

On Automating Editions

The Affordances of Handwritten Text Recognition Platforms for Scholarly Editing

Melissa Terras, University of Edinburgh

DOI: [10.55520/W257A74E](https://doi.org/10.55520/W257A74E)

Joe Nockels, University of Edinburgh

Sarah Ames, National Library of Scotland

Paul Gooding, University of Glasgow

Andy Stauder, READ-COOP SCE

Günter Mühlberger, University of Innsbruck

Introduction

How can scholarly editors best make use of recent developments in Handwritten Text Recognition (HTR), where the products of automated text recognition can appear as completely automated online editions? What tools and features should be embedded into these automated editions for them to best serve the scholarly community? How will the answers to these questions impact the labor and production pipelines of scholarly editing? This paper responds to an absence within scholarly editing literature regarding text recognition technologies and the platforms that are now utilizing them. HTR is a machine-learning approach that creates accurate transcriptions of images of handwritten documents. We highlight current developments, demonstrating that automated standardized editions generated using the inputs (images) and outputs (accurate transcriptions) of HTR are both a reality that textual scholars should consider and a technology that needs to reflect the practices inherent within scholarly editing to realize their full potential.

We advance this argument by engaging with a case study of a standardized online edition of the manuscripts of Marjorie Fleming (1803–1811), a Scottish child author who became posthumously famous for her precocious diaries, now in the National Library of Scotland. The HTR platform Transkribus (<https://transkribus.eu>) is one of many available approaches to generate near-accurate machine-processable transcripts at scale. In addition, a recently launched feature of Transkribus can automatically create standardized digital online editions from previously uploaded images of texts and their HTR-generated transcripts. We use images of Fleming's diary uploaded to Transkribus servers, and their HTR outputs, to generate an automated standardized online edition, hosted by Transkribus's read&search functionality: <https://transkribus.eu/r/marjory-fleming/>. This approach bypasses previous online edition procedures and disrupts and democratizes established processes for data formatting, hosting, and delivery of digital editions while providing a modest yet reliable tool for browsing, reading, and searching well-described documents. However, the technical features provided so far do not engage with or facilitate many standard approaches and labor practices inherent to scholarly editing, including textual examination, textual alignment, the creation of a critical apparatus, or the ability to move between multiple textual sources in a multi-witness edition. There is much opportunity for automated tools to facilitate the entire scholarly editing process, including the creator and user requirements of digital editions, and to rethink the future role of machine learning and its related technologies in this space.

In this paper we consider how standardized, near-automatic editions can best align with scholarly editing activities, such as the need for editors to engage with machine-learning, and technological providers to engage with scholarly editors, to symbiotically develop tools that can introduce efficiencies while still respecting and including the characteristics, features, and labor that will move automated digital editions further toward the expected functionality of scholarly critical digital editions. We see the potential for automated editions to democratize the practice of digital edition creation, making it easier to publish editions of underrepresented voices in the scholarly canon. However, it is necessary to develop best-practice frameworks to ensure that the products of such an HTR pipeline are both useful and respected. In particular, the tools need to be expanded to facilitate scholarly textual editing. There is much to learn from the scholarly editing community in how best to deliver such platforms. Furthermore, the deployment of automated text recognition-based editions should be engaged in a way that allows future scholarly editors to develop the necessary skill sets to work alongside and with automated HTR processes, enjoying efficiencies and using available labor to increase the scholarly quality and information available within digital editions, rather than being replaced by them. Those who digitize content at scale, those who maintain HTR-based edition infrastructure, and those who undertake scholarly editing all need to scrutinize which material can be served by this technology, to ensure that traditionally marginalized voices can have the chance to benefit from it. We demonstrate that it is important for the scholarly editing community to understand the potential of automated editions to contribute to the modern knowledge environment, as well as the need for technology providers to engage with and support established academic practices in scholarly editing, learning from established community expertise as these technologies mature.

HTR: A Current Snapshot

The automatic generation of useful machine-readable transcriptions of digital images of handwritten material has long been an ideal of both researchers and memory institutions, and it is understood that successful Handwritten Text Recognition (HTR) will transform access to mass-digitized manuscript content.¹ Accurate HTR-generated transcriptions will allow the searching of vast manuscript repositories and enable the readability of many documents whose hands are impenetrable to modern audiences, extending the scale of encounters with primary sources and the voices represented within digitized cultural heritage,² while also allowing textual information to be mined, visualized, and analyzed using various advanced digital humanities techniques,³ including providing machine-processable text for use within scholarly editions.

HTR and its allied technique, OCR (Optical Character Recognition, the means by which images of printed text can be transformed into machine-processable format), both have long histories.⁴ In recent years, OCR has integrated machine-learning techniques into its processes, improving the quality of generated texts, although it still struggles with complex layouts, fonts, or media,⁵ and its accuracy with handwriting is dependent on characters being spatially separated. Recently, HTR techniques that use deep neural networks to extract visual features and recognize characters and words in a segmented line of text via the calculation of overlapping probabilities have become more stable, accurate, and efficient.⁶ However, this complexity demands an increase in computational power. Since the late 2010s, HTR has become more accurate and has been used by both libraries and archives as well as scholarly projects across the heritage sector. HTR can be implemented in a range of ways. Some projects choose to use in-house bespoke computational approaches. Some projects engage with the facilities of major publishers, many of whom are beginning to offer machine-learning approaches to data analysis. Another approach is to use Transkribus (<https://transkribus.eu>), currently the primary user-facing platform for applying HTR to digitized content, which is the outcome of years of European Union–funded research activity.⁷ The functionality and infrastructure of Transkribus is fully documented elsewhere,⁸ but it employs machine-learning processes where users can train HTR upon a subset of material before deploying an improved model across a range of digitized content (while also improving outputs for future users). Best-case results from Transkribus generate a character error rate of below 5% on handwritten material and below 1% on print material. HTR is increasing the speed and efficiency at which transcriptions can be generated from historical content,⁹ and therefore providing more machine-processable text upon which to build further digital scholarly work, including digital editions. Recent developments in HTR have also opened up the possibility for the creation of *automated* editions, where the products of the HTR pipeline are immediately transformed into a public-facing online resource, allowing transcription and images to be viewed line by line. Technological development enabled accurate line segmentation of handwritten documents, meaning that manual alignment of images and their text is no longer necessary, saving effort while also allowing users to more easily evaluate the automated text generated.¹⁰

