Digital Scholarship in the Humanities, 2025, i46–i65
https://doi.org/10.1093/llc/fqae018
Advance access publication 16 April 2024
Full Paper

OXFORD

Are Digital Humanities platforms facilitating sufficient diversity in research? A study of the Transkribus Scholarship Programme

Joseph Nockels [1,*], **Paul Gooding**[2], **Melissa Terras** [3]

[1]Department of English Literature, The University of Edinburgh, Edinburgh, EH8 9JU, United Kingdom
[2]Information Studies, University of Glasgow, Glasgow, G12 8QH, United Kingdom
[3]Design Informatics, Edinburgh College of Art, The University of Edinburgh, Edinburgh, EH3 9DF, United Kingdom

*Corresponding author. Joseph Nockels, University of Edinburgh, School of Literatures, Languages and Cultures, 50 George Square, Edinburgh, EH8 9JU, United Kingdom. E-mail: j.h.nockels@sms.ed.ac.uk

Abstract

To what extent do Digital Humanities (DH) platforms support access to diverse user cohorts? We take, as a case study, the Automated Text Recognition (ATR) platform Transkribus and its Transkribus Scholarship Programme (TSP), which provides free processing credits to eligible users. Using a mixed methods approach we address the following questions: What are the demographics of those using the TSP scheme? What work is enabled by such a scheme? How can this inform more equitable access to DH platforms? The findings demonstrate how ATR tools are currently used and made accessible. TSP applicants are overwhelmingly students ($n = 111/156$, 71.15 per cent) drawn from universities and research institutes, mostly in Europe, but are globally distributed; representing institutions that do not hold shares in Transkribus, and indicating a diverse user pipeline. Further work is required to increase potential benefits of the scholarship and to ensure sustainability. Increased dialogue between the Recognition and Enrichment of Archival Document-COOP and applicants would assist in the calculation of processing costs. We show financial—or in-kind—support is necessary to increase access to paid-for platforms, ensuring a diversity of DH research. We also provide recommendations for platform providers and funding bodies regarding access and the impact this can have, including locating a sustainable balance between absorbing the costs of maintaining DH or digital scholarship tools and providing sufficient support and training to further enable diverse research.

Keywords: Automated Text Recognition (ATR); sustainability; Transkribus; Transkribus Scholarship Programme (TSP); freemium models.

1. Introduction

1.1 Inequality in DH platforms

The Digital Humanities (DH) has, with increased computer power and connectivity, gradually shifted towards more elaborate tasks where historical materials are mediated through computational technologies (Pons 2020: 22–28). This work is enabled, in part, by algorithmic methods like Automated Text Recognition (ATR),[1] supporting the image-to-text recognition of handwriting and complex printed materials. However, despite this shift, access to materials, tools, and infrastructures remains profoundly unequal. This inequity echoes Hawkins' (2021: 1) considerations surrounding the control of data in collaborative DH digital initiatives. Financial resources, language, hardware availability, and access to training act as both isolated and combined factors to inaccessibility. As a result, Global North perspectives remain dominant (Risam 2015). Galina (2013) presents DH as slow to theorize issues surrounding access, equity, race; gender, class, and environment. With DH being primarily driven by Global North institutions in the US and UK,[2] Jenkins and Myers (2022) argue that this slow reactivity affects how research is controlled, presented, and disseminated. This is beginning to be rectified, with Bordalejo and Murray stating that '… conversations about diversity have been increasingly visible in the digital humanities community' (2016: 80). However, they also argue that we must go beyond positioning diversity as the simple inclusion of a range of perspectives, and instead as a crucial intellectual aspect to the discipline (Bordalejo and Murray 2016: 80).

In specific DH projects, diversity has clearly been seen as necessary for cultivating new values and

perspectives, reducing the knowledge divide felt in areas like the Global South[3] (Roig-Sanz 2018). Recent work has addressed cultural and linguistic nuances within DH (Gil and Ortega 2016) and international cross-cultural representation on projects (Galina 2013; Risam 2019). For instance, Murray's (2020) work explores how non-Western cultural concepts and intellectual categories might redefine DH methodologies. Elsewhere, discussions of Global South experiences in DH have proposed the need to rethink language to change research agendas (Fontolan 2020: 191). Projects like these, as well as Casanova and Sarmiento's (2021) analysis of DH projects in the Philippines, present a promising future for DH research, with a greater diversity of research avenues opening, growing, and expanding (Roig-Sanz 2022). For this to continue, tool design must consider this diversity of cultural experience, providing meaningful openness and accounting for, as Hawkins (2021: 2) states the 'network of interdependencies and power relations' seen in many DH initiatives, presenting a need to incorporate cultural protocols within access. One necessary step is, therefore, to understand how, and to whom, tools are made accessible.

In response, this article asks whether and how free processing schemes help to make DH platforms more accessible and equitable. We explore the extent to which the free processing scheme of the ATR platform[4] Transkribus enables access among groups of more precarious users. Recognition and Enrichment of Archival Documents (READ), who provide Transkribus, states that the Transkribus Scholarship Programme (TSP) is aimed at students, those conducting workshops, and those lacking funding for their research.[5] The TSP is open globally to students and teachers from any discipline, although the scholarship must align with a specific project.[6] Although not mentioned explicitly by READ in initial publicity, we broadened the latter group to include early career researchers (ECRs), as these users became a sizeable presence in our study of TSP requests.[7] The TSP continues to be publicized through READ's website[8] having been initially mentioned within the Transkribus members' mailing list (Walcher 2020). Whether knowledge of the TSP remains reliant on existing connections to the READ-COOP is answered in our findings section. As such, this article presents an analysis of who is actually making use of the scheme, via content analysis of both online applications to the free credit scheme and conference proposals to the 2022 Transkribus User Conference (TUC), and a survey of TSP applicants and READ staff. Our findings demonstrate who currently benefits from the TSP. In doing so, we reveal the extent to which current free processing schemes are facilitating diverse research and identify inequalities in access. We then provide recommendations for other platform providers and funding bodies regarding access to DH and Digital Scholarship (DS) platforms, and the impact this can have on promoting equal rights, inclusion, and treatment. This is essential in ensuring objectivity and respect in academic spaces, identifying identity-based advantages and barriers, working to correct and address imbalances (Dewidar, Elmestekawy, and Welch 2022: 1).

1.2 Transkribus background

Transkribus is the largest consumer-level ATR platform, allowing for automated image-to-text recognition, broadening access to historical collections. Users can produce machine-processable transcripts of images, which can be presented in a variety of formats such as TEI marked-up resources (Mühlberger et al. 2019: 960–961). As of 6 January 2024, Transkribus' internal data reported 400 interacting users per day, uploading on average 25,000 images and training 15 models for recognizing text daily. In total, 51.5 million images of historical documents have been processed through the platform since 2015 by 171,307 active users. A sum of 25,558 ATR models has been trained. Transkribus projects reflect the breadth of text-based DH research, with recognition models initially being trained on the manuscripts of 18th century philosopher Jeremy Bentham (Terras and Causer 2014), materials from Ethiopia/Eritrea (Universitat Hamburg 2022) as well as 18th–19th century Bengali print, documents written in Malayalam, and 19th century Devanagari scripts (READ-COOP 2021).[9] Whereas OCR relies on character isolation, Transkribus uses a Convolutional Neural Network to process entire lines of text, scanning them in various directions and putting the gained data into a sequence (Liakos 2019). In this way, the ATR platform incorporates statistical information to make predictions in recognizing strings of text (Strauss, Weidemann, and Labahn 2017: 5). About 15,000 words of manual transcription is recommended by the platform to produce enough ground truth data,[10] to build an accurate model tailored to a specific collection.

Transkribus is not the only ATR provider. MONK, developed at the University of Groningen, has produced accurate results using pattern matching since 2009, helping scholars identify text, its date (style-based dating), and author (writer identification) (Schomaker 2020: 221–226). eScriptorium, an interface for the Kraken OCR system,[11] is another ATR software using neural networks, although it remains reliant on character isolation, tagging glyphs with set characteristics, and using corresponding variant image files for model training.[12] Launched in April 2023, and developed by the Digital Infrastructure

Department at the KNAW (Royal Netherlands Academy of Arts and Sciences) Humanities Cluster and the Nationaal Archief of the Netherlands, Loghi is an open-source text recognition software, aimed at making scanned historic documents digitally readable and searchable through neural networks similar to Transkribus.[13] That said, Transkribus, due to its user base, now pay-to-use model, and allied free processing scheme,[14] provides a highly relevant case study of equitable access. We therefore hope that our findings will allow the developers of other DH and DS platforms to draw parallels and, in turn, make access to their products more equitable.

1.3 Transkribus and sustainability

Transkribus was funded initially by an EU FP7 grant as tranScriptorium,[15] releasing its first product in 2015, thereafter the system was maintained and developed by READ supported by an EU Horizon grant until late 2020 (Mühlberger et al. 2019: 957). With this grant ending, Transkribus is now maintained by the democratically organized READ-COOP, which forms a European Cooperative Society (SCE) of 175 shareholders across more than 30 countries (at time of writing) (Stauder 2024).[16] READ-COOP members pay for a certain level of shares,[17] gaining processing discounts, as well as being included in discussions regarding development general of the software, including future feature design, through monthly members' meetings and the TUC. In contrast, the TSP provides a bundle of entirely free credits, with no prior purchase of shares needed. This removes financial barriers to Transkribus use, although free processing recipients have no direct say in the ATR tool's development. On 19 October 2020, Transkribus transitioned from a free-to-use model to a pay-for credit structure, with computational actions corresponding to a required cost.[18] This move to a pay-for-credit model was done to ensure the platform's sustainability, defined by Madsen and Hurst (2018: 17) as the maintenance of a service or product at a certain rate and level for an extended period of time without interruption. For a platform to be sustainable, various functional requirements are needed: appropriate metadata; a place to store data such as a content management system, a way to render objects in a browser with permanent links, a method of engaging users, and a way of updating the data provided. Of course, solid funding and a great deal of human resource is also necessary (Madsen and Hurst 2018: 35).

Transkribus' mission is to make documents from the past legible to new audiences, enabling collections to become more accessible. However, its broad applicability brings with it a necessity for both scalability and long-term sustainability. Owens and Padilla (2021: 338–339) warn that platforms can easily lose focus when they attempt to reach broader audiences through standardization and the building out of tools. Broader all-encompassing strategies for sustainability have also been challenged, with Van Zundert (2012: 165–166) calling for more '… flexible small-scale research focused development practices', producing specific solutions for specific demands and ascribing projects meaning when they are aligned with particular user requirements. This relates to previous ideas surrounding planning DH projects, with Cummings et al. suggesting that work '… should be driven by local needs. Rather than trying to do everything in an unfocused manner, strategize, and be selective. A narrow and focused mission should guide priorities' (2020: 15). As such, sustainability and scalability, while connected, are often in conflict. The needs of specific communities must be considered if DH research is to be supported and diversified. This is juxtaposed, however, by the financial requirements of sustainability, with both technical infrastructures and professional research activities needing resourcing.

