



Open Library of Humanities

Jugaad CoLab: Decolonial Poetics, Minimal Computing, and South Asian Digital Humanities

Rohan Chauhan, University of Delhi, India, chauhan.rohan01@gmail.com

Agniv Ghosh, University of Delhi, India, agnivghosh@gmail.com

Zahra Rizvi, Jamia Millia Islamia, India, zrizviasuke@gmail.com

Saniya Irfan, Indian Institute of Technology Delhi, India, sin22197@gmail.com

Steven S George, Jamia Millia Islamia, India, stevensg96@gmail.com

Abdul Rehman, Independent Scholar, Pakistan, ch.abdulr3hman@gmail.com

Vikalp Kumar, University of Leiden, Netherlands, wikalpedu@gmail.com

A. Sean Pue, Michigan State University, United States, pue@msu.edu

Frequently translated as “hack” or “quick fix,” the North Indian term jugaad refers to a method of solving problems by “making do” with “what is at hand.” The mixed materialities and/or sources of these inventive solutions often destabilize conventional orders of representation. This article describes the formation of an affinity group, the Jugaad CoLab, that identifies three research foci—decolonial poetics, minimal computing, and pirate care—as central concerns at the intersection of digital humanities and Indian and South Asian Studies, especially outside of English, that benefit from jugaad-as-method. These concerns are reflected in the initial batch of projects, involving Bengali, Hindi, Punjabi, and Urdu materials, undertaken by group members as summer fellows, all of whom are graduate students and/or citizen scholars in India and Pakistan. Adapting technologies to diverse, multilingual contexts with uneven access to resources, the projects take up the cultural practice of jugaad as efficient and creative methods for building open-access, sustainable, and future-proof sites of knowledge, especially in this moment of crisis.

Souvent traduit par « hack » ou « solution rapide », le terme indien jugaad désigne une méthode de résolution des problèmes qui consiste à « faire avec » ce que l'on a sous la main. La mixité des matériaux et/ou des sources de ces solutions inventives déstabilise souvent les ordres de représentation conventionnels. Cet article décrit la formation d'un groupe d'affinité, le Jugaad CoLab, qui identifie trois domaines de recherche—la poétique décoloniale, le calcul minimal et le « pirate care »—comme des préoccupations centrales à l'intersection des humanités numériques et des études indiennes et



sud-asiatiques, en particulier en dehors de l'anglais, qui bénéficient du jugaad en tant que méthode. Ces préoccupations se reflètent dans le premier lot de projets, impliquant des documents en Bengali, Hindi, Punjabi, et Ourdou, entrepris par les membres du groupe en tant que boursiers d'été, qui sont tous des étudiants diplômés et/ou des chercheurs citoyens en Inde et au Pakistan. En adaptant les technologies à des contextes divers et multilingues où l'accès aux ressources est inégal, les projets reprennent la pratique culturelle du jugaad en tant que méthode efficace et créative pour construire des sites de connaissance en libre accès, durables et à l'épreuve du temps, en particulier en cette période de crise.

This paper argues for jugaad (*jugāṛ*, /d͡ʒu.gɑːr/)—a broadly North Indian term for “hack” or “making do,” using what is at hand—as a method for digital humanities research involving Indian and South Asian studies. Jugaad is a contested term, claimed across the ideological spectrum. In the wake of India’s economic liberalization, jugaad has been widely popularized in business lexicons to define a form of corporate strategy: “‘innovation’ as the ‘good enough solution to get the job done’” (Rajdou, Prabhu, and Ahuja 2012, quoted in Patel 2017, 54). Geeta Patel’s *Risky Bodies & Techno-Intimacy* links the usage of jugaad by economic elites to the era of late capitalism, with its “flexible labour and time shortages” that produce conditions in which “one uses only what one can immediately lay one’s hands on” (Patel 2017, 54). Such elites might use jugaad to describe short-term, self-interested moves of crafty individuals, such as a politician’s fix or a consumer’s acquisition of a coveted object. And yet, as she writes, this is a misuse of the term, for “anyone who has spent time in South Asia and seen the vehicles, cars, motorcycles, machines, clothes, lovingly made over from the parts at hand that were repurposed, has seen first-hand the tender and pragmatic invocations of *jugād* at work” (Patel 2017, 54). A truer form of jugaad, Patel argues, is grounded in an ethics of care that forms an appropriate basis for the production of art, writing, and theory in the humanities.

Following Patel’s lead, we suggest jugaad as a method governing digital humanities practices engaging with cultural materials from India and South Asia. In this domain, as we will argue, jugaad involves blending materialities, flexible regional attachments, collective as well as individual efforts, concerns with constraints and reproducibility, and the potential disruption of established orders of representation. We describe the formation and research areas of a transnational affinity group, the Jugaad CoLab, which for administrative purposes formed as a lab in the Indian and South Asian Studies program at Michigan State University. The CoLab takes on jugaad-as-method for digital humanities scholarship. The initial batch of research projects conducted by graduate students and citizen scholars in India and Pakistan involves the study and preservation of cultural materials across audio, digital, textual, and visual media in Bengali, Hindi, Punjabi, and Urdu. The conclusion relates these projects to this understanding of jugaad digital humanities and addresses the challenges and opportunities of this collaboration.

Jugaad digital humanities

As a model of digital humanities practice involving India and South Asian studies, jugaad suggests a number of principles. First, by prioritizing getting things done over purpose-built solutions, jugaad allows the digital humanist to blend materialities. It therefore suggests improvising and designing a workflow to suit technical

proficiencies and hardware constraints (Rizvi et al. 2022). Second, jugaad’s widespread comprehensibility across India and South Asia suggests a broad regional focus on cultural materials emerging from the member nations of the South Asian Association for Regional Cooperation (SAARC)—Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka—and their diaspora, the world’s largest. Included in this potential corpus are the local and diasporic digital cultures of a quarter of the world’s population, conducted in Indian and South Asian languages other than English (cf. Pue, Zaidi, and Aqib 2022; Zaidi and Pue 2023). Thirdly, jugaad can be both collective as well as individual, flexible as well as makeshift. Fourthly, jugaad suggests considering and working through the constraints—be they of language, script, hardware, software, or electricity—faced by a larger scholarly community, which includes academics as well as citizen scholars. Finally, jugaad opens the possibility that the combination of forms, assembled with care, can work through juxtaposition to disrupt existing orders of representation.