HTR and Scholarly Editing

Scholarly editing literature is quiet on the topic of HTR and how it may impact or change digital editorial practices, possibly because HTR was only recently successfully deployed, and because of the long production processes of academic publishing.¹¹ However, coverage of the transformational aspects of these technologies flourishes both in digital humanities literature and in library and information science literature more broadly.¹² *Scholarly Editing: A Guide to Research*,¹³ published in 1995, makes no mention of any text recognition technologies (even though offline OCR software was available by the mid-1990s: ABBYY was founded in 1989 and had produced a popular product by 1993).¹⁴ Hockey's 2000 *Electronic Texts in the Humanities* was the first scholarly editing text to mention OCR, although Hockey focuses on its limitations and recommends keyboarding for data entry: "Handwriting is impossible because of the joined letters and in inconsistency in the letter shapes."¹⁵ *A Guide to Documentary Editing* was published in 2008 before the recent increase in accuracy in HTR, although its authors do cautiously mention the possibility of using OCR as the basis upon which to build a scholarly edition: "the transcription must be viewed as a very rough one, demanding serious verification and proofreading by senior staff."¹⁶ The use of OCR had reached the mass market by then—Tesseract had released its code openly in 2005, with Google hosting and developing this open-source cloud-based system since 2006,¹⁷ and Adobe Acrobat had begun to include support for OCR using ClearScan in 2008.¹⁸ There is mention of OCR used as the basis for digital editions in 2016's *Digital Scholarly Editing: Theories and Practices* but not in a positive light: "editions may be the result of uncorrected ('dirty') OCR taken from old, out of copyright editions, and may therefore bear little resemblance to their originals."¹⁹ Despite this warning, since the early 2010s the text recognition community has said that "the OCR of modern clean documents is effectively a solved problem" although "older degraded documents present difficulties."²⁰ HTR is mentioned only in passing in *Digital Scholarly Editing: Theories and Practices* as research that "has not yet produced reliably working products, but much more is to be expected in the coming years."²¹ However, there was an implication that digital editions that depend on any such automated processes are not "proper,"²² which may speak to issues of scholarly ownership of the entire editing process, given that the mechanisms of many OCR and HTR technologies remain black-box.

The 2016 *MLA Statement on the Scholarly Edition in the Digital Age* does not include any reference to the possibilities, appropriateness, or usefulness of either OCR or HTR to assist the scholarly editing workflow. However, this broadly framed comment is apposite:

[A digital scholarly edition] must note its technological choices and be aware of their implications, ideally using technologies appropriate to the goals of the edition, in recognition of the fact that technologies and methods are interrelated in that no technical decisions are innocent of methodological implications and vice versa.²³

In practice, the scholarly editing community already utilizes text recognition. TEI-compliant XML (which underpins most online scholarly editions) is already one of the formats that can be generated by HTR engines, including Transkribus,²⁴ and HTR-generated XML can be used as input text for digital editions, even though discussions and best-practice guidance remain at an early stage.²⁵ It is evident, from a review of papers and book chapters about particular projects, that outputs of OCR and HTR are already actively used as efficient means of generating machine-processable texts for creating scholarly editions,²⁶ even if those commenting on scholarly editing practices overall have not fully documented this shift. In addition, the implications of using HTR, not only as the textual source for digital editions but also as an infrastructure for building automated editions, is then worthy of consideration to understand the abilities, possibilities, and limitations of this environment when used in knowledge creation.²⁷

It should be noted that there are many ways in which the digital edition publishing process has already embraced a level of automation to enjoy efficiencies, particularly in the transformation of encoding. For example, the outputs of ABBYY XML, produced by OCR, were transformed into simple TEI encoding in the IslandLives project.²⁸ Accurate encoding of punctuation from more advanced automated processes was explored in the Early New High German Marco Polo version of the *Devisement dou Monde*.²⁹ Digital humanists routinely use some level of transformation in digital encoding when creating a digital edition, via bespoke scripts or available tools. Aspects of automation of publishing of digital editions use digital encoding as well. Since 2015, TEI Publisher has assisted in the publication of digital editions by transforming XML text into different outputs, displaying them as web pages, and providing tools for search and navigation.³⁰ Scholarly editing literature has not fully considered our focus on fully automated online editions—where the outputs of machine-learning HTR are automatically ported into a public-facing digital edition, with no human intervention. We argue that the community should be involved in driving these innovations, not just merely receiving them.

The Fleming Diaries: A Read&Search Edition

We present here a case study of an automatic online edition of the manuscripts of Marjorie Fleming (1803–1811), a Scottish child author who, after her tragic death from measles followed by meningitis, became posthumously famous for her precocious diaries. Written mainly in Edinburgh and the holiday retreats of Ravelston and Braehead, they provide an authoritative record of Fleming’s life, schoolwork, and passing. These diary pages are filled with witticisms, moralisms, and copybook notes, all completed under the watchful gaze of Fleming’s tutor and older cousin, Isabella Keith (1787–?).³¹ The collection also contains a few letters between Keith and Isabella Rae, Fleming’s mother, about Marjory’s death. The attention given to these diaries in the later Victorian period saw her feted by men of letters,³² including Robert Louis Stevenson, Mark Twain, and Leslie Stephen.³³ Fleming’s distant relation to Walter Scott became exaggerated in the process, with John Brown suggesting that Scott believed her to be “the most extraordinary creature” he ever met.³⁴ As a result, Fleming’s diaries became embedded in the public consciousness, and (especially given the paucity of female children’s writing) are a key source for those seeking a situational history of early nineteenth-century Scotland.

Three volumes of Marjory’s diary (1810–1811) were presented to the National Library of Scotland in 1930. They were digitized in-house, in the National Library of Scotland studios, in 2014. The digitization was internally funded. The diary was chosen specifically because of its out-of-copyright status, its clear and uncomplicated hand, and the potential for it to be suitable for crowdsourced transcription. In 2021 the Fleming diary transcripts were the first dataset created in the National Library using HTR with the Transkribus platform (<https://www.transkribus.org/>).³⁵ The dataset itself (both text and images) is published under a public domain open license on the National Library of Scotland’s Data Foundry (Fleming, 2021), providing the opportunity for others to reuse them for further analysis as “collections as data.”³⁶ The uploaded images and HTR outputs were used to generate an automated standardized online edition, hosted by Transkribus’ read&search functionality (<https://transkribus.eu/r/marjory-fleming/>). The read&search tool was launched in beta testing by Transkribus in late 2020 (<https://web.archive.org/web/20211203140223/https://readcoop.eu/readsearch/>). It has undergone development since, automating the process of turning HTR-generated text into a functional digital edition, aligning previously uploaded digital images with their HTR-generated transcripts. In 2024, the name of the service was changed to Transkribus Sites (<https://readcoop.eu/transkribus-sites-publishing-documents-made-easy/>).

While the costs for creating a read&search edition were not made public, given that the business model for this was still evolving (costs also depend on the volume of material being transformed and hosted, and so each edition was costed independently),³⁷ these costs have settled into the Transkribus Sites Subscription Plans (https://www.transkribus.org/plans/sites?_hstc=175940285.8830bf10881107c36409b6192d80640b.1716130761238.1717164014927.1717248170726.4&_hssc=175940285.1.1717248170726&_). The infrastructure has both initial setup costs and ongoing hosting and support costs. The Transkribus Sites pricing model now offers different levels of service, depending on whether the sites are Small (up to 10 thousand pages), Medium (up to 100 thousand pages) or Large (up to 500 thousand pages). There are, of course, sustainability considerations when outsourcing the hosting of editions, although creators of these editions can export all underlying assets, to back them up or to host the content elsewhere.

At the time of writing, there are now seventeen different presentational editions hosted, for example, the Hanse Quellen,³⁸ an overview of scientific texts written in Latin during the modern era from the Noscemus project,³⁹ and Franciscan Cadastre documents from Tyrol (1865),⁴⁰ both projects indicating that read&search is scalable to mass-digitized content. While the read&search tool and its features were developed by the READ-COOP team in Innsbruck, this testing of and reflection on the Fleming edition (one of the smaller editions in read&search) constitutes part of a UK Arts and Humanities Research Council–funded Collaborative Doctoral Award PhD, via the Scottish Graduate School for Arts and Humanities (<https://www.sgsah.ac.uk/>), working with both Transkribus and the National Library of Scotland. We present the functionality at time of writing of read&search by examining the Fleming diaries in more detail, before considering what we have learned by engaging in the automated edition process.