Fenlon (2020: 3) highlights that DH work is often funded on short cycles aimed towards technological innovation, experimentation, and development, often relying on bespoke and fragile infrastructure. The maintenance of inactive or archived projects often becomes the responsibility of an IT department, the institution's library, or a DH centre, likely juggling maintenance of a portfolio of projects, which should be a core activity, alongside their other day-to-day responsibilities (Boyles et al. 2018: 694). As such, few projects achieve the level of interest needed to hire full-time maintenance staff, rarely enduring 'beyond the interest and involvement of their initial creators' (Fenlon 2020: 3). As Posner (2016) states, many want to believe that DH '… can be agile and innovative, like Silicon Valley says it is, with short-term grants, app contests, and temporary labor … But this is not how one supports careful, enduring scholarship and teaching'. We should also be cognisant of what types of projects are being privileged by funded schemes, and particularly the imbalance in available research funding between the Global North and Global South.

The READ-COOP is an outlier when it comes to sustainability in DH projects, building commercial mechanisms to enable the support of Transkribus infrastructure beyond its EU funded period. The organization has employed additional staff and improved its digital services.[19] As such, the READ-COOP has shifted from project-funded development to a self-sustaining organization achieved via initial financial support, the introduction of a paid-for model, and the impact of securing major processing contracts. With Transkribus' operational costs entirely covered, the

organization can use profit—opposed to adopting a not-for-profit—motive, to focus on research and development, coupled together with providing a service users want and are willing to pay for; whereas other projects with less sustainable business models understandably opt for less experimental approaches (Cummings, Roh, and Carraway 2020: 8).

1.4 Transkribus and freemium models

Sustainability can also be seen in terms of consumption, with a project replenishing its resources as quickly as it uses them (Drucker 2021: 87). The READ-COOP's adopted credit model forms a freemium structure, whereby basic functions are given to the user for free and additional features are repeatedly sold to generate revenue (Ross 2018: 127–128). The success of freemium models varies (Gu, Kannan, and Ma 2018). Deng, Lambrecht, and Liu (2021: 3) also stress that '[T]he free version may allow consumers to sample the product before making a purchase decision and subsequently increase demand of the paid version, but it may also cannibalise demand of the paid version'. That said, freemium strategies are seen to be a way of moving beyond one-off customer interactions towards long-term engagement (Ross 2018: 128). Most aspects of Transkribus remain free, apart from ATR model training (which requires added processing from READ servers). Every new account receives 500 free credits (enough to process approximately 500 handwritten pages) and must then purchase additional credits. These free credits also allow users to judge the accuracy of outputs before deciding whether to spend money on the tool. For individual researchers working on smaller collections, these free credits could be enough to train an accurate model for keyword spotting (kws), facilitating research and discoveries, making the need to purchase additional features unnecessary. Therefore, whether users purchase processing credits is dependent on the size of their collections and ambitions.

Terras's survey of the user community (2022: 197) found that some users were worried that the shift to a freemium model would undermine the values of the platform, with a third of respondents citing worries that it would jeopardize the tool's transparency, accountability and semi-open licensing. In our previous work, consulting institutional users of Transkribus in 2021, the move to a pay-for model was also highlighted as divisive: 46.87 per cent ($n = 15$) of users agreed that the credit-model had impacted their access.[20] It should be stressed that READ was established with public EU funds (but was not eligible to receive public funds for sustainability, despite its prize-winning status).[21] In this case, the establishment of the cooperative was not simply altruistic, but also necessary to keep infrastructure functioning. As we show, the TSP is emblematic of the READ-COOP providing a degree of financial sustainability. By supporting precarious users, the TSP enables the READ-COOP to partially ameliorate user concerns around equitable tool access, while also focusing on continued sustainability and profitability by increasing long-term engagement with Transkribus.

1.5 Financial models for common DH tools

This section outlines the different models used in resourcing common DH tools. Following resources from Southern Connecticut State University[22] and the Massachusetts Institute of Technology (MIT), which provided an organized initial list of DH tools and their application (taken as random examples of the kind of lists common on institutional webpages),[23] we found numerous cases of open-source structures, defined as tools which allowed the easy modification of source code and had licenses allowing for derivatives.[24] Although not nearly a full list, these tools ranged from those supporting text analysis,[25] machine learning,[26] collection management,[27] network analysis,[28] and the plotting of geographic information.[29] While open-source software was prevalent, these tools provided source code or a simple web environment without other infrastructures like cloud storage or user-friendly interfaces. Table 1 shows the applications found with an associated cost, which tend to provide access to a platform or service, supporting a broad range of uses in the humanities.

Several freemium structures were found, with costings based on trials expiring, processing, and storage, usually in the form of monthly and annual subscriptions. Cloud storage, regular platform promotion, and maintenance, all have associated costs, seen with tools basing their subscriptions on these factors. Reviewing this, the TSP appears an uncommon structure, providing free processing for students, teachers, and those with limited funding, resourced by READ-COOP membership and Transkribus' credit scheme. No other free scholarship schemes were found, warranting further study and consideration in how DH platforms are facilitating research, especially from groups with limited ability to pay. With the TSP being our case study, the following section explains our methods in analysing whether the READ-COOP is successful in facilitating access to communities with low financial resources, impacting the breadth of research being carried out using the ATR platform.

1.6 Methodology

This study used a mixed-methods approach, involving a content analysis of emails automatically sent from READ staff detailing the TSP applications received to

Table 1. Common DH tools with associated costs as of 5 May 2023.

Tool	Description	Model	Cost (individual/ institutional)	Scholarship scheme
ChatGPT[a]	Conversation AI tool using Large Language Model (LLM)	Freemium (ChatGPT Plus), faster processing	$20 per month for ChatGPT Plus/no specific institutional plan	No
From the Page[b]	Crowdsourcing platform for description, indexing and transcription	Freemium, 200 page upload free	$100 per month researcher subscription/ $360 small institution, $600 large institution	No
Jisc Online Surveys[c]	Online survey builder	Freemium, basic tool free	Not accepting new registrations	No
Mendeley[d]	Reference management software	Freemium, basic storage (2 GB) free	$55 per month (5 GB), $110 (10 GB), $165 (Unlimited)/no specific institutional plan	No. 5 GB plan billed as 'great for students'
Omeka (Classic, .net)[e]	Web publishing platform for digital collections	Open-source, fees for added support (hours per month) and hosting (size of collection and version)	Support $1,250–$5,000 per annum. Hosting, $2,500–$12,500. .net lowest option $35 per annum (2 GB storage)/platinum model marketed for institutions £1,000 (50 GB storage)	No
Oxygen XML editor[f]	Suite for editing, developing and authoring XML documents	Freemium, 30-day free trial	$229–$317 for 6 month subscription, $405–$568 year/departmental license $4,696–$5,380, site license $11,481–$13,153	No. Academic license offer for twenty-five terminals using Oxygen XML editor. $1,108 for 1 year with support and management
Survey Monkey[g]	Online survey builder	Freemium, basic tool free	£31.42–£94.26 per month (depending on version)/enterprise version for organizations by request	No. student and educator discount, $30.16–$87.98 per month.
Zotero[h]	Reference management software	Freemium, basic storage (300 MB) free	$20 annually (2 GB), $60 (6 GB), $120 (Unlimited)/organizational licenses starting at $30 per user	No

[a] ChatGPT, 'Introducing ChatGPT Plus', accessed 5 May 2023.
[b] From the Page, 'Pricing', accessed 5 May 2023. https://fromthepage.com/pricing
[c] Jisc Online Surveys, 'Features—Pricing', accessed 2 May 2023. https://www.onlinesurveys.ac.uk/features- pricing/
[d] Mendeley, 'Pricing', accessed 2 May 2023. https://www.mendeley.com/settings/billing/v2/upgrade/
[e] Omeka, 'Services', accessed 5 May, 2023. https://omeka.org/services/
Omeka.net, 'Pricing', accessed 4 January 2024. https://www.omeka.net/signup
[f] Oxygen XML, 'Buy Oxygen XML Editor', accessed 5 May 2023. https://omeka.org/services/
[g] Survey Monkey, 'Plans and Pricing', accessed 2 May 2023. https://www.surveymonkey.co.uk/pricing
[h] Zotero, 'Storage', accessed 2 May 2023. https://www.zotero.org/storage

identify key themes. However, as this data was not enough to understand broader contextual factors around user needs, we used a survey to understand the experiences and expectations of TSP applicants. This followed previous surveys of the Transkribus user community first by Terras after the 2018 TUC, published in 2022, and then by the lead author (2021): though the latter focused on institutional users of the ATR platform. Alongside this, we performed a short survey of READ staff and content analysis of proposals sent to the September 2022 TUC. Bringing together these data streams was essential in providing

recommendations concerning free processing schemes in DH and enabling diversity in research. The following sections will outline the capture and nature of the email dataset, explain our content analysis method, and outline details of the surveys undertaken.

1.6.1 Content analysis of free processing requests

We first analysed free processing requests to the TSP. Those applying to the TSP must fill out an online form,[30] the fields gathered are given in Table 2, indicating the number of credits they expect to need, their home institution where applicable, a short description of their work, and their current academic affiliation. If accepted, credits are then allocated, and an email notification is sent from READ. Credits can be checked and monitored through Transkribus' online portal.[31] Applicants can request up to 3,000 credits, equating to the recognition of 3,000 handwritten or 18,000 printed pages (which requires less processing from READ's servers).[32] The applicant also has the option to be included in Transkribus' 'Success Stories' blog page,[33] if they are willing to provide more details of their project upon completion.

These emails were supplied by the Managing Director of READ, automatically through Outlook whenever a request was submitted, between 7 November 2020 and 16 March 2022. They were then indexed in Excel: where they could be refined, tabulated, and visualized.[34] In total, 156 requests were reiteratively collated and analysed. This process also allowed for the free processing requests to be aggregated and anonymized, in accordance with University of Edinburgh research ethics approval processes, granted on the 31 May 2021.

Content analysis was then used on these indexed emails to provide an 'objective, systematic and quantitative description of manifest content of communication' (Roland and Bawden 2012: 220). Our methodology followed previous studies of archived email content, such as Norhayati's (2014) indexing of the online correspondence of participants at the UN World Summit on the Information Society. This allowed for the exploration of the individual cognitive processes of applicants, related to the characteristics of requests (Riffe, Lacy, and Fico 2014: 8), enabling conclusions to be drawn from content (Stempel, Weaver, and Cleveland Wilhoit 2003: 209). Following Krippendorff (2004: 71), a set of procedures making valid inferences from the free processing requests was formed, presenting replicable results. Additional columns were added, supplying a summary of projects, capturing their contents, using a mix of *in vivo* codes, quoted directly from applicants (Saldana 2009: 10), and process codes (Corbin and Strauss 2015: 283) used to gain a sense of the actions users were performing with Transkribus. A close reading of the sentiment of free processing applications, following Jaillant (2022) was also conducted, deducing whether applicants displayed any uncertainty in the number of credits they requested. With the sample being relatively small, the coding was conducted by the lead researcher. Students, those conducting workshops, and those who mentioned limited funding for their research, were easily identified. ECRs were slightly harder to pinpoint through a simple reading of the free processing requests, with many applicants only offering details about the contents of their research and not information about their position. Using survey methods of scholarship applicants explained in Section 1.5.3, this missing information was established. This presented initial data to direct survey questions, gaining a better insight into the experiences had by scholarship applicants. Such a content analysis of email requests also provided demographic data, concerning the geographical location of applicants, their level of study and occupational position, and their research domains.