Blending materialities

One of the popular uses of jugaad refers to makeshift automobiles, which are made by combining multiple vehicles and repurposing parts. Drawing on such associations, technology theorist Pankaj Sekhsaria emphasizes jugaad’s dependence on “reconfiguring materialities” of existing tools to “overcome obstacles and find solutions” (Sekhsaria 2013, 1153). For him, jugaad provides a means to overcome the kinds of conditions common to Indian and South Asian universities, especially technical and hardware constraints (cf. Thangavel and Menon 2020). Sekhsaria considers how, for over twenty years, training in physics also required training in the making and modifying of “indigenously built” scanning tunnelling microscopes, which involved “junk markets, scrap materials, small time spring makers and second-hand dealers” (Sekhsaria 2013, 1152).

The digital remix, like the remixed film song, is “a material practice.” As Kuhu Tanvir and Ramna Walia have argued about the latter:

be it the disc-scratching that DJs fall back on to introduce an electric scratching sound to songs, or actually separating elements of a song wherein the melody, the beats, the lyrics that have formed a cohesive union in the original are stripped from the whole, and readied for new combinations and thus new musical objects, remixes are all physical practices. By definition, therefore, a remix is reliant upon an existing sonic media object, leaning not just on its elements but on its memory. (Tanvir and Walia 2021, 166)

For Indian and South Asian studies, the remix of source materials holds the potential of similar disruptions, revealing the cultural object in a new form, distinct from the original.

Broad Indian and South Asian regional focus

Jugaad, as a concept, is used throughout North India, as well as Pakistan and in the diaspora, and as such lends itself to a broad cultural specificity. Building on Sekhsaria's account, digital humanities and design scholars Padmini Ray Murray and Chris Hand insist that jugaad "should be understood in its culturally and historically specific contexts," rather than being more simply assimilated into the English term "hack" (Ray Murray and Hand 2015, 152). In English, they note, "hack contains within it both the meaning of subverting the authority of proprietary systems through some sort of destructive action" and "to come up with a quick solution" (Ray Murray and Hand 2015, 144). The "aim of jugaad," on the other hand, "is almost always constructive," and "is truly born out of necessity" rather than the desire to undermine capitalist systems that often drives "DIY" and "Maker's" movements (Ray Murray and Hand 2015, 144). Noting that the terms of debate in the digital humanities have "thus far almost exclusively been determined by scholarly work done in American, Europe, and Australia" (Ray Murray and Hand 2015, 142), the authors argue that "approaches that privilege the local should be seen as extending the limits of digital humanities practice," even if doing so appears, "albeit superficially, to contradict the universalising impulse of the discipline" (Ray Murray and Hand 2015, 152–153). As noted above, we understand the local context for jugaad to be transnational and multiethnic, including the contemporary SAARC nations and the Indian and South Asian diaspora—a context based more in a commonality of position than of identity (cf. Spivak 2008).

Collective and individual, flexible and makeshift

By considering the collective and flexible practices of jugaad, this usage challenges a common negative understanding of the method as "part and parcel of India's systemic risk," "a product of widespread poverty" that "underpins path dependencies stemming from dilapidated infrastructure, unsafe transport practices, and resource constraints," and, as such, is "wholly unsuitable both as a development tool and as a business asset" (Birtchnell 2011, quoted in Mukherjee 2020, 110). Game studies researcher and digital humanist Souvik Mukherjee argues against the celebration of jugaad, particularly because, in his formulation, it emphasizes individual ingenuity over collective, purpose-driven effort. We challenge this understanding of jugaad as

productive of inadequate, unreproducible “quick fixes” to argue, in digital humanities, for the ways and means that can be both collective and individual while also flexible in addition to makeshift.

Concern with constraints

In emphasizing jugaad in digital humanities practice, we grapple with the constraints of our larger research community, understood to include researchers and citizen scholars in the SAARC nations, the diaspora, and in the fields of Indian and South Asian studies in general. Digital humanist Roopika Risam identifies jugaad’s closest affinities in digital humanities to be with minimal computing, which she defines as an “approach to digital humanities design and method that eschews high-performance computing in favour of practices that are more accessible around the world,” and “draws on a range of cultural practices that privilege making do with available materials to engage in creative problem-solving and innovation” (Risam 2019, 43). As she writes, jugaad is just one among other cultural practices that find expression in minimal computing, such as “*gambiarra* in Brazil, *rebusque* in Colombia, *jua kali* in Kenya, and *zizhu chuangxin* in China” (Risam 2019, 43). Minimal computing focuses on developing DH projects that accommodate resource-poor environments characterized by low bandwidth and slow speeds, and therefore privilege open-source platforms and static web pages. She also identifies within minimal computing a perspective that considers “the social impacts of computing through postcolonial lenses,” and that understands the powerful difference between minimizing computing out of necessity and choosing to consider issues of obsolescence and e-waste from the beginning of projects (Risam 2019, 43). In short, “minimal computing within digital humanities embraces postcolonial critiques of globalization and technology, modeling responsiveness to issues of access, wealth, and uneven development” (Risam 2019, 43).

In a recent special issue on minimal computing, Roopika Risam and digital humanist Alex Gil elaborate on its definition. They write that minimal computing “advocates for using *only* the technologies that are necessary and sufficient” under certain constraints, whether they be “lack of access to hardware or software, network capacity, technical education, or even a reliable power grid” (Risam and Gil 2022). They note that “minimal” does not necessarily imply “ease,” and Gil and Risam find most troubling that a “reduction in computation” often requires more “technical education” (Risam and Gill 2022). Minimal computing “gestures towards a decision-making process driven by the local contexts in which scholarship is being created,” especially its constraints, be they of “resources (e.g., financial, infrastructural, and labor)” or “freedoms (e.g., movement and speech)” (Risam and Gil 2022). The constraints on Indian and South

Asian studies scholars in the region, diaspora, and elsewhere are such that minimal computing approaches appeal as a method of increasing participation in the field.

Representational disruption through juxtaposition

Finally, we understand jugaad in digital humanities as a method for destabilizing established orders of representation, especially through productive juxtaposition. This principle motivates Geeta Patel's focus on the role of care in jugaad, noting that its practices often produce an aesthetically uneven result, without the seamlessness that might come from purpose-made replacement parts or matching fabrics. These jugaad are "lovingly made" and deemed "good enough"—in the sense of the phrase as applied in popular psychology to the "good enough mother" who meets her child's needs, even as she does not perfectly fulfill the child's every rising desire. By extension, Patel describes jugaad theory, whether in art or writing, as involving the sometimes-clashing aesthetics of a multiplicity of elements and techniques that is most visible in "the playful, unexpected, sometimes cordial, occasionally dissonant juxtapositions" that "tease open the fraying edges of disciplines, so that disciplines begin to shade into something else" (Patel 2017, 343). These juxtapositions incite the observer and the practitioner "to push and pull against the conventions of representation" (Patel 2017, 54), in a "good enough" bricolage-style argumentation that upsets conventional orders of representation and of narration. We hold this possibility open for jugaad digital humanities: using the affordances of the digital, including juxtaposition, to challenge established understandings.