Constructing the Fleming Read&Search Edition

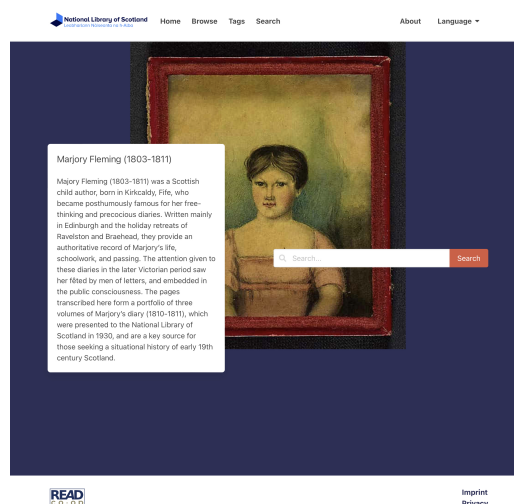


Figure 1: The opening page of the Fleming digital online edition, via read&search: <https://transkribus.eu/r/marjory-fleming/>.

The inputs for the Fleming online edition were the digitized images of the manuscripts and their corresponding HTR-generated transcripts. First, the JPEG files of the Fleming diary, which were previously digitized at the National Library of Scotland, were ingested into Transkribus using its upload function (the documents are already ordered into a meaningful collection, uploaded onto a web service that provides the basis for the online edition, thus saving future effort). To run HTR, the automatic layout analysis tool was used, recognizing the main body of handwritten text on each page as a separate region and segmenting it, allowing computational analysis of each document's structure and identifying transcription lines that corresponded to those within the manuscript's image. Certain identities could be tagged as transcription began, such as mentions of Fleming's tutor Isabella or the place-name Kirkcaldy and flagging of abbreviations (which are then expanded and shown to the reader), alongside structural elements of the text such as strikethroughs and superscript. Normalization of tags is an important step to be accomplished by an editor. For example, Isabella is known in the diaries as Isa, Issa, Isabella, Isabella Keith, and Isa Keith. The resolving of these entities shows that to make the most out of the process, only semiautomation is possible, indicating where human labor and checking are best deployed. These tagged entities are then later searchable in read&search, corresponding to the page numbers of the document, and (as with all aspects of the HTR) are dynamically updated in the presentation edition when changed and saved in Transkribus itself, being baked into its XML outputs that underpin the read&search functionality. It is worth considering this act of tagging at the point of transcription, in light of digital scholarly editing's approach to markup. The markup necessary for the creation of digital scholarly editions is normally undertaken post-transcription or post-creation of the electronic text, not during, because "The first requirement is to prepare the text for processing, which means, in many workflows, to transcribe."⁴¹ Instead, Transkribus embeds these tags inside the encoding from the point of creation, during the act of transcription, producing an enriched transcription. While the embedding of tags within "cultural heritage texts has been criticized in the belief that if one embeds markup and interpretation into a text, then this text cannot be reused by other people, affecting therefore the possibility of exchanging and interchanging encoded files,"⁴² the fact that Transkribus's editor-generated tags can be exported into TEI-compliant XML (allowing the user to choose which zones and type of line breaks to include) can assist further encoding. Within HTR and the read&search workflow, the timing of when editorial interpretation of the text occurs is therefore at an earlier stage than in the creation of many digital editions. It would be worthy of further exegesis to consider how and if the early intervention and symbiosis between scholar and assistive technology affects the scholarly editing process, and how HTR-enriched transcription sits alongside encoding and markup as normally framed within the scholarly editing literature, given that this does require editorial analysis and input.

The Fleming diaries are a smaller dataset than normally used to train HTR; 15,000 words of text are generally recommended as a training set, before applying to a full run of manuscript material, but the whole of the Fleming dataset is only 2,115 lines and 12,167 words, in Fleming's neat, consistent hand. We did train a model (using 11,944 words, leaving a small set of words (223) to be able to validate the character error rate (CER, the percentage of characters that the model fails to recognize correctly). This produced a CER of only 1.85% on the training set and 11.26% on the validation set. It is possible to attempt HTR using an existing model (and many editors may have chosen to do so rather than spend the time necessary to create a model for this small set), although training a bespoke model leads to better outcomes and less human-led correction once the HTR has been completed, and models can also be made public and shared for others to use. Transkribus English Handwriting M2, which is based on more than 1.2 million words aggregated from various projects and models in Transkribus, achieved 14.1% CER out-of-the-box, when used on the Fleming dataset, and was therefore utilized as a base model due to the limited number of pages in the Fleming manuscripts.

We describe read&search editions as standardized, as currently the default user interface for different sets of images and their transcripts is the same, although features may be optional and configurable in future iterations of the tool. The functionality of the read&search tool is such that it directly uses the data (segmented .jpg images and XML transcripts) from the Transkribus platform to automatically produce a presentational online digital edition, via CSS and JavaScript (using Vue⁴³ and Buefy⁴⁴ for user interface elements and Node.js⁴⁵ to access the Solr⁴⁶ server for search functionality). Images are presented from an IIIF⁴⁷ server and can therefore also be accessed by other applications. The main feature of the site is that the digital image of text is placed alongside transcription (a feature that assists in understanding documents, particularly when historical hands are unknown to the reader). Importantly, this process is not static, but dynamic: if the outputs of HTR are corrected or edited, the resulting online presentation is immediately updated (an improvement on many static digital editions, which can be time-consuming and difficult to update when errors are spotted or corrections needed). A search function operates across the complete manuscript (driven by Solr full-text search).⁴⁸ In its description, each edition can link to external sources, such as the institution that created it. The platform is mobile-friendly (according to the Google Mobile-Friendly test),⁴⁹ and the browser produces a suitable printable version of the content, if necessary.

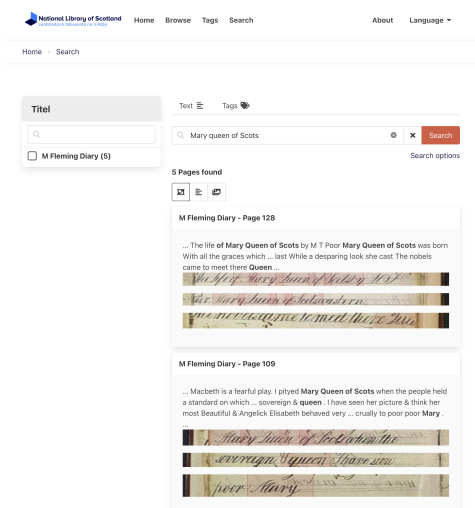


Figure 2: A search for “Mary Queen of Scots” in the Fleming read&search edition. Five pages contain this reference, with results returned via both transcript and image. Tagged entities can also be searched. There is an option to change the default precision for search terms, allowing or declining fuzzy matching; the tool offers exact-match, medium-precision, or high-precision search. <https://transkribus.eu/t/marjory-fleming/?#/search?view=combined&f=1&t=Mary%20Queen%20of%20Scots&p=1>.