Table 2. Template of TSP text fields gathered from READ via Outlook.

University/institution
Programme or discipline
Pursued degree
Credits needed for project
My material is (handwritten/printed/both)
I intend to finish my project on
Please provide a short description of yourself and your project INCLUDING the title of the thesis and why you would like to apply for free processing.[a]

[a] This field only appeared in student proposals.

1.6.2 Content analysis of TUC proposals

Our work also includes a content analysis of conference proposals made to the TUC, held on 28–30 September 2022. We studied sixteen proposals received by READ for the conference's scholarship presentation section, which was set aside for recipients of the Transkribus Scholarship. Applicants wrote an abstract detailing the contents of their research and how Transkribus was utilized. These proposals were given an average score taken from two reviewers formatted as a .csv file. In accordance with GDPR, this file was anonymized before being provided to the lead researcher for content analysis and coding. Reviewers were also anonymized. Findings in this article have been generalized where applicants to the TUC can be identified through research interests. This, like the larger content analysis of free processing requests, was conducted solely by the lead researcher.

These TUC proposals provided essential data on how those who gained funding from READ utilized their free credits. These proposals bolstered the demographic findings reached through analysing email requests and conducting surveys of scholarship applicants.

1.6.3 Survey of Transkribus Scholarship applicants

This content analysis informed a survey of scholarship applicants, hereafter referred to as research survey one (RS1), ascertaining if TSP was meeting their research needs: eliciting responses from those 156 applicants to the scheme, via a mailing list held by READ. The survey was hosted on Jisc Online Surveys,[35] between 1 August 2022 and 1 January 2023. The survey comprised of twenty questions and returned a 21.12 per cent response rate ($n = 34/161$). Questions were largely optional, resulting in a degree of difference in terms of survey engagement. Results provided key demographic data concerning the geographical location of applicants, and their occupations and research domains. The survey also provided more detail surrounding the scale, date, and makeup of collections applicants were researching with support from Transkribus.

1.6.4 Survey of READ staff

Alongside our main survey, a survey of six questions was sent to four members of READ staff, regarding the TSP, hereafter referred to as RS2. Respondents produced a collective response and have given permission for this to be used. This brief survey offered background context to how the TSP operates, the practicalities of facilitating it, and whether the READ-COOP had any plans in changing its structure. In addition, answers from READ staff provided data concerning the aims of the TSP.

The following findings section synthesizes the data streams explained above, answering to what extent the TSP is enabling diverse DH research, making access to ATR technologies more equitable.

2. Findings

2.1 Demographic data

2.1.1 Institutional information

The TSP free processing requests came from a total of ninety-nine institutions (see Fig. 1). Unsurprisingly, applicants predominantly came from universities and research institutes ($n = 91/99$, 91.92 per cent) as the scholarship is aimed specifically towards students and those teaching tutorials. Similarly, 88.23 per cent ($n = 30/34$) of RS1 responses came from universities. Responses from other educational environments included two requests from high schools, national and state libraries, publishers, and even a botanical garden. It appears that the TSP is being primarily used by content holding institutions and university researchers, with the free processing scheme only reaching beyond these groups minimally. This coincides with the aims stated by READ staff, in reply to RS2, of facilitating research, contributing to '… the spread of digital methodologies in the area of historical documents', for the intended groups laid out in READ's initial publicity surrounding the TSP: students, those with limited funding and workshop leaders.[36]

2.1.2 READ-COOP membership

Most scholarship requests came from those not affiliated with READ-COOP member institutions ($n = 146$, 93.59 per cent). Reiterating this, 26.47 per cent ($n = 9/34$) of survey respondents can be seen as already benefitting from reduced processing as part of the READ-COOP. The proportion of READ-COOP representatives rises to 50 per cent ($n = 8/16$) when looking at the 2022 TUC proposals, explained by the nature of the event being tailored to those heavily involved in the platform. TSP recipients across all datasets represented a total of nineteen READ-COOP member organizations, seventeen (89.47 per cent) of which were universities, one (5.26 per cent) national library, and another a research institute. Nonetheless, we remain mindful that transcription work is often conducted in collaboration across universities and cultural heritage organizations, which is not fully reflected in free processing requests. Nonetheless, the TSP is well-used outside of the READ-COOP suggesting that TSP awareness is not reliant on a degree of prior involvement in the Transkribus user community.

READ staff, replying to RS2, addressed the question of READ-COOP-affiliated researchers accessing the TSP. They suggested that such researchers have asked for free processing support '… which we didn't find ideal as such institutions should be able to acquire the desired amount of credits in our opinion, however we did end up granting them the scholarship (as an exception)', citing the aim of the SCE of supporting its members. The Transkribus Scholarship, then, sits between various other structures established by the READ, namely the inclusion of institutions who want to be active members of the COOP. With internal funding streams being harder to access in some cases, the TSP is still necessary for those working within READ-COOP member institutions.

Regardless of whether an institution is a member of the READ-COOP, contributing financially to the organization, unequal distribution of institutional funding means that the resources of an institution do not always trickle down to researchers. We note that a researcher's connection with the READ-COOP does not determine their expertise level in using recognition

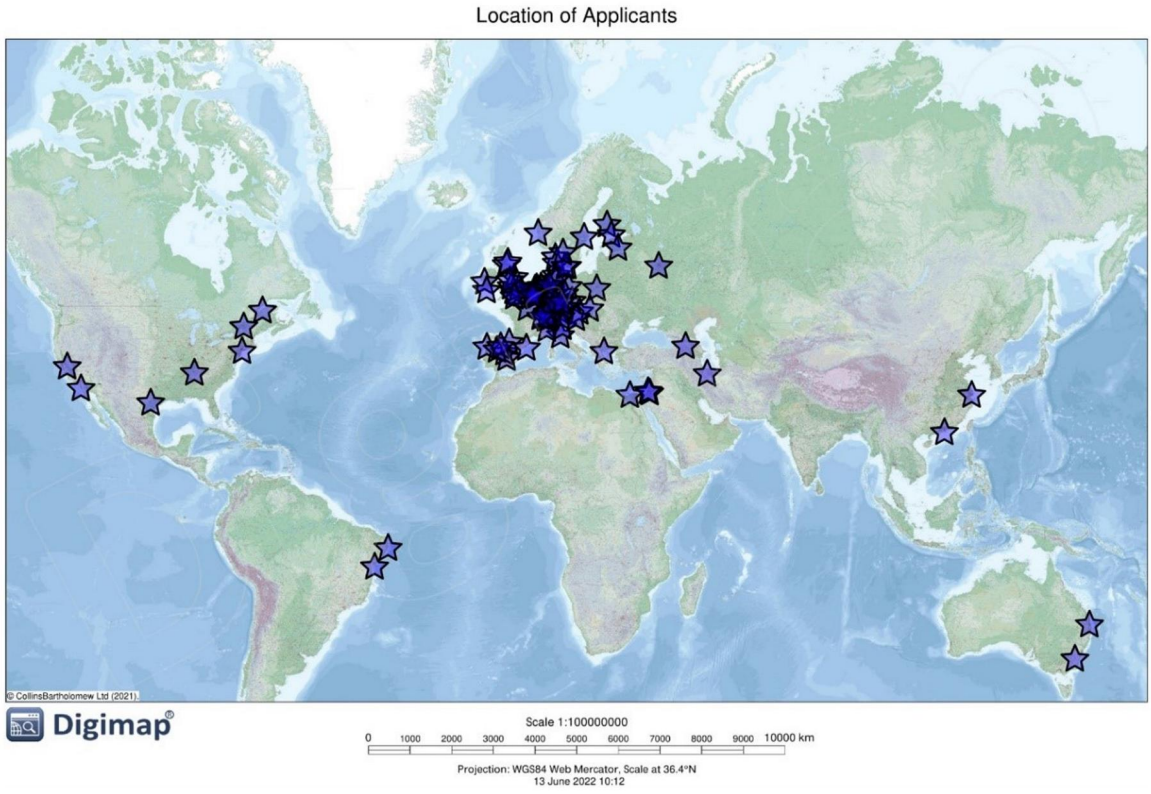

Figure 1. Map showing the location of Transkribus Scholarship applicants, based on their institutional affiliations.

technologies. This is clarified through RS1, with 66.66 per cent ($n = 22/33$) of respondents having previous experience with similar tools to Transkribus. That said, the survey also showed that while most respondents had used text recognition, they had not worked with Transkribus, as a reduced amount ($n = 17/33$, 51.51 per cent) came from institutions making use of the platform already. Where respondents came from institutions using Transkribus, this work had been sustained, with 68.75 per cent survey participants representing organizations having used the ATR tool between two and three years ($n = 11/16$) and 18.75 per cent even longer ($n = 3/16$). From this, it can be ascertained that the majority of those engaging with the Transkribus Scholarship hold a workable knowledge of text recognition, though not always related to Transkribus specifically.

2.1.3 Geography

Looking at the 156 free processing requests, READ received requests from a total of 33 countries. As Fig. 2 shows, Germany ($n = 29$, 18.59 per cent); the Netherlands ($n = 18$, 11.54 per cent), the UK ($n = 16$, 10.26 per cent), and Switzerland ($n = 12$, 7.69 per cent) were the most common. This is replicated in our survey (Fig. 3): Germany ($n = 10/34$, 29.41 per cent), Switzerland ($n = 5/34$, 14.71 per cent) and the UK ($n = 3$, 8.82 per cent). That said, input from those in the Netherlands was comparably lower ($n = 2$, 6.25 per cent). Presented alongside Nockels et al. (2022), where formal published research activity mentioning Transkribus was indexed, informal experimentation, research, and tutorials using Transkribus also gravitate heavily towards Western Europe.

An initial glance at the Eurocentric nature of the dataset may suggest the READ-COOP's focus on enabling common transnational activities and benefiting user needs regardless of geographical location is still in progress. However, with scholarships enabling students from outside Europe to attend European institutions[37] accessing non-Western materials remotely, these figures only present a partial picture of the diversity of ATR researchers and their interests. Among TUC proposals, 12.5 per cent ($n = 2/16$) referenced the use of non-Western materials, despite the researcher being based at a European institution. In another case, the researcher had undertaken a research project on a range of materials working between a European and an Asian institution. Remaining conscious that work beyond Western canons is occurring within Global North institutions is

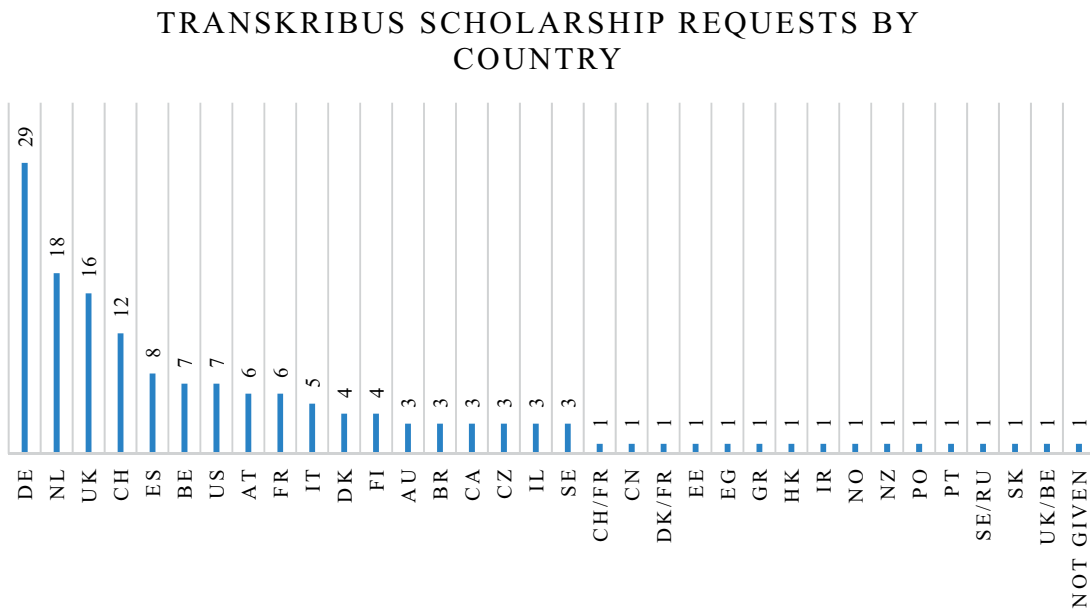

Figure 2. Transkribus Scholarship requests by ISO 3166 Alpha 2 country code.