Jugaad CoLab

The Jugaad CoLab affinity group grew out of a multi-year collaboration between the Department of English at Jamia Millia Islamia (JMI) and Michigan State University (MSU), funded by the Scheme for the Promotion of Academic Research Collaboration (SPARC) (SPARC 2023), awarded to JMI, focused on "Digital Apprehensions of Poetics." The SPARC program aimed to increase the research capacity of India's public central universities through collaboration with top foreign institutions. The collaboration between JMI and MSU aimed to expand the capacity for digital humanities research in India through a focus on digital approaches to the study of Indian language poetics while also extending the study of Indian language poetry to include its digital forms. Poetry remains an extremely vibrant literary form for personal and collective expression across Indian and South Asian languages, circulating more widely than ever before, especially through digital platforms (Pue, Zaidi, and Aqib 2022). The SPARC program enabled graduate students from Indian universities to work abroad, and it

supported workshops, conferences, and collaborative research projects between the two institutions. The conferences, in particular, made clear the wide range of interest in digital humanities research among contemporary scholars in India and the diaspora, as well as the constraints and challenges faced by this community (Zaidi and Pue 2023).

The Jugaad CoLab was formed under the direction of A. Sean Pue, the international PI on the SPARC grant, to further these research goals. Housed in the Indian and South Asian Studies program at Michigan State University, the lab avoids the perceived administrative difficulty of facilitating further partnerships across South Asian nations, and particularly between universities in India and Pakistan, following the standard practice for supporting cross-SAARC scholarly collaboration in the region. Initial funding for the CoLab came from MSU matching funds promised as part of the SPARC fellowship, provided by the endowed Delia Koo fellowship program at MSU's Asian Studies Center. Koo funds supported research projects based in South Asia through a series of summer research fellowships awarded through a competition for "Digital Apprehensions of Poetics" projects. The competition was open to students, faculty, and alumni of central Indian universities. An additional fellowship competition, not restricted to India's central public university affiliates, was funded through an Andrew W. Mellon New Directions Fellowship. Several of the recipients used their fellowship funds to compensate additional team members, such as student data editors, meaning that the summer fellowship model proved advantageous in growing the project's network of participants.

Research foci for Indian and South Asian digital humanities

In dialogue with each other and members of our larger research community of academics and citizen scholars involved in digital humanities scholarship related to Indian and South Asian studies, we have identified three primary research foci for the Jugaad CoLab: decolonial poetics, minimal computing, and pirate care.

Decolonial poetics

Our collaboration initially focused on Indian-language poetics and was driven by a relatively traditional understanding of poetics as the study of poetry's form and content (cf. Rizvi et al. 2022), but as our affinity group took shape, we adopted the more expansive understanding of poetics grounded in an idea of language as a "representational system" in which "we use signs and symbols—whether they are sounds, written words, electronically produced images, musical notes, even objects—to stand for or represent to other people our concepts, ideas and feelings" (Hall 1997,

6, 3). This capacious understanding of poetics is appropriate to the multimodal cultural materials and archives that Indian and South Asian research involves.

Decolonial poetics acknowledges the effects of colonial knowledge and power relations on poetics and their study. Historical accounts have described how both British officials and Indian social reformers decried South Asian poetics as morally bankrupt when contrasted to English models (cf. Pritchett 1994). Our work considers the effects of cultural hegemony, defined as conditions in which “certain cultural forms predominate over others, just as certain ideas are more influential than others” (Said 1979, 7), on South Asian poetics. Decolonial poetics attends to such discursive histories and is concerned with “the way representational practices operate in concrete historical situations, in actual practice” (Hall 1997, 6).

Minimal computing

As championed by Alex Gil and Roopika Risam, minimal computing involves approaches to computing that address local constraints, such as those of language, access, and resources. In practice, Jugaad CoLab developers focus on reproducible methods using a stack of technologies available to our larger research community that are free, open access, open source, multilingual, and collaborative. As Maya Dodd has argued, “[t]he development of decolonial thinking needs a real decolonization of how we will know, and the digital offers this as a unique possibility to enhance access for use and for re-creation” (Dodd 2023, 30). We understand minimal computing, then, as part of our consideration of “how we will know” that emphasizes accessibility, creativity, reproducibility, localization, reuse, and environmental impact.

Pirate care

The final research focus, pirate care, brings the CoLab into alignment with a transnational movement concerned with the crisis and criminalization of care in the contemporary era. Two features of this movement are an agreement on the need for a “common care infrastructure” and for an “exploration of the mutual implications of care and technology” (Graziano, Mars, and Medak 2019). While pirate care’s goals are wide-ranging, they motivate the CoLab’s goals of preserving and engaging the human cultural record of a quarter of the world’s population. Within the domain of digital humanities, pirate care involves contributing to the “digital commons,” using the “affordances of the Global South and the tools at hand” and working and designing for constraints while privileging collaboration (Grallert 2022).

We adopt these research foci as three central concerns for Indian and South Asian digital humanities. Our focus on a “common care infrastructure” should allow digital humanities groups to bring together technical and infrastructural components and expertise from other domains to meet local constraints and to seek efficient and creative methods for building open-access, sustainable, and future-proof portals of knowledge. At the same time, these solutions enable the potentially transformative juxtapositions associated with jugaad.

Research projects

The research projects of the Jugaad CoLab illustrate and build out jugaad as a method for digital humanities scholarship. The first batch of fellowships, described below, involves materials in Bengali, Hindi, Punjabi, and Urdu. They use jugaad to address the CoLab’s research areas of decolonial poetics, minimal computing, and pirate care.

Project 1: OCR of historical South Asian documents

Rohan Chauhan’s project at the CoLab addresses the lack of easy-to-use, open-source optical character recognition (OCR) technology for historical documents in South Asian languages by creating ground truth datasets that can be used to train and test cutting-edge OCR/HTR engines. These datasets are designed to broadly reflect the particularities of early printed texts in three South Asian languages with different scripts: Urdu (اردو, right-to-left in the Nastaliq script), Bengali (বাংলা, left-to-right), and Hindi (हिन्दी, left-to-right).

While OCR of contemporary fonts in European languages is considered a solved problem, historical documents are yet to truly benefit from this technology (Reul et al. 2019). Historical documents in several South Asian languages, even those printed as late as the 1940s, remain even further removed from these technological developments. This restricts the proliferation of high-quality, machine-actionable textual repositories in South Asian languages, which, in turn, limits the application of computational tools and methods on South Asian historical print. This technological imbalance keeps historical print in South Asian languages outside the purview of the most ubiquitous practice of our times, that is, search, which Ted Underwood rightly explains as nothing but “algorithmic mining of large electronic databases” that “has been quietly central to the humanities for two decades” (Underwood 2014, 64).