The view for each individual page in read&search aligns the digital image with the transcript, line by line; the transcription and the image are linked so that clicking on a word in the image brings up the corresponding word in the transcription and vice versa.⁵⁰ Structural elements that were marked up during the training and transcription process remain and can be used by the users to navigate the text. In the case of the Fleming edition, we realized we needed to return to these structural elements and increase them, to make the most of the navigational elements of the digital edition, dynamically updating the edition as we corrected. Each page’s XML transcript and digital image are downloadable (although this is an optional feature, due to any potential copyright or licensing issues).

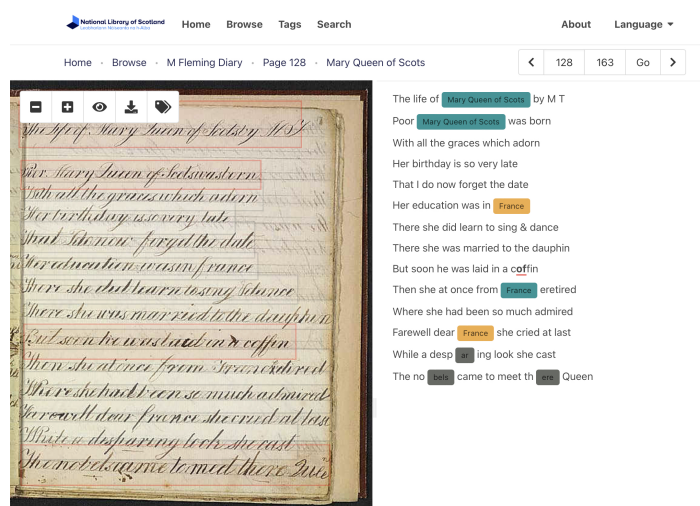


Figure 3: Page 128 from the Fleming read&search edition, showing the opening of Fleming’s poem regarding Mary Queen of Scots, and highlighting tagged entities. <https://transkribus.eu/t/marjory-fleming/#/documents/475423/pages/128?t=Mary%20Queen%20of%20Scots>.

What is immediately produced by read&search, then, is a modest, functional online presentational *digital edition*, created directly from previous HTR inputs and outputs with minimal effort, using a commercial platform that acts as an interface and search engine to a manuscript. This exponentially speeds up the process and ease of generating an online presentational edition of a manuscript, disrupting prior models, costs, labor, and time frames of transcription, markup, correction, hosting, and display, and devolving issues of sustainability to the commercial platform used to deliver and host (for a fee). The read&search tool was launched as a paid service by the READ-COOP in late 2020. (It should be stressed that users do not have to be members of the READ-COOP to access read&search; only a Transkribus user account is needed, in addition to being able to cover costs for hosting, although READ-COOP membership does make users eligible for a discount on all Transkribus services.) There is, then, a cost barrier to accessing these tools, admittedly a financial burden to those with little institutional or grant funding. These tools may then potentially privilege richer institutions, with the costs stymying the potential to open up the creation of digital editions upon and for a broader range of texts and audiences. However, the read&search approach does follow the golden “open access” publishing concept of payment in advance to publish, for free access to a wide readership, replicating the financial models now used to underpin much openly published scholarship (although these also reduce affordability).

With regard to Transkribus, since it moved from a “free at the point of use” to a credit-based paid for system in 2020,⁵¹ active mechanisms now exist to provide free access to its platforms for students and those being trained upon the system,⁵² and its cooperative business model endeavors to work with, rather than extract from, the scholarly community,⁵³ with income supporting the ongoing functionality of the platform. The READ-COOP will continue to monitor who can—and cannot

—access this functionality, to avoid the replication of the historical canon in the digitized environment,⁵⁴ hoping to encourage and support a plurality of voices in the digital editions hosted;⁵⁵ other technology providers should also endeavor to do the same.

In the case of the Fleming edition, we have demonstrated the ease of the process in generating an online edition for a voice that is not featured elsewhere in the digital edition landscape, given that few extant historical texts capture the experience of (female, Scottish) children in their own words. This process, then, demonstrates that automating at least some of the transcription, formatting, upload, and hosting process of building digital editions may allow new historical, linguistic, and literary focuses to be explored by a wider group of readers who may have an interest in creating or consulting them, if they can meet the costs for HTR processing with the likes of Transkribus and also hosting the edition with a tool like read&search.

However, editors can offer little intervention beyond the creation and checking of the HTR transcript itself (including enriching the transcription with tags) and resolving identification or disambiguation of individual entities in the marked-up text. This complication suggests that scholarly tasks normally associated with editors, such as producing a critical apparatus—expanding abbreviations and glossing unfamiliar words, for example—are not yet possible. In the future, commenting features will allow further exegesis of the text (see below) but for now, this automated process precludes many of the editorial tasks commonly associated with online digital editions. We highlight the following needs: for the editing and scholarly communities to critique these types of automated processes, for features to be expanded to reflect the user needs of the scholarly community, for these automated, enriched editions to be used as the vehicle for *critical digital editions*, and for the scholarly community to benefit from their efficiencies.

Read&Search as Digital Edition

At time of writing, the read&search tool was under development (in January 2024 it was renamed as Transkribus Sites).⁵⁶ Automated editions produced by read&search meet the requirement to be considered as a digital textual edition “broadly defined as a reconstruction of a literary text inserted in the text’s print or manuscript tradition and which makes use of digital technologies to reproduce any level of detail of that text.”⁵⁷ Indeed, the read&search tool contains much functionality that is desirable for users of digital editions:

At the most basic level of interest, users are looking for simple browsing functionalities. To satisfy these users, editors will want to present the materials within an attractive and intuitive interface. At a more advanced level of interest, users will want to research the materials the DSE [Digital Scholarly Edition] has to offer, and access them in non-linear ways. To reach those users, editors will need to provide indexes, advanced search options, advanced textual comparison options, to open the corpus up for analysis in a standardized format, etc. Finally, at the highest level of interest, there are metausers, who want to re-use the DSE’s data for their own purposes: to write their own transcriptions of the DSE’s facsimiles (and publish the results), to build their own interface around the data the DSE provides, or to perform functionalities the DSE does not (yet) offer (and publish the results).⁵⁸

There is no universally accepted definition of *digital scholarly edition*, although much research has been published on the topic.⁵⁹ The functionality of read&search meets Sahlé’s definition of a *scholarly edition* if we consider the HTR training and markup/enrichment process as an act of critical examination, especially given this process is not fully automated but depends on human intervention.