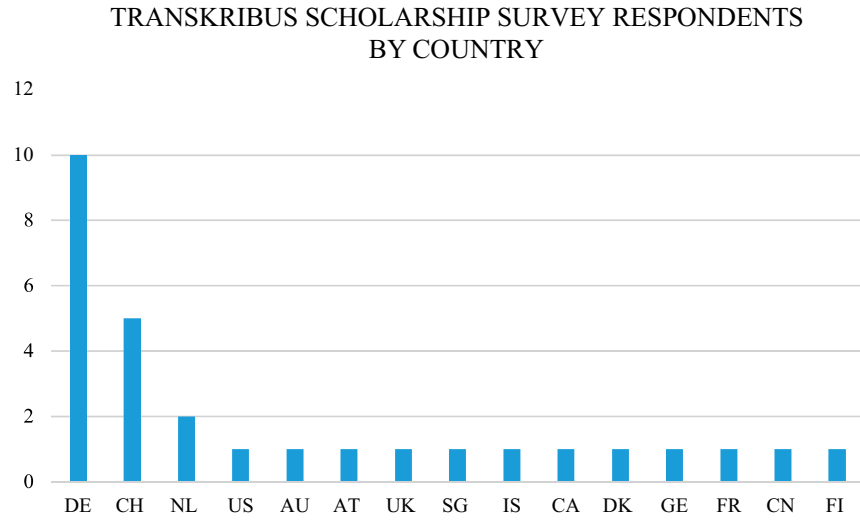

Figure 3. Transkribus Scholarship survey respondents by ISO 3166 Alpha 2 country code.

needed, as Roig-Sanz (2022) maintains in her analysis of Global South DH projects, to fully depict the range of ATR research conducted under the TSP.

Set against Nockels et al. (2022: 372–376), a wider geographical distribution was found from scholarship applications than scholarly research into ATR. Brazil ($n = 3$, 1.92 per cent), Canada, Israel; China ($n = 1$, 0.64 per cent), Egypt, Estonia, Iran, Hong Kong, New Zealand, Portugal, and Slovakia were all present in TSP requests but not found in the published research gathered by Nockels et al. (2022). Looking at regions, South America ($n = 3/156$, 1.92 per cent) and Asia ($n = 2/156$, 1.28 per cent) were both represented among free processing requests (Fig. 4), despite no published work from either being previously identified (Nockels et al. 2022). Requests also came from applicants based in Oceania ($n = 4/156$, 2.48 per cent) the Middle East/West Asia ($n = 4$, 3.21 per cent) and Africa ($n = 1$, 0.64 per cent).

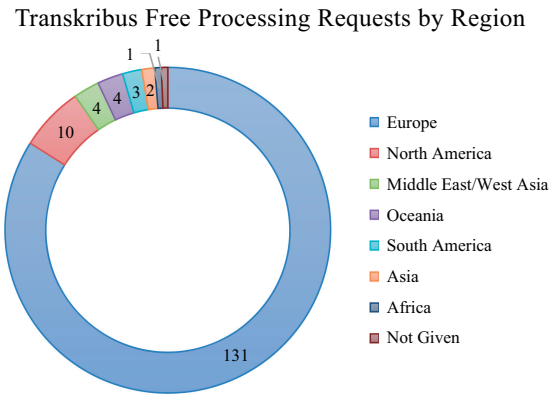

Figure 4. Transkribus Scholarship requests by continental region.

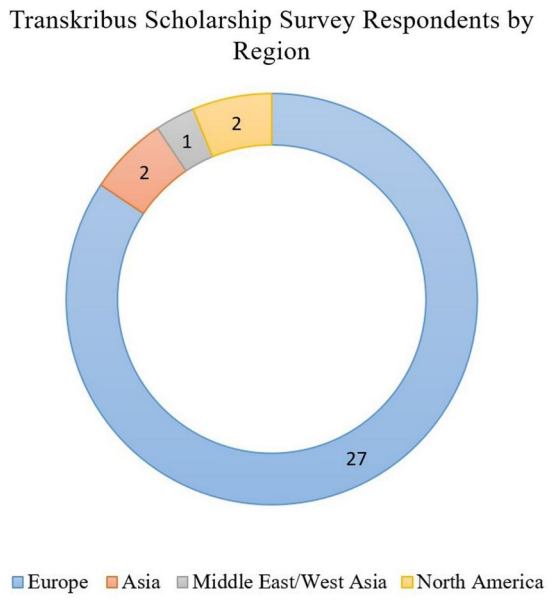

Figure 5. Transkribus Scholarship survey respondents by continental region.

As a result, North America did not dominate free processing requests ($n = 10$, 6.41 per cent). This is also shown in RS1 responses (Fig. 5), with respondents from North America only making up 5.88 per cent of participants ($n = 2/34$). Asia ($n = 2$, 5.88 per cent) and the Middle East/West Asia ($n = 1$, 2.94 per cent) were also represented among survey respondents. Here, China ($n = 1$, 2.94 per cent), Iceland ($n = 1$), and Singapore ($n = 1$) were accounted for but not present in Nockels et al. (2022).

The free processing requests demonstrate a greater linguistical and geographical diversity than we found to be included in the published record (Nockels et al. 2022) with DH research occurring outside the field's main Western hubs. As suggested, critical work on a diverse, non-canonical materials is also being conducted within Western institutions (Roig-Sanz 2022). Therefore, this mapping of requests and survey responses only indicates geographical diversity rather than the nature of the research being done. Requests from European institutions dominate the dataset ($n = 131/156$, 83.97 per cent, see Fig. 4), RS1 responses ($n = 27/32$, 84.38 per cent see Fig. 5), and especially in submitted TUC proposals, which all came from European institutions (although Covid-19 restrictions and growing climate awareness affected international conference travel).[38] Western European nations are more heavily represented than Eastern Europe in both the credit request dataset and survey responses. For the TSP requests, 5.13 per cent came from Eastern Europe ($n = 8/156$). From those responding to the survey of scholarship recipients, only 2.94 per cent came from Eastern Europe ($n = 1/34$).

2.1.4 Research domains

Table 3 shows the full list of domains self-described by scholarship applicants. The main fields represented by these requests were history ($n = 80$, 51.28 per cent) and the DH ($n = 12$, 7.69 per cent). READ staff, replying to RS2, suggested that the free processing scheme was primarily aimed at those using '… digital methodologies in the area of historical documents'.

The range of applicant domains is varied, reaching across disciplinary 'trading zones' between scientific and humanities subjects (Kemman 2021). This breadth of research interests can also be seen in applicants' project summaries, providing a greater level of detail in how they are making use of Transkribus. Those working in history chose to provide greater detail in terms of time periods, with research ranging from the 17th century to the 20th century. A similar range of research interests and domains was observed in the TUC conference proposals. These abstracts included projects on philosophy, transliteration, history, architecture, religion, war studies, and colonial studies. Few scholarship applicants mentioned the types of material they were working on in detail, with many providing only the relevant time period, though some were identified as coming from scholars working on personal archives of individuals ($n = 13$, 8.33 per cent). This suggests that ATR technology can support access to archives that are of direct interest to individual scholars, as opposed to mass digitization projects which have often focused upon materials of more general and mainstream interest. Despite being hard to distinguish, with projects having multiple aims, some applicants placed training a model on a specific language ($n = 5$, 3.21 per cent), font or writing style ($n = 3$, 1.92 per cent), or aspect of document layout ($n = 3$, 1.92 per cent) as the centre of their

Table 3. List of domains reported by applicants to the Transkribus Scholarship.

Domain	Number
History	80
DH	12
Linguistics	7
Economics	4
Archival Science	3
Computer Science	3
Library and Information Science	3
Sociology	3
Slavonic Studies	3
Computational Linguistics	2
Data Science	2
German	2
Hispanic Studies	2
Literature	2
Publishing	2
African Studies	1
Architecture	1
Biology	1
Classical Philology, DH	1
Classics	1
Comparative Religion	1
Forestry Studies	1
French	1
Genealogy	1
German Philology	1
Global Studies	1
Greek Literature	1
Japanese Linguistics	1
Journalism	1
Law	1
Mathematics	1
Medieval Literature	1
Modern Languages	1
Nonfiction Writing	1
Philology	1
Romance Philology	1
Russian	1
Spanish Literature	1
Swedish Language Studies	1
Talmud Studies	1
Theology	1
Translation and Intercultural Studies	1
Not Given	1
Other	1
Grand Total	**156**

project. Emphasizing the range of work being undertaken with Transkribus, requests came from researchers using an array of language materials (Table 3).

2.1.5 Audiences engaged by free processing

The TSP is reaching ECRs, students, and those conducting workshops with limited funding. About 58.82 per cent of survey respondents described themselves as PhD students or postgraduates ($n = 20/34$), 17.65 per cent as ECRs ($n = 6$), and 5.88 per cent as graduate students ($n = 2$). Outliers from respondents when asked about their position included: military officer, digital publications manager, and retired engineer.

About 12.82 per cent of free processing request applicants ($n = 20/156$) provided no employment information. Nonetheless, from the remainder of these requests, the TSP is being taken up by students ($n = 111/156$, 71.15 per cent), ranging from undergraduate level to PhD candidates. ECRs are also represented ($n = 5$, 3.21 per cent). Applicants performing tutorials using Transkribus, at whatever level, also utilize the TSP ($n = 19$, 12.18 per cent). Those attending workshops would also gain free credits as part of the scheme, although would not appear in this study. In teaching-related processing requests, Transkribus appeared to be viewed more as a pedagogical aid than technological tool, with students being asked to transcribe personal materials and connect their heritage to historical events.