A representative example is Bichitra: The Online Tagore Variorum (Bichitra 2023), a project funded by India’s Ministry of Culture focused on Rabindrinath Tagore’s writings. The project team attempted to use OCR, but they reported that results “were often so different that we virtually had to create new files from scratch” (Chaudhuri

2015, 60). The task of rekeying “kept most of the team of 30 busy for 18 months and more, a few of them till almost the end of the 2-year project” (Chaudhuri 2015, 60). Citizen scholars and government-funded initiatives have also attempted to address the dearth of high-quality, machine-actionable textual repositories in South Asian languages. These include the work of different language communities on Wikisource (Wikisource 2023), an effort particularly noteworthy for Bengali; Pakistan’s Iqbal Cyber Library for Urdu (Iqbal Academy 2023); and West Bengal government’s Department of Information Technology and Electronics for Bengali (Department of Information Technology & Electronics 2023). Such initiatives, while valuable, are not terribly extensive and are practically nonexistent in some major South Asian languages, including Hindi. More importantly, freely available labelled ground truth remains lacking for both handwritten and printed South Asian languages, which makes their reusability impossible in a variety of tasks (Smith and Cordell 2018).

State-of-the-art OCR engines use trainable models to perform two consecutive tasks that produce machine-actionable transcriptions. First, they segment line-coordinates on the page, and second, they transcribe the text by each segmented line. To train OCR models in any cutting-edge OCR engine, we need training data. This training can either be synthetic (e.g., automatically generated artificial images from contemporary computer fonts aligned to existing transcriptions) or real data, prepared manually as line-by-line annotation and transcription of digital scans. As preparing synthetic training data can be automated and takes significantly less time in comparison to preparing real ground truth datasets, synthetic data is often preferred. However, models trained with this approach usually yield poor transcriptions of historical documents, as computer-generated fonts do not represent the sheer variety of historical typefaces, rendering the OCR output impractical for humanities use (Springman and Lüdeling 2017).

Historical typography is particularly complex in South Asian languages. Urdu scholar Frances Pritchett has evocatively captured the situation in Urdu: “Words are written with no spaces between them, words are often crammed irregularly into every available cranny for reasons of space and/or aesthetics, two or more words are written together” (Pritchett 2006). Historical print in Bengali and Hindi also poses a set of challenges in terms of typographical complexity, specifically with their large repository of consonant clusters. Variation in historical spellings or irregularities in representations of particular glyphs and word boundaries complicate the situation further, for instance, the treating of *do-chasmi he* (دوچاسمی ہے) as a variant of *choti-he* (چوٹی ہے) for aspiration in Urdu or the irregular spacing between nouns and their case markers in all three languages. South Asian historical documents often feature more than one language in their text, such as English in addition to the main language, but also Urdu *nastaliq* in Hindi texts, for example. In addition, document layout of historical print in

South Asian languages has significant irregularities as well, particularly multi-column and other complex layouts prevalent in periodicals, dictionaries, or books with in-text images, ornamental features like borders, and paratextual elements like footnotes and marginalia. This, in turn, limits a cutting-edge OCR engine in correctly predicting the position of the text on the page before it even gets to the text recognition task.

Overcoming these constraints for humanities research requires providing OCR engines with annotated ground truth datasets. To prepare training data for this project, Chauhan relied on eScriptorium—a multi-institutional initiative led by textual scholars and committed to the principles of open science—which provides an interactive digital text processing pipeline for printed and handwritten documents (Stokes et al. 2021). It tightly integrates Kraken, a language-agnostic OCR engine designed for humanists (Kießling 2019), and offers significant flexibility and openness over closed services, such as Transkibus, by allowing the user to export and import both datasets and models (Figure 1). Furthermore, users can customize controlled vocabulary in eScriptorium to label diverse elements of the page layout, which bodes well for addressing automatic detection and classification of text blocks or regions in historical documents with complex layouts. It also assists in annotating line types, which is particularly useful for capturing an extensible number of features for both region and line types, for instance, the title of an item or author signature in periodicals or verse and prose in literary histories.

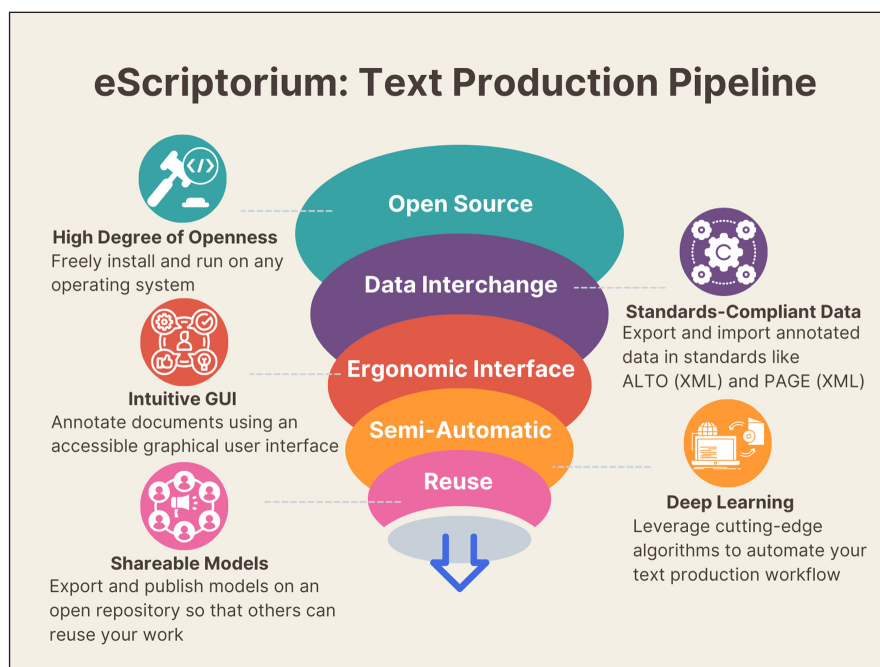


Figure 1: An Overview of eScriptorium.

Automatically analyzing layout, particularly in documents with complex layouts, involves training state-of-the-art segmentation models that locate and label region and baseline types on each page of a text. As the output of the segmentation step is the input of the recognition step, faulty segmentation can negatively impact recognition, which is why it is particularly crucial to get this step right. Moreover, segmentation is also important for automatically converting OCR'd texts outputted as ALTO (Analyzed Layout and Text Object) into minimally encoded but logically structured TEI (Textual Encoding Initiative) files (Janes et al. 2021; Pinche, Christensen, and Gabay 2022). Segmentation also reduces the challenge of efficiently extracting the content of an OCR'd document from region types, such as the main body, marginal notes, title, numbering, tables, figures, or illustrations. This readily applies to a variety of text types, including periodicals, where it has the potential to open up the presently “intractable” archive (Bode 2018, 124) of the scanned imagery of South Asian periodicals for digital inquiry. This step is most demanding in terms of editorial decisions, especially for incrementally complex layouts, which are quite common in Urdu and in particular text types, such as periodicals, dictionaries, textbooks, and so on.