A scholarly edition is the critical representation of historical documents, defined as:

1. “historical documents”: editing is concerned with documents that already exist. In this wide sense of “historical” the definition includes documents relevant for all subjects, history as well as literature or philosophy. Scholarly editing goes back to and starts from existing documents. To edit (to publish) a new document (which does not refer to something preexisting) is not scholarly editing.
2. “representation”: covers (abstract) representation as well as presentation (reproduction). As I used to say: transmedialization (representation by data) and medialization (presentation by media). Publishing descriptive data (e.g., metadata) without reproduction is not critical editing. A catalogue, a database, or a calendar is not an edition.
3. “critical/scholarly” (erschließend): reproduction of documents without critical examination is not scholarly editing. A facsimile is not a scholarly edition.⁶⁰

It is in point 3 where we must carefully consider how the act of the human editor using HTR within the preparation and pipeline of an automated digital edition expands its criticality. It has long been argued that markup embedded into digital scholarly editing is a critical act, creating “a regime of representation that is self-consistent and self-validating.”⁶¹ If this is the case, does the critical and careful act of applying HTR in a consistent manner through training and correction, and the resolving and disambiguation of entities within the document via embedding and enriching markup, then using the inputs and outputs to automatically generate a digital edition, not provide a pipeline for a digital scholarly edition, much simplifying the labor and process of production, and allowing it to open up for a broader constituency? In this respect, we take inspiration from authors such as Rockwell and Sinclair who argue that digital work, or building, in the humanities is itself a form of high-level scholarly inquiry.⁶²

That being said, read&search does not present a full range of functionalities that could be described as advanced scholarly edition features. Particularly, it does not yet provide a space for commentary or apparatus, allow textual comparisons, or generate indexes (although these functions will soon be available). It is the current lack of the apparatus (and the hosting of intellectual editorial thought behind an apparatus, allowing editors to actually *edit*) that stops the read&search tool from being defined as a *critical edition*, given “no classical text can be read responsibly without one.”⁶³ Feelings on this point run strong, particularly in *Digital Scholarly Editing: Theories and Practices*:

[W]ithout the apparatus the reader cannot have any real idea what he or she is actually reading. One could—and people regularly do—argue that the availability of these mutilated texts is better than nothing, but in many ways these texts are actually worse than nothing, since they are

misleading and fuel the idea that texts exist outside the dialectic between documents and editors, and that editions can possibly establish texts once and for all, undermining in this way the very survival of textual scholarship itself.⁶⁴

HTR-generated digital editions of manuscripts are no longer mere “mutilated texts,” given their increased potential for access and the support in generating accurate transcriptions (from potentially complex sources) that underpins the editions; given the advances in the technology, generated transcriptions are no longer “a very rough one.”⁶⁵ Many of these transcriptions are also enriched with tagging that is embedded into the underlying XML, as discussed above, extending their presentational value, although the transcriptions do not yet support many of the functionalities users would expect from a digital critical edition, such as multi-witness comparison and a critical apparatus.

It is better to consider the intellectual and technical roles of the scholarly editor here, and what they can bring *in addition* to automatable aspects of digital edition creation, such as critical thinking regarding textual and digitization selection and the bounds and scale of an edition, including:

- the selection (and omission) of texts;
- examining and aligning texts and their specific aspects;
- addressing issues of bias and representation and ensuring diverse voices are included;
- training and improving an HTR model;
- providing critical annotations, glosses, and apparatus; designing multi-witness editions;
- markup of textual features; cross-referencing and linking to other related digital editions, and information sources;
- user testing, and incorporating user feedback about content as well as navigation;
- quality control throughout the process;
- feature development;
- promotion of the digital editions in apposite venues to reach potential users;
- utilizing the resulting digital edition to generate new and meaningful insights into manuscripts, authors, histories, and cultures;
- and sharing those insights to show the worth of digital editions overall.

Authorial opinion in the literature previously cited is determined not to be replaceable by the machine even for mere transcription generation. However, the usefulness and accuracy of HTR, and the tools that can now be built upon it, necessarily propel us beyond that.

Improving Read&Search Functionality

It is worth considering the necessary feature development that would expand the read&search functionality, allaying some fears and increasing the criticality of its editions. Moreover, we should also reflect upon the ways in which automated editions change labor practices in the construction of digital editions, and impact sustainability and peer review.

In 2016–2017, in “A Catalogue of Digital Editions,” Franzini, Terras, and Mahony identified forty-nine features commonly used in scholarly editions.⁶⁶ Using this framework to analyze the read&search tool, we can see that it can be improved in a range of ways to meet scholarly edition users’ needs. Firstly, the epistemological value of these automated editions could be enhanced by adding standardized descriptive informational fields, including:

- making clear where the source material is kept, and its provenance and origin;
- declaring the language and temporality of the digital edition;
- including a philological statement about editorial practices, an account of textual variance, and the value of witnesses;
- providing an epistemic statement about technical aspects and practices used to make the edition;
- offering statements about relationships and linking to other projects or data;
- supplying contact information for users to make comments or suggestions;
- and providing license information about images and edition overall.

In regard to digital editions, this material offers metadata that will enhance the read&search tool’s scholarly value.

Secondly, using Franzini, Terras, and Mahony’s categories, a range of feature developments could expand the functionality of read&search as its currently stands. The most important of these is the necessity of providing the means to host a critical apparatus, including glossary, comments, and/or notes. This undertaking will allow read&search to host *critical* editions. This feature is planned for the near future and will enable a mechanism for editors to engage with the substance and meaning of the text in the digital edition. However, we are planning additional features to improve the tool, depending on the commercial uptake of the system. They include:

- expanding the download feature to export XML-TEI;
- where necessary, providing a mechanism for hosting translation of source material;
- offering a feedback mechanism for users to make comments or corrections;
- hosting a tool that allows both visual and textual comparison between different manuscripts; and
- applying DOIs or PIDs to each individual edition to allow them to be citable and findable.

Further mechanisms (not mentioned in Franzini, Terras, and Mahony’s categorization) being considered for future read&search features include:

- flexibility in hosting metadata for any scheme standard (TEI, MODS, etc.),⁶⁷
- the ability to highlight key sections and provide reading excerpts,
- the visualisation of structure tags (for marginalia, etc.),
- the automated visualization and mapping of location tags,
- the automated visualization of person tags as networks,
- and linking to WikiData and using WikiData for dynamically enriching the edition using semantic technologies.

The presentational digital edition we show here is therefore the initial phase of integrated, automated online digital editions. As the editions hosted on read&search multiply, we should consider the varied ways in which end users can navigate and find different editions, such as through faceted search.

Likewise, the RIDE Criteria for Reviewing Scholarly Digital Editions can also supply a template by which to evaluate read&search implementation.⁶⁸ Issues not covered above in Franzini, Terras, and Mahony’s framework but mentioned in the RIDE criteria are sustainability and the user focus, including benchmarks of scholarly quality. Sustainability for read&search editions is currently the responsibility of the READ-COOP. Depending on a platform provider to ensure the longevity of a digital resource may alarm some scholarly editors, because of concerns about outsourcing the hosting of any online resource (even if underlying assets can be exported for future use elsewhere).⁶⁹ It may be useful for the developers of the HTR platforms hosting such editions to consider digital preservation aspects—for example, those set out in the Endings Project’s “Endings Principles for Digital Longevity”—and to produce an Ending Principles–compliant export format.⁷⁰

The RIDE review criteria prioritize the user viewpoint; they propose that we evaluate the relative trustworthiness of editions and that we assess their scholarly quality (in an act that is itself one of peer review). It should be noted that read&search is a form of self-publishing; there is no peer review needed before editions appear on this platform, and publishing a digital edition in this manner is therefore limited to—and may privilege—those who can draw upon resources to access the technology and have texts already digitized in the first instance. This has always been the case for digital scholarly editions, however. Many project leaders are reluctant to discuss resourcing,⁷¹ and there has been very little consideration of the peer review before publication or evaluation after online publication of projects. That said, many digital editions would have gone through rounds of user testing. It is not clear that there is currently much to test with read&search because of the standardized functionality of its early release.