Although the Transkribus Scholarship is predominantly being used by ECR groups, a few credit exemption requests came from established academics in their fields. In one case, the free processing request read more as a biography of the academic's credentials, instead of providing information about the project being worked on and how it related to Transkribus. In addition, some RS1 responses came from seasoned researchers, with two participants (5.88 per cent) answering that they were professors. These profiles do not seem to meet that of needing financial aid through the Transkribus Scholarship, although with complexities surrounding accessing resources in higher educational institutions, funding through such free processing schemes could still be needed by established academics.

2.2 Disseminating and publicizing the free processing scheme

This section presents information about how the TSP scheme is currently disseminated and whether any beneficial changes could be made. A large portion of RS1 respondents ($n = 13/34$, 38.24 per cent) heard about Transkribus from their supervisors or colleagues, with others ($n = 6$, 17.64 per cent) hearing generally through word-of-mouth, online ($n = 5$, 14.71 per cent), social media ($n = 4$, 11.76 per cent), or through workshops ($n = 4$) or conferences ($n = 2$, 5.88 per cent). Two participants detailed that they were closely involved with READ, as project members, again highlighting that the TSP is being used by those with an already strong knowledge of the software and part of existing communities. The breakdown of how those granted free processing requests heard about the TSP differs, with 58.62 per cent of survey respondents ($n = 17/29$) citing online publicity, 27.59 per cent ($n = 8$) recommendations from supervisors and colleagues, 6.90 per cent from workshops ($n = 2$) and 3.45 per cent ($n = 1$) from social media and through

word-of-mouth. In both cases, however, existing links with READ were important—with one participant stating that they discussed the scheme with Günter Mühlberger, the Founding Director of Transkribus. In order to broaden publicity surrounding the TSP, one respondent suggests that '… Transkribus should consider advertising to universities directly, rather than users (like Gale or other big academic softwares)'. Whether this approach would increase ATR use, and the diversity of research outputs, is beyond the scope of this piece—although such a response clarifies that more work is needed to bring the Transkribus Scholarship to researchers needing support.

2.3 Frequency of requests

Figure 6 shows the frequency in which Transkribus Scholarship requests were received by READ, plotted by the date of automated emails. The first requests arrived soon after Transkribus's transition to a paid-for model, with ten applicants in November 2020, rising to thirteen in December 2020. The biggest slump in requests can be identified between May 2021 ($n = 17$) and June 2021 ($n = 6$), possibly caused by research and work breaks over the summer.

Requests from June 2021 fall further ($n = 4$), after those relating to conducting workshops, which occur year-round, are taken out. This assertion that fewer requests are received during breaks in the academic calendar could be corroborated via ongoing study. Making more publicizing individual researchers' success via the TSP could also affect the frequency of applications.

No major changes have occurred in the structure of the scheme since its launch in October 2020: the current web environment and credit calculator remain the same. In reply to RS2, READ staff stated that the TSP only takes up one hour of staff time per week, '… although this naturally depends very much on how many requests are coming in' resulting in a small cost in terms of human resourcing which needs to be absorbed by the READ-COOP. Taking inspiration from Terras and Causer's (2014) quantitative study from the Transcribe Bentham project, comparing the cost of setting up crowdsourcing infrastructure and enabling the work of volunteers with the alternative of hiring paid interns, we estimated the cost of supplied free credits compared to the time resources of READ staff, using the Transkribus credit calculator. In total, if all applicants' credit demands were met, providing those who requested more than the maximum amount with 3,000 credits, the TSP provided c. 285,632.25 free credits. Using the Transkribus pricing website this corresponds to 58,841.39 EUR. The TSP thus deploys a sizeable amount of resources, while using little staff time. This picture is, of course, more complex, with resourcing offset by income share-holding READ-COOP members holding shares and other paying users. Therefore, this generosity and support, facilitating diversity in research using ATR, requires necessary planning, especially around financial sustainability. As such, other DH platform providers must lay the groundwork, ensuring regular income, if their efforts to support diversity and inclusion are to be successful and sustainable.

2.4 Requested credit amount

When asked about the number of credits issued, READ staff responded to RS2 that: 'Only on occasion have applicants had to amend requests, requesting more than the limit … however most applicants were

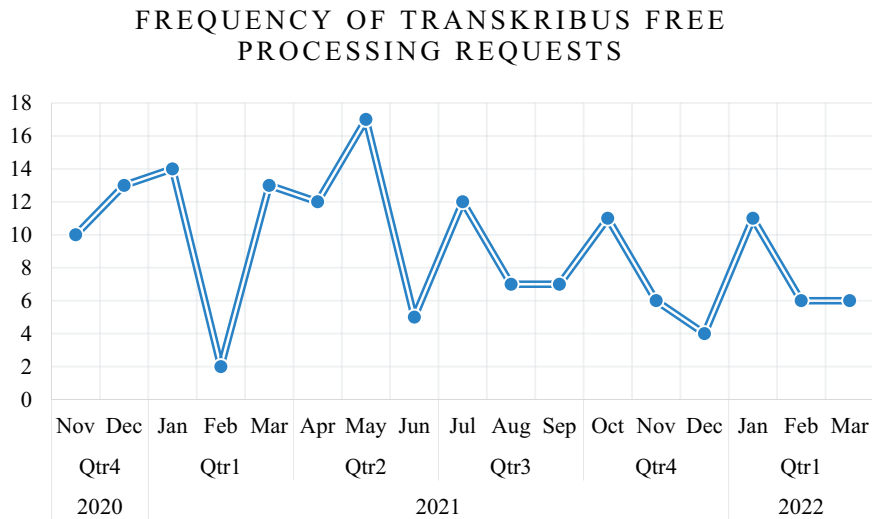

Figure 6. Graph showing the frequency of Transkribus Scholarship requests (November 2020 to March 2022).

content to still take the maximum amount'. With READ being established using public funds, there is an expectation to support researchers; this was shown through the user community's response to the paid-for funding model (Terras 2020, 2022: 197). A tension arises, between the recognition that the READ-COOP must raise enough funds to support and develop Transkribus infrastructure, and the expectation that a tool which was fully funded until late 2020 can and should remain free. This is seen with READ staff giving less focus on research and development post-EU Horizon funding, repositioning resources to ensure that Transkribus continues to be functional and available, while actively looking for other external funding (Terras 2020).

To provide a fuller picture of the resources deployed by the TSP, Fig. 7 breaks down applicants' free processing requests by amount, tied to the cost of model training within the Transkribus environment. The majority of requests were for 0–1,000 credits ($n = 66$, 42.31 per cent), followed by those requesting between 1,001 and 2,000 ($n = 23$, 14.74 per cent). The next largest group requested a significant rise in credits, with fifteen applicants requesting above 10,000 (9.62 per cent). Fourteen applicants (8.97 per cent) requested whatever READ staff determined was needed, with some uncertainty about precise requirements. The spread of credit amounts from survey respondents when asked 'How many credits did you request from Transkribus/READ-COOP?' appeared similar to that of the free processing requests. 27.27 per cent ($n = 9/33$) replied that they requested 1,000 and under, with the next largest groups requesting between 2,001 and 3,000 ($n = 8$, 24.24 per cent) and above 5,000 ($n = 8$, 24.24 per cent). It was expected that those working with handwriting would request larger credit amounts, as this requires additional ATR processing over deploying OCR on printed materials. A mean average credit amount was taken from the free processing requests based on the nature of applicants' material. The outliers mentioned below were taken out of such equations. Taken across the 100 applications using handwriting, an average of 3,843.97 credits were requested per applicant. As expected, the twenty-eight applicants working solely on print requested fewer credits—2,328.57 per applicant. Scholarship requests working both on handwriting and print had a lower average of 4,600.00 credits, although only three projects fit this description.

These amounts sit above the limit set by the READ-COOP, although are likely to have been inflated from the outliers. Some outlier credit requests were identified: one applicant asked for over 50,000 credits, and another asked for 300,000. This can be translated, according to READ-COOP estimations and credit calculator to a project working on at least 240,000 pages (provided they are working on handwritten text).[39] In this case, the requested amount was likely a mistake, as it came from a student working on a small collection of documents. In some cases, such a credit amount is needed by PhD students or ECRs, leading to greater dialogue with READ staff. One RS2 respondent stated that 'depending on what documents you want to transcribe, 3,000 credits (according to them the max) is not nearly enough. If I want to do a keyword search with a high number of, for example, books with hundreds of pages, the credits would stack up really fast. This hinders this broad type of research/finding of sources'. Another respondent stated that they would appreciate having the ability to request more credits

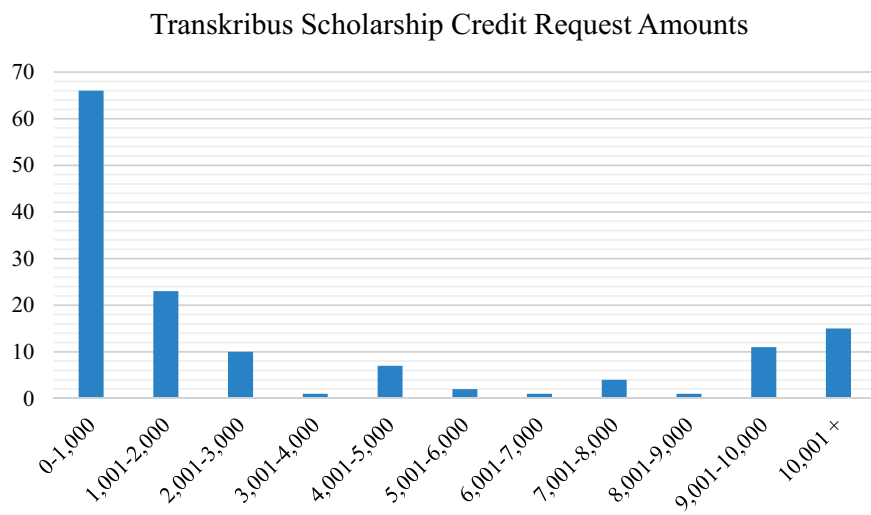

Figure 7. Chart showing the range of credit requests made by applicants to the Transkribus Scholarship.

from the READ-COOP when necessary. Dialogue between READ staff and applicants, beyond filling out the scholarship form, is essential to manage expectations while also providing sufficient support for a wide range of research.

READ staff reported to RS2 that 'most applicants are very informed about Transkribus and have been working with it for some time, so there are not many questions regarding the general use of Transkribus from them …'. Despite this, they noted that many applicants found it difficult to calculate compute requirements: '… uncertainty regarding the amount of credits is quite common, many applicants tend to overestimate how many credits they need for the number of pages they want to process'. Despite the READ-COOP's provision of a Credits Calculator on the Transkribus website,[40] only 33.33 per cent of survey respondents ($n = 11/33$) replied that they used it to calculate their required credits. 60.61 per cent ($n = 20$) estimated based on page number, 15.15 per cent ($n = 5$) asked for help in their application and 3.33 per cent ($n = 1$) asked help from a colleague. In one case, the respondent replied that they '… just thought to get the maximum allowed …' since they didn't know how many errors they would make.