For his project, Chauhan adapts the controlled vocabulary or ontologies provided in the SegmOnto project to annotate region and line types on each page within eScriptorium (Gabay et al. 2021). He then retains these annotated datasets to train machine learning models, for which he has explored two approaches. The first uses Kraken—the OCR engine associated with eScriptorium—to train a segmentation model that is able to infer both region and line types. It treats region detection as a pixel classification task and is prone to merge neighbouring regions of identical types into a single region. The second approach involves YALTAi (Clérice 2022), a framework that treats the region detection task as an object detection task by repurposing the YOLO (You Only Look Once) family of computer vision models for document layout analysis. YALTAi further integrates Kraken for inference and provides a pipeline to train and apply the fine-tuned YOLO vision models for region detection and Kraken’s segmentation model for baseline detection. The output, which is in ALTO, can then be imported into eScriptorium for manual correction. Preliminary results suggest segmentation with YALTAi has an edge over segmentation with Kraken; however, this gain in performance comes at the cost of the segmentation task being hardware intensive in YALTAi (Figure 2).

Accounting for complexity in typographical features, such as variation in the representation of a glyph or historical spellings, requires transcribing diplomatically, that is, exactly as the text is represented in the underlying scan of the document. Specifically, the ground truth datasets for character recognition training for each language developed in Chauhan’s study comprises 30–70 pages per book drawn

model in Kraken for the specificities of a particular text by retraining it on a sample of 100–250 lines from that text. This process of increasing—or resizing, in Kraken’s terminology—the knowledge of an existing model for book-specific needs drastically improves the efficiency of the OCR by 2–3% in terms of CAR or up to 5% in terms of WAR. Since fine-tuning a generalized model requires far fewer examples of lines than training from scratch, it is also efficient in terms of the hardware, electricity, and labour needed to produce high quality transcriptions. Retaining the ground truth prepared during document-specific fine-tuning to retrain a generalized model from scratch further updates the generalized model for a wider coverage of typefaces. Following this virtuous cycle of iteration, we build efficient OCR models that are capable of handling a wide variety of typefaces across the print history of our target languages and make them freely available to researchers for reuse via an open GitHub repository (Figure 3).

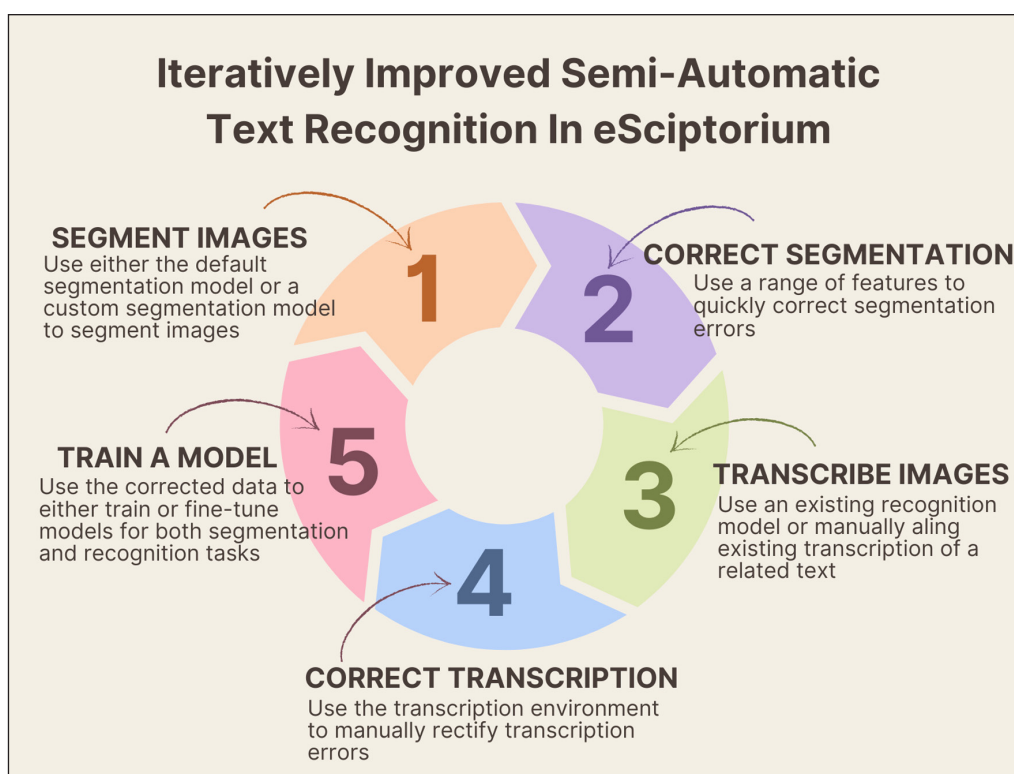


Figure 3: Typical Workflow in eScriptorium.

Moving forward, the CoLab publishes these ground truth datasets for reuse in a variety of contexts, including humanities research. For this, we benefit from standards established by the eScriptorium community as well as others, especially those

concerning restrictive copyright clauses attached to the underlying digital imagery (Lassner et al. 2021). Research is also ongoing in better capitalized initiatives for Urdu (e.g., Open Islamicate Text Initiative [OpenITI]), and the CoLab hopes that our work can benefit their work as theirs has benefitted ours.

Project 2: Metadata for Bengali periodicals

Agniv Ghosh and Rohan Chauhan's project works to improve item-level metadata for Bengali periodicals. It is also a capacity-building collaboration with the goal of promoting frugal innovation in underserved Indian language classrooms to improve existing standards of transparency, credit-sharing mechanisms, data ownership, and data management, while developing new literacies and transferable skills among students. Specifically, this project builds on the resources created in an academic writing course that Agniv Ghosh has been teaching to students of the M.A. Bengali program in the University of Delhi's Department of Modern Indian Languages and Literary Studies since 2017. By working closely with students of Ghosh's class, this collaboration improved the reusability of the resources produced, and provided a suitable workflow to meet the objectives of this course by "making do" with simple and ubiquitous digital tools.