Discussion: On Automation and Digital Editions

At the time of initially writing this piece (mid-2022), read&search was the only platform automating the process of creating a digital edition (and it has since morphed into Transkribus Sites⁷² by time of publication in 2024). However, other platforms are sure to follow—particularly if publishers see that this can be monetized. The read&search tool is an emerging type of digital edition, one that is produced directly from both the inputs and outputs of Handwritten Text Recognition, without the need for further processing, data transformation, or design and hosting. Taken together, they will eventually allow critical components to be added to this basic functionality. The development of these platforms is deserving of careful consideration regarding access and opportunities, enabling a wider community to create digital editions with relative ease, especially as the functionalities of automated editions develop.

HTR injects efficiencies, for both transcription and the development of presentational digital editions, with automated tools such as read&search. However, as with many machine-learning technologies, its rollout will change procedures, ownership, and power dynamics, disrupting established processes and economies, and employment and skills development.⁷³ Although choices regarding the design and implementation of standardized editions narrow in standardized interfaces, so do editorial opportunities to engage in the infrastructure of the digital scholarly edition, both of which will limit scholars’ ability to develop the technical skill sets needed to suggest interventions and improvements. We should view these matters with care to ensure the next generation of paleographers, digital humanists, scholarly editors, and librarians enjoy opportunities to develop the crucial skill sets needed to work alongside and with near-automated platforms, while also ensuring that the developers of tools and platforms can be informed by scholarly expertise and practice. It is possible to envisage a future of supervised machine learning that can aid in the *editing* of digital editions as well as their creation, assisting scholarly editors in their task, and expanding their purviews and the volume and scale of what they can achieve.

A new “dialectic” needs to be acknowledged, between documents, editors, and artificial intelligence, if we are to make the most of the potentials and useful affordances of machine learning and related infrastructures for cultural heritage. Additionally, if we can save time in transcription and generation of digital editions, how else can we use our thought cycles to drive forward the use and understanding of historic and literary manuscript material in the digital space? How can we best use this supportive technology to democratize and increase the diversity of archival voices that are foregrounded in the building of digital editions? How will the facilitated creation of digital editions through automated processes affect the scale and range of digital editions available, and our access to, understanding of, and pedagogical approaches to the past?

Conclusion

The read&search tool demonstrates the feasibility of a presentational online scholarly digital edition of text, that is, one that is generated directly and automatically from uploaded images of texts and their HTR transcription. The tool falls short of the functionalities that would enable it to be considered as a full critical scholarly edition. In addition, the variety of technical features does not also allow for the act of scholarly editing—and the intellectual labor needed to gather witnesses, align texts, and prepare a critical apparatus. However, we are at an early stage of this type of tool's development and testing, and read&search can be thought of as a proof of concept demonstrating that the seamless technical integration of HTR to online publishing has now been accomplished. This fundamental achievement presents ramifications for the scholarly editing community that textual scholars must consider in order to understand the epistemic affordances of this developing technology at the juncture of machine learning and online technologies.

Automated editions are now available to a broad constituency, removing procedural barriers regarding transcription, formatting, and hosting of digital editions (although to date, editors can only utilize this opportunity if they can afford to use commercial services). As these technologies and platforms mature, opportunities will arise for the formation of many more online editions than before; they will expand the range of voices that are chosen, affording the editing and exegetical attention necessary, in the preparation of scholarly digital editions. Automated digital editions will continue to evolve, hopefully with the input of the scholarly community. The affordances and characteristics of critical editions will be best embedded into them with continued dialogue, experimentation, and testing of these features and functions by scholarly editors and their audiences.

Image-based web services linked to mass digitization and automated HTR transcription disrupt the labor and pipelines contained within the creation of online editions. As with many applications of machine learning in the scholarly sphere, finding where this technology can save on labor, in order to allow human activities to concentrate on the parts of the process that are unlikely or unable to be automated, should allow richer and more expansive intellectual activity, and increase the multiplicity of digital editions created by the individuals and institutions. By trialling, critiquing, engaging with, and developing these novel pipelines, the scholarly editing community has much to contribute to the development of machine-learning-related tools that can automate the publishing of images of historical documents and their related texts.

-
1. Melissa Terras, "Inviting AI into the Archives: The Reception of Handwritten Recognition Technology into Historical Manuscript Transcription," in *Archives, Access and AI: Working with Born-Digital and Digitised Archival Collections*, ed. Lisa Jaillant, 179–204 (Bielefeld, Germany: Transcript Publishing, 2022). [↗](#)
 2. Tessa Hauswedell, Julianne Nyhan, Melodee H. Beals, Melissa Terras, and Emily Bell, "Of Global Reach Yet of Situated Contexts: An Examination of the Implicit and Explicit Selection Criteria that Shape Digital Archives of Historical Newspapers," *Archival Science* 20 (2020): 139–65, <https://link.springer.com/article/10.1007/s10502-020-09332-1>. [↗](#)
 3. Johanna Drucker, *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship* (New York: Routledge, 2021). [↗](#)
 4. Herbert F. Schantz, *The History of OCR, Optical Character Recognition* (Technologies User Association, 1982); and T. L. Dimond, "Devices for Reading Handwritten Characters," Eastern Joint Computer Conference, December 9–13, 1957. [↗](#)
 5. Ryan Cordell, "'Q-i-jtb the Raven': Taking Dirty OCR Seriously," *Book History* 20, no. 1 (2017): 188–225, <https://muse.jhu.edu/article/674968>. [↗](#)
 6. Byron Leite Dantas Bezerra, Cleber Zanchettin, Alejandro H. Toselli, and Guiseppe Pirlo, eds., *Handwriting: Recognition, Development and Analysis* (Hauapauge, NY: Nova Science Publishers, 2017). [↗](#)
 7. Terras, "Inviting AI into the Archives"; and Joe Nockels, Paul Gooding, Sarah Ames, and Melissa Terras, "Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts: A Systematic Review of Transkribus in Published Research," *Archival Science* 22 (2022): 369–92, <https://doi.org/10.1007/s10502-022-09397-0>. [↗](#)
 8. Guenter Muehlberger, Louise Seaward, Melissa Terras, Sofia Ares Oliveira, Vicente Bosch, Maximilian Bryan, Sebastian Colutto, et al., "Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study," *Journal of Documentation* 75, no. 5 (2019): 954–76, <https://www.emerald.com/insight/content/doi/10.1108/JD-07-2018-0114/full/html>. [↗](#)
 9. Terras, "Inviting AI into the Archives." [↗](#)
 10. Tobias Grüning, Roger Labahn, Markus Diem, Florian Kleber, and Stefan Fiel, "READ-BAD: A New Dataset and Evaluation Scheme for Baseline Detection in Archival Documents," in *2018 13th LAPR International Workshop on Document Analysis Systems (DAS)*, 351–56 (Vienna: IEEE, 2018). [↗](#)
 11. Since this paper was accepted for publication, a conference at the University of Rostock on Machine Learning and Data Mining for Digital Scholarly Editions was held on 9–10th June 2022 (<https://www.i-d-e.de/aktivitaeten/veranstaltungen/machine-learning-and-data-mining-for-digital-scholarly-editions/>). The forthcoming proceedings are in press at time of publication: Geiger, Bernhard, Ulrike Henry-Krahmer, Fabian Kaßner, Marc Lemke, Gerlinde Schneider und Martina Scholger (Hrsg.) (Forthcoming): *Machine Learning and Data Mining for Digital Scholarly Editions*. Schriften des Instituts für Dokumentologie und Editorik 18. Norderstedt: Books on Demand. [↗](#)
 12. For example, see A. R. D. Prasad, "Application of OCR in Building Bibliographic Databases," *DESIDOC Journal of Library & Information Technology* 17, no. 4 (1997): 17–19; and Cordell, "'Q-i-jtb the Raven'." [↗](#)
 13. David C. Greetham, ed., *Scholarly Editing: A Guide to Research* (New York: Modern Language Association of America, 1995). [↗](#)
 14. Mixergy, "ABBY: How a Bulletin Board Post Changed Everything," March 20, 2013, <https://mixergy.com/interviews/david-yang-abby-interview/>. [↗](#)

