. (London), 190–200, UCL (2018).

[4] B. Heinisch, Citizen Humanities as a Fusion of Digital and Public Humanities?, *Magazén* **1** (2020) 143.

[5] L. Tauginienė, E. Butkevičienė, K. Vohland, B. Heinisch, M. Daskolia, M. Suškevičs et al., *Citizen Science in the Social Sciences and Humanities: The Power of Interdisciplinarity*, vol. 6, Palgrave Communications (2020).

[6] M.R. Mondardini, U. Roffler, T. Eliseeva, O. Höhener, D.M. Kretzer, R. Lenart-Gasiniec et al., *Practicing Citizen Science in Zurich: Handbook*, Citizen Science Center Zurich, Zurich (2021), 10.5167/uzh-207987.

[7] M. Ridge, Crowdsourcing our cultural heritage, in: *Routledge International Handbook of Research Methods in Digital Humanities*, eds. K. Shuster and S. Dunn, Routledge Handbooks Online (2020), 10.4324/9780429777028-26.

[8] B.S. Bloom and D.R. Krathwohl, *Taxonomy of educational objectives: The classification of educational goals. Handbook I, Cognitive domain*, David McKay Co., Inc. (1956).

[9] V. Miettinen, Digitalkoot: electrifying the Finnish cultural heritage, in *Proceedings of the 4th ACM workshop on Online books, complementary social media and crowdsourcing*, (New York, NY, USA), pp. 55–56, Association for Computing Machinery, 2011, 10.1145/2064058.2064071.

[10] L.T. Jovian and O. Amprimo, OCR correction via human computational game, in *2011 44th Hawaii International Conference on System Sciences*, pp. 1–10, IEEE, 2011, 10.1109/HICSS.2011.319.

[11] Y.N. Golumbic, D. Orr, A. Baram-Tsabari and F. Barak, Between Vision and Reality: A Study of Scientists' Views on Citizen Science, *Citizen Science: Theory and Practice* **2** (2017) 1.

PoS(CitSci2022)003