Students in Agniv Ghosh's class are expected to focus on nineteenth- and twentieth-century Bengali periodicals to develop their individual projects. The objectives of this course are to introduce students to the fundamentals of archival research, and to help them develop thematic bibliographies according to their individual projects. By working directly with historical sources, the students are encouraged to rethink the received categories through which Bengali culture is typically conceived. Here, each student chooses a few issues of a particular periodical and transcribes each item in a basic bibliography format. They then collate everyone's work in a single file and select the items they want to use for their dissertation based on a theme or a genre. Before the implementation of this grant, Ghosh's class was using Microsoft Word to manually transcribe these records, which was neither ideal for their research needs nor sustainable in the long term.

The fellowship from Jugaad CoLab gave Agniv Ghosh and Rohan Chauhan the opportunity to streamline this approach, while also developing a reproducible workflow accessible to the typical Indian academic. Between June and December 2021, the team organized the project in a shared Google Drive folder with the help of graduates of the M.A. Bengali program housed at their home department in University of Delhi. Google Drive offered a versatile and costless collaborative working environment, and the easy learning curve associated with it made it fairly straightforward for the less technically

inclined team members to master quickly. It also allowed the team to preserve some level of version history for the files and to track student contributions transparently, particularly for credit attribution in case of reuse. The team maintained a to-do list in the form of a Google Doc to organize the project. This Google Doc was at the centre of the project's workflow. The team used this document to add tasks and post insights or problems arising during the work in the following format: a brief descriptive heading, date of entry, name of the person who made the entry, and details of the entry. The team tracked tasks and their related issues by assigning hierarchized heading tags to tasks and issues within them, so that they were easy to traverse using the outline feature in Google Docs. The fellows found the format of these entries particularly useful for transparent credit attribution. For everyday communication, they maintained a WhatsApp group—a ubiquitous instant messaging service in India, hence ideal for the needs of the team—and relied on Google Meet to conduct team meetings.

The first problem tackled in this project was in the shape of several Microsoft Word files where the item-level records were originally transcribed by students in Ghosh's Bengali academic writing class. These files had a number of problems. Firstly, records of items appearing in a particular issue of a periodical were transcribed as a list in the MS Word file and only documented information for author, title, genre, and page number, while entries on publication date, volume, and issue number were provided as a heading to each issue's content list. Secondly, there was no consistent separator across the several thousand items it contained. Thirdly, the publication date format was according to the Bengali calendar, which was not compatible with the bibliography software Zotero, which the team wanted to use for delivering this data to other researchers. Based on the solar year, this date format, while fairly easy to understand for students of Bengali, posed its own set of challenges for machine-parsing, especially in spreadsheets. And lastly, apart from missing entries for book reviews published in these periodicals, the existing metadata did not include information on fields such as the segment of a periodical under which a particular entry featured, or information on the source in case of translation or republication from another source.

The team began by transforming the data in Word files into comma-separated values. They first copied this data into text files and converted inconsistent separators into commas using regular expressions. At this stage, it was tempting to simply change the extension of the file from .txt to .csv; however, this would have required manually supplying data fields for issue number, volume number, date of publication in Bengali calendar along with its Gregorian conversion, and the name of the editor and the periodical for nearly 20,000 entries for the journal *Prabasi* alone. To mitigate this, they put together a Python script that essentially segmented each record in an

item at a consistent separator as different attributes corresponding to a set of entities broadly defined to mimic the structure of a bibliography, that is, author, title, genre, segment name, page number, issue number, periodical, date, and editor, and appended the data accordingly. Working with Pandas (McKinney 2010), a Python library, they were able to easily tidy the data by transforming columns wherever required, which was particularly convenient for converting the publication dates in the Bengali Calendar to Gregorian Calendar. After this, they imported the data as a CSV file into a shared Google Drive folder, where the data editors edited the file as a spreadsheet based on a set of predefined guidelines. The fellows also experimented with Bengali OCR to get these records from existing analog bibliographies. Here, they worked with *Kallol* (1923–1929) using Jibendra Sinha Roy’s study of the periodical that includes a list of items published in its entire run and is available from the Internet Archive (Roy 1960). The plain text output of an existing analog bibliography turned out to be an extremely flexible and time-saving way to convert into a CSV file using a mix of automation and manual correction.

The CoLab grant also enabled the fellows to pay stipends to the data editors, who teamed up on this project from August 2021 to December 2021. These students came to this project straight out of their M.A. Bengali course and had no exposure to the field of digital humanities. They also lacked any experience in the technologies the fellows wanted to use in this project (i.e., spreadsheets, OCR, text editors, and text encoding). Ghosh and Chauhan progressed by organizing online learning sessions on technologies they wanted to use in the workflow, and briefly contextualized their usage within digital humanities. Here, like Cordell suggests, the fellows refrained from defining DH but allowed the student to define the field for themselves through their direct engagement with what they were making. Ghosh and Chauhan were apprehensive to assign readings from the long list of “what is DH” definitional pieces to the data editors, not just because the data editors do not read English well—the language in which most of these pieces are written—but also because they see the field as infinitely local, where the definition will necessarily vary both due to a discipline’s (literary studies, media studies, history) own peculiarities and the local context of its practitioners (Cordell 2016). Instead, the fellows chose to organize sessions to discuss the affordances of what the team’s labour will result in, that is, an indexable itemized bibliography of Bengali periodicals, an area of study all the data editors were interested in. These sessions got the data editors excited about the value of their labour. They were most enthusiastic that their contributions will organize and future-proof the written heritage of their research language. They were also expressive about how their work will improve the discoverability of primary material, not just for their research needs, but also for those of other researchers. The

student data editors also integrated the deliverables of this project, along with their learning during their engagement with it, into their own projects, which Ghosh and Chauhan found most rewarding.

Between January and May 2022, Agniv Ghosh extended this approach in his class. Unlike previous iterations of this course, his students now work in a shared Google Drive folder and use spreadsheets instead of MS Word to prepare the bibliographic records. Wherever applicable, the students use the OCRed text of an existing analog bibliography, edit the output, and convert the same into a CSV file. The ability to track changes and contribution history in Google Drive enables each student contributor to maintain ownership of their work. As the students work in a cloud-based storage solution, their data remains protected from redundancies of drive failure while enabling them to collaborate with their peers in real time. More importantly, their work in this class also equips them with useful transferable skills, thereby improving student outcomes in the course. Since this data is now stored as comma separated values, it is also flexible for a number of affordances such as searching, filtering, and pivoting. Apart from the 40,000 items in *Prabasi* (1901–1930), *Bharatbarsha* (1913–1930), and *Kallol* (1923–1929) that the team transformed for this project, Ghosh's class has collaboratively added another 16,000 items from 32 different periodicals published roughly between 1843 and 1940 during this year's iteration of the course from January to May. For each item published in a particular periodical, the itemized bibliography provides a record for the following fields: author, title, genre (primarily copied from the digital surrogate, e.g., the periodical, or editorial interpolation in case of its absence in the data source), segment name, source (if a translation or republication), issue number, volume number, bound number, page number (based on the original source used for data validation), editor, and periodical. Furthermore, Ghosh and Chauhan are presently exploring publication of this data using Zotero with data both in the Bengali script and Roman transliteration, prepared using a Python-based transliteration service for a variety of Indian or Indian scripts, Aksharamukha (Rajan 2021).