15. Susan Hockey, *Electronic Texts in the Humanities: Principles and Practice* (New York: Oxford University Press, 2000), 21–23. [↵](#)
16. Mary-Jo Kline and Susan Holbrook Perdue, *A Guide to Documentary Editing*, 3rd ed. (Charlottesville: University of Virginia Press, 2008), 114, <https://gde.upress.virginia.edu/>. [↵](#)
17. Luc Vincent, “Announcing Tesseract OCR,” Google Code Blog, August 30, 2006, <http://googlecode.blogspot.com/2006/08/announcing-tesseract-ocr.html>. [↵](#)
18. AdobePress, “Adobe Acrobat 9 How-To #61: Extracting Active Text from an Image in Acrobat 9,” 2008, <https://www.adobe.com/articles/article.asp?p=1272051>. [↵](#)
19. Matthew James Driscoll and Elena Pierazzo, *Digital Scholarly Editing: Theories and Practices* (Cambridge, UK: Open Book Publishers, 2016), 6. [↵](#)
20. William B. Lund and Eric K. Ringger, “Error Correction with In-Domain Training across Multiple OCR System Outputs,” in *2011 International Conference on Document Analysis and Recognition*, 658–62 (IEEE, 2011). [↵](#)
21. Driscoll and Pierazzo, *Digital Scholarly Editing*, 6. [↵](#)
22. Driscoll, and Pierazzo, *Digital Scholarly Editing*, 6. [↵](#)
23. MLA Committee on Scholarly Editions, “MLA Statement on the Scholarly Edition in the Digital Age,” 2016, 1, <https://www.mla.org/content/download/52050/1810116/rptCSE16.pdf>. [↵](#)
24. READ-COOP, “How to Export Documents from Transkribus: Export as TEI,” 2021, <https://readcoop.eu/transkribus/howto/how-to-export-documents-from-transkribus/>. [↵](#)
25. TEI Wiki, “Transkribus,” 2019, <https://wiki.tei-c.org/index.php?title=Transkribus>. [↵](#)
26. See Melina Jander, “Handwritten Text Recognition—Transkribus: A User Report,” Electronic Text Reuse Acquisition Project (eTRAP) Research Group (Institute of Computer Science, University of Göttingen, Germany, 2016), https://www.etrapp.eu/wp-content/uploads/2016/11/TrATN-Transkribus_User_Report-2016.pdf; Tiziana Mancinelli, “Early Printed Edition and OCR Techniques: What Is the State-of-Art? Strategies to Be Developed from the Working-Progress Mambriano Project Work,” *Historias Fingidas* 4 (2016): 255–60, <https://commons.datacite.org/doi.org/10.13136/2284-2667/65>; Wout Dillen and Dirk van Hulle, “Digital Scholarly Editing: Theory, Practice, Methods,” Conference of the European Society for Textual Scholarship in Conjunction with the Digital Scholarly Editions Initial Training Network (DiXiT), University of Antwerp, October 5–7, 2016; Christiane Fritze, “Where to Go with the Digital Scholarly Edition? A Contribution from the Perspective of the Austrian National Library,” *Bibliothek Forschung und Praxis* 3, no. 43 (2019), <https://www.degruyter.com/document/doi/10.1515/bfp-2019-2068/html>; Elena Spadini, “Producing Scholarly Digital Editions,” in *Reassembling the Republic of Letters in the Digital Age: Standards, Systems, Scholarship*, ed. Howard Hotson and Thomas Wallnig, 251–60 (Göttingen: Universitäts-verlag Göttingen, 2019), https://serval.unil.ch/resource/serval:BIB_0EB07AAA45BD.P001/REF; Marie-Laure Massot, Arianna Sforzini, and Vincent Ventresque, “Transcribing Foucault’s Handwriting with Transkribus,” *Journal of Data Mining and Digital Humanities* (2019), <https://doi.org/10.46298/jdmdh.5043>; Marie-Laure Massot, Arianna Sforzini, and Vincent Ventresque, “Transcrire les fiches de lecture de Michel Foucault avec le logiciel Transkribus: compte rendu des tests” (2018), HAL Id: hal-01794139, <https://hal.science/hal-01794139v1>; Regis Schlagdenhauffen, “Optical Recognition Assisted Transcription with Transkribus: The Experiment concerning Eugène Wilhelm’s Personal Diary (1885–1951),” *Journal of Data Mining and Digital Humanities*, 2020, <https://jdmdh.episciences.org/6736/pdf>; Hizkiel Mitiku Alemayehu, “Handwritten Text Recognition Best Practice in the Beta maṣāḥaṭ Workflow,” *Journal of the Text Encoding Initiative*, 2022, <https://journals.openedition.org/jtei/4109>. [↵](#)
27. Lina Markauskaite and Peter Goodyear, *Epistemic Fluency and Professional Education: Innovation, Knowledgeable Action and Actionable Knowledge* (Dordrecht: Springer, 2017), 185. [↵](#)
28. Kirsta Stapelfeldt and Donald Moses, “Islandora and TEI: Current and Emerging Applications/Approaches,” *Journal of the Text Encoding Initiative* 5 (June 2013), <https://journals.openedition.org/jtei/790>. [↵](#)
29. Elisa Cugliana and Gioele Barabucci, “Signs of the Times: Medieval Punctuation, Diplomatic Encoding and Rendition,” *Journal of the Text Encoding Initiative* 14 (April 2021), <https://journals.openedition.org/jtei/3715>. [↵](#)
30. TEI Publisher: The Instant Publishing Toolbox, <https://teipublisher.com/index.html>. [↵](#)
31. Frank Gent, “Marjory Fleming and Her Biographers,” *Scottish Historical Review* 26, no. 102 (1947): 93–104. [↵](#)
32. John Brown, *Marjorie Fleming: A Sketch, Being the Paper Entitled “Pet Marjorie: A Story of Child-Life Fifty Years Ago”* (Boston: Fields, Osgood, & Co., 1869), 14. [↵](#)
33. S. L. Clemens, “M. Twain’s Marjorie Fleming, the Wonder Child,” in *Europe and Elsewhere* (New York: Harper and Brothers, 1923) [repr. from *Harper’s Bazaar*, 1909]. [↵](#)
34. Brown, *Marjorie Fleming*, 16. [↵](#)
35. Joe Nockels, “Automatically Transcribing the Marjory Fleming Diaries,” blog, April 27, 2021, National Library of Scotland, <https://blog.nls.uk/automatically-transcribing-the-marjorie-fleming-diaries/>. [↵](#)