Although the online calculator can be used as an initial indicator of the scale and nature of transcription work, close reading of requests shows that added dialogue between READ staff and applicants regarding credit calculation is an essential support function, with additional advice needed for those new to the ATR platform. Calculating the amount of processing needed in any ATR project regardless of the platform can be tricky, particularly in comparison to OCR projects using fixed licenses based on page limits. Even the approximate number of pages in need of processing is hard to deduce in the funding stages of a project, especially if more material is digitized. RS1 saw 81.82 per cent ($n = 27/33$) respond that the process of calculating credits was easy. However, a minority of two respondents cited errors they had made when applying for credits. '… I initially miscalculated the number of pages I would need to process'; 'I realized after the application that I had not understood that there was a difference between handwritten models and print models. Thus, I had applied for far too many …' Another saw the process as easy but time consuming as their material had random empty pages, resulting in the user having to 'correct layout analysis beforehand and check how many … there were'. In comparison, 11.80 per cent of TSP applicants ($n = 19/161$) did not list any sort of credit amount on their application. Instead, these requests simply stated that the applicant would take whatever amount they needed. Others displayed a level of guesswork: 'uncertain', 'I don't know if 10k credits will be enough', 'about … maximum', 'I'm assuming I won't need more …', 'I really don't know how many credits to ask for as a reasonable number to distribute while giving workshops. Advice would be welcome (or am I supposed to think of a number of credits for my own practise in preparation for the workshop as well … ?)' While processing credits and the submission of applications can easily be automated, providing efficiency in the immediate term, keeping humans in the loop and forming a strong dialogue with users is likely to provide greater benefit in the long run, especially surrounding the clarification of user needs. Future research could quantify how much staff time these interactions account for more broadly.

Despite challenges in calculating credit requirements, most respondents felt that the support of the TSP met their research needs. The majority of RS1 respondents gained enough financial support for their projects ($n = 21/32$, 65.63 per cent), though some also found they had leftover credits ($n = 6$, 18.75 per cent). In those cases, recipients used the balance for personal research projects, demonstrations, and/or work on similar genres of material to, as one survey respondent stated, '… greatly enhance the utility of the work done in the current project'. One respondent's excess credits, initially for a seminar, went to supporting an entire doctoral degree. Although recipients having excess credits could be interpreted as case for the clarification of how to calculate credits, these credits are not wasted but form a social good, extending the research enabled by the Transkribus Scholarship. Some ($n = 5/33$, 15.63 per cent) responded that they did not gain sufficient credits. This could highlight that better communication is still needed between users and READ staff, with applicants designing projects around the READ-COOP's capability to provide free access.

2.5 Project completion

RS1 asked participants whether they could have completed their research without receiving free Transkribus processing, with responses being almost an even split. 51.72 per cent ($n = 15/29$) responded that their work would have been infeasible without free processing, citing the scale of their work, the volume of dirty OCR and need for clean-up of texts, and/or lack of finances. The other 48.28 per cent ($n = 14$) replied that they would have been able to carry out their research. That said, this picture is muddled when reading respondents' explanations, with many stressing that, even if they could afford credits, the scheme meant they did not have to scale back their plans. One survey respondent stated that 'without receiving free credits, I would have had to minimize the text corpus to complete my research, which would minimize the outcome and its reliability'. Another mentioned

undertaking manual transcription efforts in place of using Transkribus, due to the '… substantial financial burden'. Whether survey respondents could afford the cost of processing depended greatly on the nature of their project, with one respondent's project being predominantly concerned with Transkribus' layout analysis (an ongoing free aspect of the tool). Some responses reflected wider structural and attitudinal issues in the academy, surrounding the worth given to students' work. One respondent stated that 'there is no funding available for postgraduate taught dissertations at my institution', a problem that is common across the academic sector. With academic institutions often allocating few resources to support postgraduate research, DH tool providers can offer assistance when possible. Through this non-institutional support, students can benefit from learning advanced computation techniques (although balances must be set, with providers likely unable to provide free support to every research project, so indicating policies need to be developed regarding who to support). There is no easy distinction between those able to afford research costs and those unable without the TSP. Sustained interaction with users, beyond the submission of applications, can serve to provide better detail for platform providers. In the case of Transkribus, READ-COOP members, who already gain discounted processing, can ensure they publicize ATR internally, ensuring students at all levels are able to access platforms.

3. Discussion

Our article forms an initial collection of TSP applicants' real practice. The TSP fulfils its purpose of supporting audiences in need: students, those conducting workshops, and those in need of funding, including ECRs. This is also the view of READ staff in RS2: 'Currently, it works very well'. RS1 respondents agreed, with the vast majority gaining an outcome in the expected time frame ($n = 31/32$, 96.90 per cent). In researching the TSP, this article fulfils the aim, cited by READ staff, of checking 'on the results of the work supported by the scholarships in order to show the community all the interesting projects that were made possible'. User consultation needs to continue, fuelling more equitable access to Transkribus and producing a greater diversity of research. This follows Feliciati's (2022: 388) call for user-centred design, involving stakeholders in the design process and 'answering as closely as possible … their needs and behaviours'. With the TSP ongoing, this work should be further 'included as part of the iterative process of measurement and improvement of service quality' (Feliciati 2022: 392). Through this process, the READ-COOP (and other platform providers) can continually build on previous work, focusing and refocusing final outputs on achieving the expectations and requirements of their target communities. This approach will ensure that a growing body of users are supported and, in turn, aid in the READ-COOP's sustainability.

Some applicants have opened dialogue with READ staff involving technical matters, shown in RS1, '… regarding specific models', as stated by one survey participant. Another survey respondent stated that 'If I have questions, I always write emails'. In some cases, those applying for free processing were also READ-COOP members, re-emphasizing that the Transkribus Scholarship is part of a wider enabling structure, described by one survey participant as providing a 'constant exchange with READ employees about technology/processing etc' However, a large amount of applicants have little to no contact with READ employees, beyond confirmation that they have gained free processing ($n = 13/32$, 40.63 per cent). Respondents want a more fluid method of contact, with one survey respondent calling for the 'ability to request more credits from READ-COOP when necessary'. Sustained communication is also needed to gain a better sense of how to benefit users of Transkribus, for instance showing whether further developments to the programme, such as a better way of feeding back projects' progress to the wider scholarly community, could increase interest in the TSP. As suggested, in terms of the READ-COOP remaining financially sustainable, having a solid dialogue with customers is also essential. Therefore, working with ATR requires new approaches both to historical material but also to public engagement.

4. Recommendations

This section provides a set of recommendations for DH and DS platform providers on how to best develop free processing schemes to increase research access, in turn diversifying the work undertaken using said platforms. We first recommend that platforms develop their own policies and procedures for accessing free use of platforms which provide suitable balance between revenue generation and support for a diverse cohort. This may mean focusing on supporting particular user groups, or those from particular geographic areas, or economic backgrounds. Thought should be put into these frameworks, and they should be published openly and transparently, and advertised directly to prospective recipients. We also suggest that platforms should publish high-level overviews of who is successfully accessing free schemes, to support transparency of these programmes.

Our finding that fewer requests for free support were received during breaks in the academic calendar

can advise providers on the timing of their scheme's promotion, ensuring that their intended audiences are reached. This recommended approach would also focus resources on specific parts of the year, striking a balance between providing free access and servicing paid-for customers.

We recommend that free processing is provided alongside robust mechanisms for reporting project success, such as the Transkribus 'success stories' blog,[41] which can facilitate publicity of the platform and its free-access scheme. This supports some of the suggestions seen in the 2020 TUC, notably the inclusion of more inclusive communication channels, such as a messaging board within the Transkribus environment or on the main website (Terras 2020). Areas for scholarship recipients to discuss and present work, outside formal conferences, will aid the publicity of schemes and awareness of the support available.

In light of the current TSP operational structure only accounting for an hour of READ staff time a week, as stated in RS2, coupling together automated processes with human interaction, free processing schemes can also be seen as benefiting tool providers themselves. The TSP enables workshops, training users in Transkribus, and bolsters the reputation of the tool. As such, tool providers can benefit from recognizing the worth of being altruistic to their research communities.

Lastly, free processing schemes need to be coupled with training and skills development more generally. In reply to our survey, one respondent suggested that 'It might be useful to develop a training course for recipients of the scholarships … there was a big learning curve for the software'. In the case of Transkribus, how-to manuals are available[42] and workshops occur online, but these could be tailored for those applying to the TSP, providing firmer grounding before applicants use the ATR platform. Although being wary of increased resourcing, training could have an impact on scheme success, with applicants having a better sense of the process. Increasing such dialogue is crucial to ensure that user needs are met, sufficiently facilitating diversity in DH work.

5. Conclusion

In applying a lens to the mediation of access to digital cultural heritage materials, this article focused on how work in DH can be limited through a lack of funding and what can be done by tool providers to open up access. We examined the TSP, a free processing initiative established by READ in order to provide broad access to its ATR tool, Transkribus. This scheme sits alongside other structures established by READ that ensure revenue, suggesting that platform provides must balance the financial sustainability of their tools while bearing in mind user needs and the promotion of diverse research. READ's free processing scheme appears to be robust in supporting diverse research across trading zones, shown in the range of applicants' research domains. Our analysis of the demographics of those who applied for free processing shows that the scheme is being used by the intended groups, laid out in READ's publicity for the TSP: students, those conducting workshops and those who mentioned having limited funding for their research. ECRs were harder to pinpoint through a simple reading of the free processing applications, although our survey provided insights into their experiences. There is a clear need to develop more understanding around the calculation of credits, with the online calculator being a useful reference alongside human interaction with READ staff, accounting for the nature of ATR projects and better quantify project costing. With some more clarification, this part of the process can facilitate a broader diversity of research using and into ATR.

Following the recommendations in this article, DH tool providers should balance future planning in the case of financial sustainability with providing generous support, through free processing and access, to defined user groups. We encourage platform providers to transparently engage in these free access frameworks, monitoring them and relaying their successes. We also encourage providers to provide enhanced training and support in conjunction with free access schemes, to encourage a diverse range of applicants.

Collectively, this will allow for the building of more equitable DH infrastructure, coinciding with and, in many cases, enabling the broadening of research in the sector.

Acknowledgements

We would like to acknowledge the contributions of Andy Stauder, Managing Director of READ, and Günter Mühlberger, Founding Director of READ, throughout the research process, especially in making data available. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

Joseph Nockels (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, and Writing—original draft), Paul Gooding (Conceptualization and Writing—review & editing), and Melissa Terras (Conceptualization and Writing—review & editing).

Funding

Melissa Terras serves on the Board of Directors of Transkribus as Research Director. Joe Nockels' doctoral research on Transkribus is funded by the Scottish Graduate School in the Arts and Humanities, Arts and Humanities Research Council (AH/R012717/1), in partnership with Transkribus and the National Library of Scotland. This research has had full ethical approval via the READ-COOP and University of Edinburgh's disclosure processes.

Conflict of interest statement. Terras serves on the Board of Directors of Transkribus as Research Director. Transkribus is one of Nockels's industrial PhD partners. This research has ethical approval via the READ-COOP, the National Library of Scotland, and the University of Edinburgh's processes.