Digital surrogates of Bengali periodicals are available from a number of digital repositories, including the Internet Archive. The problem, particularly on the Internet Archive, arises in finding these copies through a keyword search, as they lack basic metadata (cf. Sagar 2018). It was also practical for the needs of this course to have a handy extensive record of Bengali periodicals online at a single point of reference. For this, the team organized the links to Bengali periodicals available from three digital repositories, including the Internet Archive, in a presently private Zotero group primarily to improve the discoverability of different volumes of each periodical for their work. This Zotero group, presently maintained by Agniv Ghosh with help from

the students of their home department, continues to collect and improve volume-wise metadata for digital surrogates of nineteenth- and early twentieth-century Bengali periodicals.

Additionally, the team is also curating a machine-actionable genre-specific dataset of Bengali poetry. Here, they sample poetry anthologies published by authors who feature in the periodical dataset. They also include selections of poetry and literary criticism on poetry that appeared in periodicals covered in this project, which in turn results in a specialized dataset to study the development of literary modernism in Bengali.

Project 3: Noha and other Urdu poetic keywords

Zahra Rizvi's project focuses on contributing to and managing a collaborative Urdu Poetic Keywords project with other members of the CoLab and beyond. Urdu has developed a rich and evolving terminology in parallel with its long tradition of literary criticism, which is constantly reworked in digital spaces, popular culture, and subcultures. The CoLab is interested in exploring whether digital literary cultures pose a shift in the usage of certain words that have a longstanding tradition of rigorous scholarly reflection in poetic and formal structures across Indian languages (Rizvi et al. 2022).

The keyword Zahra Rizvi is currently working on is *noha*, a short rhythmic poem, often accompanied by *matam* (lamentation) and performed at the conclusion of the *majlis* (gathering) that usually engage with themes of loss, lament, and mourning. Noha, to recall Pinault's comment on South Asian Shiism, presents a collapsing of the past and the present (Pinault 1997). As such, it is especially conducive to media archaeology, which Zahra Rizvi combines with autoethnography to approach noha as both a keyword and a key archive for scholars.

Developing from Zahra Rizvi's media archaeology project, which specifically deals with contemporary noha online, the noha poses an interesting intersection of performance, literature, and subculture. Nadeem Sarwar, one of the most popular *noha-khwans* of contemporary times, has a YouTube noha channel totalling 1,322,418,565 views alone. Noha is one of the most vibrant and hugely popular participatory poetic forms. However, due to its simple language, structural freedom, and colloquial dependence, often in the hands and voice of women, noha is mostly unexplored by scholars. Yet Zahra Rizvi's ongoing finding, as she collects and curates contemporary noha available online, is that these very qualities serve to make the noha accessible and participatively replicable, available through not just published collections and handwritten curated *bayaz* (noha notebooks), but also cassettes or CDs; legal or

illegal channels; radio; TV; official and unofficial websites; social media like Twitter, Instagram, and Facebook, where noha-khwans have an active presence; and YouTube.

One of the main challenges Zahra Rizvi has seen in the course of her project is the variants and discrepancies amongst the noha available online, especially those transliterated into Roman script, as these transliterations are often all that is available. The transcription work is usually carried out by noha readers and enthusiasts voluntarily, and, as South Asia features immense regional linguistic diversity, the transliterations, which anyway are carried out in simple forms, lose nuances from the original languages and also require a certain amount of cleaning up (in case of typos or errors).

Since noha is practiced as an oral form, the noha write-ups in themselves are not enough. The tune of the noha and rhythm of the matam might not be readily available along with the write-up. One has to find someone who knows the noha, a video on YouTube, or a recording uploaded to online archives (not all of which are indexed). Some noha might be available only as a write-up or as a recording, in which case the missing element has to be tracked down.

Noha websites also appear and disappear, as sustainable authorship is still not as accessible as would be desirable. As collection and curation continues, there is a challenge to understand platform compatibility and to design and distribute models in and for curation and archiving. All the while, there is little institutional support, and a greater need for more interconnected research, especially with citizen scholars, as collaborators are few and far between.

Project 4: Digitally mapping the Urdu marsiya

Marasi (plural of *marsiya*), or elegies, have an old tradition with roots in the pre-Islamic Arab and Persian world, where human sentiments and pathos were expressed in the form of elegiac poetry (Nicholson [1907] 1969). As a popular form of religious and cultural expression, the subject matters of most of the *marasi* are the battle of Karbala, the brutal condition of the Prophet's family, and the repercussions of the war, but there are some of them that focus on events other than these. However, the battle of Karbala remains the most popular theme amongst all (Knapczyk 2014). The primary objective of Saniya Irfan's research is to trace the evolution of *marsiya* as a poetic form and to capture the semantic shifts in the metaphorical usage of its vocabulary over a period of time and with geographical advancements. These shifts or variations will then be mapped as a model of visual representations. This mapping can be done by linguistic analysis of the vocabulary used by *marsiya nigar* (*marsiya* authors) for quick comparative models, which can even be trans-generic and trans-lingual.

Following the popularity of marsiya writing, or *marsiya nigari*, in religious gatherings since the nineteenth century, commentary on prominent poets has been done in a traditional, subjective way. Celebrated works of Meer Anees (1800–1874) and Mirza Dabeer (1803–1875) are taken up by many critics, and one such famous work of critical analysis is by Shibli Nomani (1857–1914). A notable scholar of Persian and Urdu language and literature, Shibli proposes a comparative exploration of both Meer Anees and Mirza Dabeer, emphasizing the linguistic facets of their marasi (Shibli [1907] 1936).

For her Jugaad CoLab project, Sania Irfan developed a digital corpus of marsiya and tested the claims of Shibli Nomani’s subjective investigation. While conventional literary analysis can yield similar readings, computational analysis, with its capacity to unpack extremely complex patterns and to work on large scales, can open up new ways of understanding histories, cultures, texts, performances, literary form, and so on (Pue 2017). Saniya Irfan used the software Sketch Engine to test Shibli Nomani’s claims. The widely used metric of word frequency was the basis of this scrutiny. Lists of the frequent words generated from a digital corpus of marsiya by Meer Anees and Mirza Dabeer, excluding extra words, were utilized to assess the lexical repertoire in terms of rhyme and metre, metaphor or imagery as is observed from the conventional point of view.