36. Thomas G. Padilla, "Collections as Data: Implications for Enclosure," *College and Research Libraries News* 79, no. 6 (2018): 296, <https://crln.acrl.org/index.php/crlnews/article/view/17003/18751>. ↵
37. READ COOP, "Read&Search: Request a Quote," <https://readcoop.eu/readsearch/request-a-quote/>. ↵
38. "Recesses of Low German Cities," <https://transkribus.eu/t/rezesse-niederdeutscher-staedtetage/>. ↵
39. Noscemus Read&Search, <https://transkribus.eu/t/noscemus/>. ↵
40. <https://readcoop.eu/kati/>. ↵
41. Elena Pierazzo, *Digital Scholarly Editing: Theories, Models and Methods* (London: Routledge, 2016), 124, HAL version: hal-01182162, <https://hal.univ-grenoble-alpes.fr/hal-01182162>. See also Rehbein and Fritze's "Workflow of Editing," which separates the transcribing of a text into a machine-readable transcript and the encoding of the transcript into Encoded Text. Malte Rehbein and Christiane Fritze, "Hands-On Teaching Digital Humanities: A Didactic Analysis of a Summer School Course on Digital Editing," in *Digital Humanities Pedagogy: Practices, Principles and Politics*, ed. Brett D. Hirsch, 47–78 (Cambridge: Open Book Publishers, 2012), 52. ↵
42. Although all marked-up (and indeed, unmarked-up) texts are an expression of authorial intention and editorial interpretation; see Pierazzo, *Digital Scholarly Editing*, 109, see also 109–12. ↵
43. Vue.js, <https://vuejs.org>. ↵
44. Buefy, <https://buefy.org>. ↵
45. Node.js, <https://nodejs.org/en/>. ↵
46. Solr, <https://solr.apache.org>. ↵
47. International Image Interoperability Framework, <https://iiif.io>. See also Florian Krull, Geunter Muehlberger, and Melissa Terras, "Transkribus and IIIF: Beneficial Possibilities between Image Sharing and Handwritten Text Recognition Frameworks," 2019 IIIF Conference, Göttingen, Germany, June 24–28, 2019, <https://iiif.io/event/2019/goettingen/program/26/#Transkribus-and-iiif-beneficial-possibilities-between-image-shar>. ↵
48. Solr, <https://solr.apache.org>. ↵
49. Previously at <https://search.google.com/test/mobile-friendly>, now see <https://developer.chrome.com/docs/lighthouse/overview/>. ↵
50. Hockey, *Electronic Texts in the Humanities*, 135. As early as 2000 Hockey had already pointed out the scholarly "benefits of retaining the original lineation" between image and transcription in an electronic edition, and this feature is a basic function of read&search. ↵
51. Terras, "Inviting AI into the Archives,." ↵
52. Nockels, Terras, Gooding, Muehlberger, and Stauder, "Are Digital Humanities Platforms Sufficiently Facilitating Diversity in Research?." ↵
53. READ-COOP, "READ-COOP SCE," 2021, <https://web.archive.org/web/20220120222628/https://readcoop.eu/about/>. ↵
54. See the Transkribus Scholarship Programme for an example of how various students and teachers have been supported by Transkribus (<https://www.transkribus.org/scholarship>), and Nockels et al 2024 for an analysis of how such schemes in Digital Humanities can facilitate diversity in research (<https://academic.oup.com/dsh/advance-article/doi/10.1093/llc/fqae018/7646909>). ↵
55. Nanna Bonde Thylstrup, *The Politics of Mass Digitization* (Cambridge, MA: MIT Press, 2019). ↵
56. Transkribus Sites, <https://www.transkribus.org/sites>. ↵
57. Greta Franzini, Melissa Terras, and Simon Mahony, "Digital Editions of Text: Surveying User Requirements in the Digital Humanities," *Journal on Computing and Cultural Heritage (JOCCH)* 12, no. 1 (2019): 1–23, <https://dl.acm.org/doi/10.1145/3230671>. ↵
58. Wout Dillen and Vincent Neyt, "Digital Scholarly Editing within the Boundaries of Copyright Restrictions," *Digital Scholarship in the Humanities* 31, no. 4 (2016): 785–96, 788, <https://doi.org/10.1093/llc/fqw011>. ↵
59. See Kenneth Price, "Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?," *Digital Humanities Quarterly* 3, no. 3 (2009), <http://www.digitalhumanities.org/dhq/vol/3/3/000053/000053.html>; Hans Walter Gabler, "Theorizing the Digital Scholarly Edition," *Literature Compass* 7 (2010): 43–56, <http://dx.doi.org/10.1111/j.1741-4113.2009.00675.x>; Mats Dahlström, "How Reproductive Is a Scholarly Edition?," *Literary and Linguistic Computing* 19 (2004): 17–33, <http://dx.doi.org/10.1093/llc/19.1.17>. ↵
60. Patrick Sahle, "About," *A Catalog of Digital Scholarly Editions*, 2018 (accessed January 11, 2019), <https://web.archive.org/web/20190118041936/http://www.digitale-edition.de/vlet-about.html>. ↵
61. Richard Hadden, "Textual Assemblages and Transmission: Unified Models for (Digital) Scholarly Editions and Text Digitisation" (PhD diss., National University of Ireland Maynooth, 2018), 7, <http://mural.maynoothuniversity.ie/11172/>. ↵

62. Stephen Ramsay and Geoffrey Rockwell, "Developing Things: Notes toward an Epistemology of Building in the Digital Humanities," in *Debates in the Digital Humanities*, ed. Matthew K. Gold, 75–84 (Minneapolis: University of Minnesota Press, 2012), <https://doi.org/10.5749/minnesota/9780816677948.003.0010>. ↵
63. Cynthia Damon, "Beyond Variants: Some Digital Desiderata for the Critical Apparatus of Ancient Greek and Latin Texts," in Driscoll and Pierazzo, *Digital Scholarly Editing*, 201, <https://books.openedition.org/obp/3421?lang=en>. ↵
64. Driscoll and Pierazzo, *Digital Scholarly Editing*, 6. ↵
65. Kline and Perdue, *Guide to Documentary Editing*, 114. ↵
66. Greta Franzini, Melissa Terras, and Simon Mahony, "A Catalogue of Digital Editions," in Driscoll and Pierazzo, *Digital Scholarly Editing*, 161–82, <https://books.openedition.org/obp/3416?lang=en>. ↵
67. Library of Congress, "MODS: Metadata Object Description Schema," <http://www.loc.gov/standards/mods/index.html>. ↵
68. Patrick Sahle, in collaboration with Georg Vogeler and the members of the IDE, "Criteria for Reviewing Scholarly Digital Editions," version 1.1, June 2014, Institut für Dokumentologie und Editorik, <https://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1/>. ↵
69. Gillian Oliver and Steve Knight, "Storage Is a Strategic Issue: Digital Preservation in the Cloud," *D-Lib Magazine* 21, no. 3/4 (2015), <http://mirror.dlib.org/dlib/march15/oliver/03oliver.html>. ↵
70. Endings Project Team, "Endings Principles for Digital Longevity," University of Victoria, <https://endings.uvic.ca/principles.html>. ↵
71. Franzini, Terras, and Mahony, "Catalogue of Digital Editions," nn. 48 and 49. ↵
72. <https://www.transkribus.org/sites>. ↵
73. Aphra Kerr, "Game Production Logics at Work: Convergence and Divergence," in *Making Media: Production, Practices, and Professions*, ed. Mark Deuze and Mirjam Prenter, 413–26 (Amsterdam: Amsterdam University Press, 2019), 413, <https://www.degruyter.com/document/doi/10.1515/9789048540150-031/html>. ↵