Notes

1. READ commonly refer to Transkribus as a Handwritten Text Recognition (HTR) tool. However, with its efficacy also proven over printed materials, we use the term ATR. See: READ-COOP, 'Print Multi-Language', accessed 4 January 2024. https://readcoop.eu/model/transkribus-print-multi-language-dutch-german-english-finnish-french-swedish-etc/
2. Terras's 2012 infographic showcases 114 physical DH centres in 24 countries (44 in the USA and 14 in the UK). See: https://blogs.ucl.ac.uk/dh/2012/01/20/infographic-quantifying-digital-humanities/
3. The author is aware of contentions in using the moniker Global South, due to it being coined in/from the Global North, disregarding, in some cases, the complexities of the region (Fontolan, 2020: 190). However, due to its common usage in the literature, we use the term to ensure consistency. A similar consideration is given, in this article, to the label Middle East, which denotes a decidedly European perspective of geography. The more inclusive definition of West Asia is used alongside this label (Payind and McClimans 2017: 18).
4. Transkribus enables users, through a Java-based client, to upload images; utilize automated processes like model training and search against returned transcriptions. Users can also collaborate on documents with varying degrees of access. READ-COOP, 'Transkribus', accessed 5 January 2021. https://readcoop.eu/transkribus/
5. READ-COOP, 'Supporting Future Scholars: The Transkribus Scholarship Programme', accessed 13 April 2023. https://readcoop.eu/supporting-future-scholars-the-transkribus-scholarship-programme/
6. Ibid.
7. Defined here, in line with the UKRI (2020), as those who are engaging in post-doctoral research and/or transitioning toward becoming independent academics.
8. READ-COOP, 'Supporting Future Scholars: The Transkribus Scholarship Programme', accessed 13 April 2023. https://readcoop.eu/supporting-future-scholars-the-transkribus-scholarship-programme/
9. See Nockels et al. (2022) for a systematic review of published research mentioning Transkribus.
10. Ground truth (GT) data are defined by Schofield (2014: 3) as part of the ATR training process, with ground truth being those where the layout of the document, line detection and content has been ascertained, 'the greater the quantity of ground truth that is used for training, the greater the accuracy of the machine-read text'. READ describe GT as the accurate and verified data which is used to train and test machine learning models, such as those used for automatic transcriptions. READ-COOP, 'What is Ground Truth?', accessed 11 May 2023. https://readcoop.eu/what-is-ground-truth/
11. Kraken, accessed 20 January 2023. https://kraken.re/main/index.html
12. eScriptorium, accessed 3 May 2023. https://www.escriptorium.uk/
13. 'KNAW—HUC/Loghi', accessed 2 April 2023. https://github.com/knaw-huc/loghi
14. READ-COOP, 'Transkribus Scholarship Programme', accessed 5 January 2021. https://readcoop.eu/transkribus/scholarship/
15. European Commission, 'tranScriptorium', accessed 4 January 2024. https://cordis.europa.eu/project/id/600707
16. READ-COOP, 'Members of the READ-COOP SCE', accessed 3 May 2023. https://readcoop.eu/members/
17. For institutional members of the READ-COOP 4 shares must be bought, costing 250 EUR each, for individuals or natural persons, 1 share is required for minimum participation. READ-COOP, 'Article 6—The Shares', 'Statutes of the READ-COOP SCE with limited liability', accessed 10 April 2023. https://readcoop.eu/wp-content/uploads/2020/05/Statutes_READ_COOP_SCE_current.pdf
18. Data collection occurred before the introduction of READ's tailored subscription model in July 2023. The newest plan grants 100 pages worth of free processing per month and includes a subscription plans based on users selecting an individual, scholar or organization account. On-demand purchasing is still available and the organization plan remains in beta stage. READ-COOP (2023a, 2023b, 2023c), 'Coming Soon: Exciting Changes to the Transkribus Subscription Plans!', accessed 11 August 2023. https://readcoop.eu/new-subscription-model/#:~:text=Introducing%20Tailored%20Subscription%20Plans,of%20diverse%20groups%20of%20users.
19. The READ-COOP now maintains offices separate from the University of Innsbruck, Austria, where it was previously based and supported financially for the first year.
20. Results from READ-COOP survey, accessed 31 May 2021. https://readcoop.eu/understanding-institutional-use-of-htr-your-experience-with-transkribus/
21. Transkribus was included as a finalist for the 2020 Horizon Impact Award, acknowledging 'EU-funded projects whose results have created societal impact across Europe and beyond'. European Commission, 'Horizon Impact Award 2020–10 finalists short-listed', accessed 5 April 2023. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/horizon-impact-award-2020-10-finalists-short- listed-2020-09-15_en
22. Southern Connecticut State University, 'Tools for Digital Humanities', accessed 1 May 2023. https://libguides.southernct.edu/digitalhumanities/tools
23. Massachusetts Institute of Technology Libraries, 'Digital Humanities: Tools and Resources Recommendations', accessed 1 May 2023. https://libguides.mit.edu/c.php?g=176357&p=1158575
24. Open Source Initiative, 'Open Source Definition', accessed 1 May 2023. https://opensource.org/osd/
25. Voyant Tools, accessed 6 May 2022. https://voyant-tools.org/ Antconc, accessed 4 May 2023. https://www.laurenceanthony.net/software/antconc/ Word Seer, 4 accessed May 2023. http://wordseer.berkeley.edu/announcing-wordseer-4-0/
26. MALLET (MAchine Learning for Language EToolkit), accessed 2 May 2023. https://mimno.github.io/Mallet/about
27. Tropy, accessed 2 May 2023. https://tropy.org/ Prospect, accessed 4 May 2023. https://prospect.unc.edu/faq/
28. Gephi, accessed 2 May 2023. https://gephi.org/
29. QGIS, accessed 4 May 2023. https://www.qgis.org/en/site/
30. Two forms exist, both considered here, one for students and another for teachers. READ-COOP, 'Transkribus Scholarship Programme', accessed 5 January 2021. https://readcoop.eu/transkribus/scholarship/students, https://readcoop.eu/transkribus/scholarship/teacher/
31. READ-COOP, 'Transkribus Lite Credits', accessed 5 April 2023. https://app.transkribus.eu/credits
32. These amounts were calculated using Transkribus' 'Credit Calculator'. This feature is discussed as length later in this article. READ-COOP, accessed 5 April, 2023. https://readcoop.eu/transkribus/credits/
33. READ-COOP, 'Transkribus Success Stories', accessed 11 May 2023. https://readcoop.eu/success-stories/

34. This anonymised dataset has been made publicly available through Zenodo. See: https://zenodo.org/records/10477942
35. Jisc is a GDPR compliant questionnaire platform licensed by the University of Edinburgh. Jisc Online Surveys was chosen as it allows surveys to be filled under pseudonyms, offers space for privacy information, and enables the full anonymizing of responses: as such, no personal information was stored. Jisc Online Surveys, accessed 1 June 2021. https://www.onlinesurveys.ac.uk/
36. READ-COOP, Transkribus Scholarship Programme, accessed 5 January 2023. https://readcoop.eu/transkribus/scholarship/
37. The Commonwealth Scholarship Fellowship Plan (CSFP) enables citizens from Commonwealth countries to study in the UK. Though financially aiding more precarious students' UK studies, whether such schemes border the potential re-inscription of colonial protocols forms part of a larger conversation about the role of the UK Commonwealth. British Council India, 'Commonwealth Scholarships', accessed 6 January 2024. https://www.britishcouncil.in/study-uk/scholarships/commonwealth-scholarships
38. See tweet from ATR researcher @CARomein, July 13, 2022, referring to declining invitations overseas, accessed 6 July 2023. https://twitter.com/CARomein/status/1547262238695751680?cxt=HHwWgICw_d2z_fgqAAAA.
39. Although credits can be purchased in whatever combination—to be used on printed or handwritten material, if working on printed material, this requested credit amount would enable work on one and a half million pages of text. 'READ-COOP Credits & Pricing', accessed 4 May 2022. https://readcoop.eu/transkribus/credits/
40. READ-COOP, 'Transkribus Credits Calculator', accessed 3 May 2023a. https://readcoop.eu/transkribus/credits/
41. READ-COOP, 'Success Stories', accessed 17 April 2023. https://readcoop.eu/success-stories/
42. READ-COOP, 'Transkribus How-To', accessed 17 April 2023. https://readcoop.eu/transkribus/howto/use-transkribus-in-10-steps/

References

Antconc. (2023) https://www.laurenceanthony.net/software/antconc/, accessed 4 May 2023.

Bordalejo, B., and Murray, P.R. (2016) 'If You Think You Know the Answer, You Don't Understand the Question. Boundary Land: Diversity as a Defining Feature of the Digital Humanities', *Digital Humanities 2016: Conference Abstracts*, pp. 80–1. Kraków: Jagiellonian University & Pedagogical University.

Boyles, C. *et al.* (2018) 'Precarious Labor and the Digital Humanities', *American Quarterly*, **70**: 693–700.

British Council India. (2024) 'Commonwealth Scholarships', https://www.britishcouncil.in/study-uk/scholarships/commonwealth-scholarships, accessed 6 Jan. 2024.

Casanova, R., and Sarmiento, A. (2021) 'Digital Humanities in the Philippines: Projects, Challenges and Opportunities of North-South Collaboration', *Digital Scholarship in the Humanities*, **36**: 55–67.

ChatGPT. (2023) 'Introducing ChatGPT Plus', https://openai.com/blog/chatgpt-plus, accessed 5 May 2023.

Corbin, J., and Strauss, A. (2015) *Basics of Qualitative Research*. London: SAGE.

Cummings, R., Roh, D.S., and Carraway, E. (2020) 'Organic and Locally Sourced: Growing a Digital Humanities Lab with an Eye Toward Sustainability', *Digital Humanities Quarterly*, **14**: 1–35.

Deng, Y., Lambrecht, A., and Liu, Y. (2021) 'Spillover Effects and Freemium Strategy in the Mobile App Market', *Management Science*, **69**: 1–53.

Dewidar, O., Elmestekawy, N., and Welch, V. (2022) 'Improving Equity, Diversity, and Inclusion in Academia', *Research Integrity and Peer Review*, **7**: 1–10.

Drucker, J. (2021) 'Sustainability and Complexity: Knowledge and Authority in the Digital Humanities', *Digital Scholarship in the Humanities*, **36**: 86–94.

eScriptorium. (2023). https://www.escriptorium.uk/, accessed 5 Jan. 2021.

European Commission. (2020) 'Horizon Impact Award 2020–10 Finalists Short-listed', https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/horizon-impact-award-2020-10-finalists-short-listed-2020-09-15_en, accessed 5 Apr. 2023.

Feliciati, P. (2022) 'Planning with the Public: How to Co-develop Digital Public History Projects', in S. Noiret, M. Tebeau and G. Zaagsma (eds.) *Handbook of Digital Public History*, pp. 385–95. Berlin, Germany: De Gruyter.

Fenlon, K.S. (2020) 'Sustaining Digital Humanities Collections: Challenges and Community-Centred Strategies', *International Journal of Digital Curation*, **15**: 1–13.

Fontolan, M. (2020) 'Seminar Review: I Seminário Internacional Remoto—Experiences in Digital Humanities from the Global South', *Tapuya: Latin American Science, Technology and Society*, **3**: 190–1.

From the Page. (2023) 'Pricing', https://fromthepage.com/pricing, accessed 5 May 2023.

Galina, I. (2013) 'Is there anybody out there? Building a global digital humanities community', *Humanides Digitales*, **21**: 307–316.

Gephi. (2023) https://gephi.org/, accessed 2 May 2023.

Gil, A., Ortega, É. (2013) 'Global outlooks in digital humanities: multilingual practices and minimal computing', in Crompton C., Lane R., Siemens, R. (eds). *Doing Digital Humanities: Practice, Training, Research*, pp. 1–16. London: Routledge.

Gu, X., Kannan, P.K., and Ma, L. (2018) 'Selling the Premium in Freemium', *Journal of Marketing*, **82**: 10–27.

Hawkins, S. (2021) 'Introduction', in S. Hawkins (ed.) *Access and Control in Digital Humanities*, pp. 1–20. London: Routledge.

Jaillant, L. (2022) 'Introduction', in L. Jaillant (ed.) *Archives, Access and Artificial Intelligence—Working with Born Digital and Digitized Archival Collections*, pp. 1–20. Bielefeld: Transcript Publishing.

Jenkins, B.M., and Myers, T.K. (2022) 'Digital Black Voices: Podcasting and the Black Public Sphere', *Digital Humanities Quarterly*, **16**: 1–16.

Jisc Online Surveys. (2021) 'Homepage,' https://www.onlinesurveys.ac.uk/, accessed 1 Jun. 2021.

Jisc Online Surveys. (2023) 'Features—Pricing', https://www.onlinesurveys.ac.uk/features-pricing/, accessed 2 May 2023.

Kemman, M. (2021) *Trading Zones of Digital History*. Berlin, Germany: De Gruyter.

Kraken. (2023) https://kraken.re/main/index.html, accessed 20 Jan. 2023.

Krippendorff, K. (2004) *Content Analysis: An Introduction to its Methodology*. London: Sage.

Liakos, F. (2019) 'Digitization Initiatives for Preservation & Access: The Transkribus Project As a Stepping Stone for a New Archival Era', Master dissertation, Leiden University, Leiden, Netherlands.

Madsen, C., and Hurst, M. (2018) 'Are digital humanities projects sustainable? A proposed service model for a DH infrastructure', Paper presented at the *Coalition of Networked*

Information Fall Membership Meeting (CNI18f), pp. 1–63. Washington, DC.

MALLET (MAchine Learning for Language EToolkit). (2023) https://mimno.github.io/Mallet/about, accessed 2 May 2023.

Massachusetts Institute of Technology Libraries. (2023) 'Digital Humanities: Tools and Resources Recommendations', https://libguides.mit.edu/c.php?g=176357&p=1158575, accessed 1 May 2023.

Mendeley. (2023) 'Pricing', https://www.mendeley.com/settings/billing/v2/upgrade/, accessed 2 May 2023.

MONK (no date) 'Monk Homepage', https://www.ai.rug.nl/lambert/Monk-collections-english.html, accessed 5 Feb. 2021.

Mühlberger, G. *et al.* (2019) 'Transforming Scholarship in the Archives through Handwriting Text Recognition, Transkribus as a Case Study', *Emerald Publishing*, **75**: 965–7.

Nockels, J. *et al.* (2022)' Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts: A Systematic Review of Transkribus in Published Research', *Archival Science*, **22**: 1–26.

Norhayati, Z. (2014) 'Content Analysis of Email Archival Messages: Why Use Software and How to Train Coders', *SAGE Research Cases*, **1**: 1–10.

Omeka. (2023) 'Services', https://omeka.org/services/, accessed 5 May 2023.

Omeka.net. (2024) 'Pricing,' https://www.omeka.net/signup, accessed 4 Jan. 2024.

Open Source Initiative. (2023) 'Open Source Definition', https://opensource.org/osd/, accessed 1 May 2023.

Owens, T., and Padilla, T. (2021) 'Digital Sources and Digital Archives: Historical Evidence in the Digital Age', *International Journal of Digital Humanities*, **1**: 325–41.

Oxygen XML. (2023) 'Buy Oxygen XML Editor', https://omeka.org/services/, accessed 5 May 2023.

Payind, A., and McClimans, M. (2017) *Keys to Understanding the Middle East*. Columbus, OH: Ohio State University.

Pons, A. (2020) 'The Historiographical Foundations of Digital Public History', in S. Noiret, M. Tebeau, and G. Zaagsma (eds.) *Handbook of Digital Public History*, pp. 15–30. Berlin, Germany: De Gruyter.

Posner, M. (2016) 'Money and Time', [Blog Post], https://miriamposner.com/blog/money-and-time/, accessed 20 Mar. 2023.

Prospect. (2023) https://prospect.unc.edu/faq/, accessed 4 May 2023.

QGIS. (2023) https://www.qgis.org/en, accessed 4 May 2023.

Murray, P.R. (2020) 'Decolonising Design: Making Critically in India', in M. Maya Dodd, and N. Kalra (eds.) *Exploring Digital Humanities in India Pedagogies, Practices, and Institutional Possibilities*, pp. 124–37. Routledge: Abingdon.

READ-COOP. (2020a) 'Transkribus Credits & Pricing', https://readcoop.eu/transkribus/credits/, accessed 16 Jul. 2021.

READ-COOP. (2020b) Statutes of the READ-COOP SCE with Limited Liability', https://readcoop.eu/wp-content/uploads/2020/05/Statutes_READ_COOP_SCE_current.pdf, accessed 10 Apr. 2023.

READ-COOP. (2021) 'Members of READ-COOP SCE', https://readcoop.eu/members/, accessed 20 Jan. 2021.

READ-COOP. (2022a) 'Become a Member and Co-Owner of READ-COOP', https://readcoop.eu/join/, accessed 2 May 2022.

READ-COOP. (2022b) 'Transkribus Lite Credit Manager', https://transkribus.eu/lite/credits, accessed 9 May 2022.

READ-COOP. (2022c) 'Transkribus Scholarship Programme Portal for Students and Teachers', https://readcoop.eu/transkribus/scholarship/#:~:text=The%20Transkribus%20Scholarship%20Programme%20is,from%20any%20programme%20or%20discipline, accessed 18 Apr. 2022.

READ-COOP. (2022d) 'Transkribus Success Stories', https://readcoop.eu/success-stories/, accessed 20 Apr. 2022.

READ-COOP. (2023a) 'Print Multi-Language', https://readcoop.eu/model/transkribus-print-multi-language-dutch-german-english-finnish-french-swedish-etc/, accessed 4 Jan. 2024.

READ-COOP. (2023b) 'What is Ground Truth', https://readcoop.eu/what-is-ground-truth/, accessed 11 May 2023.

READ-COOP. (2023c) 'Coming Soon: Exciting Changes to the Transkribus Subscription Plans', https://readcoop.eu/new-subscription-model/#:~:text=Introducing%20Tailored%20Subscription%20Plans,of%20diverse%20groups%20of%20users, accessed 11 Aug. 2023.

Riffe, D., Lacy, S., and Fico, F. (2014) *Analysing Media Messages, Using Quantitative Analysis in Research*. London: Routledge.

Risam, R. (2015) 'South Asian Digital Humanities: An Overview', *South Asian Review*, **36**: 161–175.

Risam, R. (2019) *New Digital Worlds—Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Evanston, IL: Northwestern University Press.

Roig-Sanz, D. (2018) 'Social Networks of the Past. Mapping Hispanic and Lusophone Literary Modernity', YouTube [Online Video], https://mapmodern.wordpress.com/mapping-modernity/, accessed 14 Apr. 2023.

Roig-Sanz, D. (2022) 'Towards a Truly Global Digital Humanities', https://trafo.hypotheses.org/36011, accessed 14 Apr. 2022.

Roland, L., and Bawden, D. (2012) 'The Future of History: Investigating the Preservation of Information in the Digital Age', *Library & Information Science*, **28**: 220–36.

Ross, N. (2018) 'Customer Retention in Freemium Applications', *Journal of Marketing Analytics*, **6**: 127–37.

Saldana, J. (2009) *The Coding Manual for Qualitative Researchers*. London: SAGE.

Schofield, P. (2014) 'tranScriptorium: A Guide for the Humanist', UCL, https://discovery.ucl.ac.uk/id/eprint/1432610/, accessed 4 May 2023.

Schomaker, L. (2020) 'Lifelong Learning for Text Retrieval and Recognition in Historical Handwritten Document Collections', in A. Fischer, M. Liwicki and R. Ingold (eds.) *Handwritten Historical Document Analysis, Recognition and Retrieval—State of the Art and Future Trends*, pp. 221–48. London: World Scientific.

Southern Connecticut State University. (2023) 'Tools for Digital Humanities', https://libguides.southernct.edu/digitalhumanities/tools, accessed 1 May 2023.

Stauder, F. (2024). *Personal Communication*. READ-COOP Members. Innsbruck, Austria: READ-COOP. 10 January, 2024, 04:34:00 GMT.

Stempel, G.H., Weaver, D.H., and Cleveland Wilhoit, G. (2003) *Mass Communication Research and Theory*. Boston, MA: Addison-Wesley Longman.

Strauß, T., Weidemann, M., and Labahn, R. (2017). *D7.11 Language Models—Improving Transcriptions by External Language Resource*. Innsbruck: Recognition and Enrichment of Archival Documents (READ).

Survey Monkey. (2023) 'Plans and Pricing', https://www.surveymonkey.co.uk/pricing, accessed 2 May 2023.

Terras, M., and Causer, T. (2014) 'Many Hands Make Light Work. Many Hands Together Make Merry Work': Transcribe Bentham and Crowdsourcing Manuscript Collections', in M. Ridge (ed.) *Crowdsourcing Our Cultural Heritage*, pp. 57–88. Farnham: Ashgate Publishing.

Terras, M. (2020) 'What Users Think about Transkribus—Transkribus User Conference', YouTube [Online Video], https://www.youtube.com/watch?v=tyRqww0BdMI, accessed 12 Apr. 2023.

Terras, M. (2022) 'Inviting AI into the Archives: The Reception of Handwritten Recognition Technology into Historical Manuscript Transcription', in L. Lise Jaillant (ed.) *Archives, Access and AI Working with Born-Digital and Digitised Archival Collections*, pp. 1–26. Berlin, Germany: Transcript Verlag,

Tropy. (2023) https://tropy.org/, accessed 2 May 2023.

UK Research & Innovation (UKRI). (2020) 'Early Career Researcher', https://www.ukri.org/what-we-offer/developing-people-and-skills/esrc/early-career-researchers, accessed 9 May 2021.

Universitat Hamburg. (2022) 'Betamasaheft Project', https://www.betamasaheft.uni-hamburg.de/, accessed 5 Jan. 2022.

Van Zundert, J. (2012) 'If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities', *Historical Social Research*, **37**: 165–86.

Voyant Tools. (2022) 'About—Voyant Tools Help', https://voyant-tools.org/docs/#!/guide/about, accessed 6 May 2022.

Walcher, J. (2020) 'Personal Communication', *[TRANSKRIBUS]—Nominated for two Awards—Newsletter*. 21 September, 2020, 14:05:00 GMT. Innsbruck, Austria: READ-COOP.

Word Seer. (2023) http://wordseer.berkeley.edu/announcing-wordseer-4-0/, accessed 4 May 2023.

Zotero. (2023) 'Storage', https://www.zotero.org/storage, accessed 2 May 2023.

Digital Scholarship in the Humanities, 2025, 40, 46–65
https://doi.org/10.1093/llc/fqae018
Full Paper