One important challenge in this work was the creation of a digital corpus of the Urdu texts. Most of the texts were available in image format with traditional print and typography, and it was difficult to convert them into a machine-readable text. Therefore, the research sample on which this work was done was considerably smaller than it will be in the future. Saniya Irfan’s project aims to study the South Asian poetic tradition of marsiya by creating a digital archive of the OCRized texts. This archive will be editable and navigable. The corpus thus made will be used for textual analysis to trace metaphorical, archetypal, and rhythmic patterns. Now a PhD scholar at the Indian Institute of Technology Delhi, Saniya Irfan is also involved in a project on Optical Character and Hand-Written Text Recognition for Indian languages project. Her current research expands to address other marsiya nigars’ contributions in shaping this genre by the method of topic modelling in the digital archive. Lastly, the archive will also include seminal studies on marsiya that will enable the user to navigate through on the basis of a keyword. This functional objective will act as a linkage between critical works and a particular *marsiya* or marsiya nigar.

Project 5: Songs of resistance: Annotating and digitally archiving the Parcham Song Squad

Steven George’s project considers the use of digital humanities tools by oppressed communities in India, specifically Dalit communities in India—that is, groups

formerly known as “untouchable”—and expands his research to other marginalized communities. In this Jugaad Digital Poetics project, he focuses on “Songs of Resistance: Annotating and Digitally Archiving the Parcham Song Squad.” As part of the CoLab, he started work on annotating and archiving a Resistance song squad based in India, known as Parcham, a Hindi word for flag or banner (Figure 4). In this digital sound project, George has documented audio-visual interviews of the resource persons from the group, worked on digitally converting the analog audio, archiving, and annotating these audio files. As an ongoing project, he is creating various types of digitally accessible content on Parcham, including an annotated history of the same.



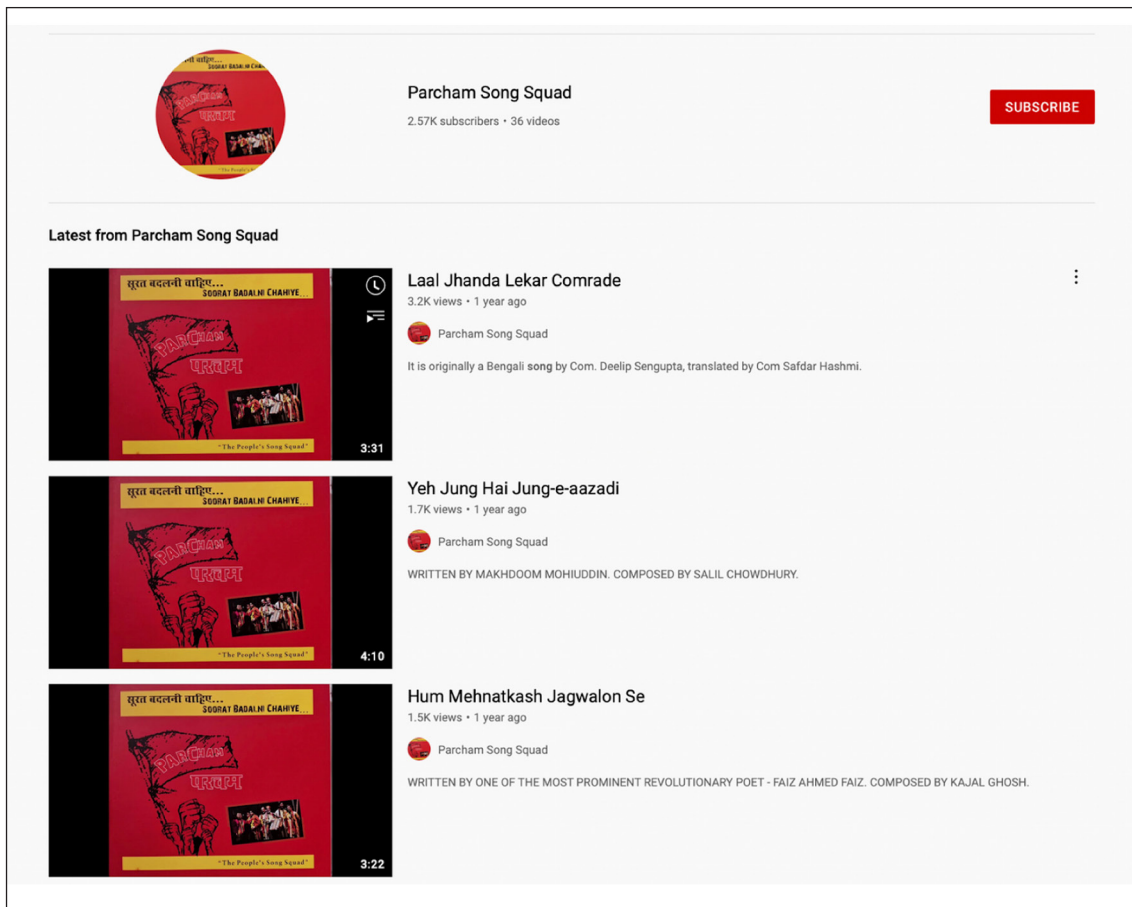
Figure 4: Image from Parcham Song Squad.

Parcham Song Squad came into existence in the years 1971–1972. The Song Squad sings songs of resistance and peace. They compose their own songs, as well as singing songs from the already existing repertoire of rich traditions of resistance songs from India and abroad. During their peak period they used to do more than 200 performances in a year. They largely worked as a cultural troupe for the Communist Party of India (Marxist). Their group has performed worldwide, in several world forums. Most of the group members have now entered their seventies. They still continue to sing songs in protests, public gatherings, and progressive–democratic events.

Despite their legacy, there is hardly anything found about them on the web. They have contributed songs to documentaries, films, and plays. For a long period, they have composed songs for Jan Natya Manch (JANAM), which is primarily a theatre group credited as pioneers of street theatre in India and is also part of the cultural wing of the Communist Party of India (Marxist). JANAM celebrated their 50th year anniversary in the year 2022. Unlike Parcham, there is significant documentation available on Janam, including two books (Ghosh 2012; Deshpande 2007). Apart from those, there are articles about them on Wikipedia, websites, newspapers, and academic journals.

The absence of information about the Parcham Song Squad online is not a strange phenomenon. Many songs and performance groups in India leave little if any documentation or historical evidence. Instead, they exist mostly in the public memory and people’s narratives. This is a gap that Kajal Ghosh, one of the members of the Song Squad, describes as a general “lack of writing and documenting practice in the performative traditions of India” (Kajal Ghosh 2021, personal communication, 8 April). Moreover, as a resistance group, whose songs and activism challenge inequalities in society, the Parcham’s Song Squad’s materials are especially at risk.

To address this gap, George started collecting their old posters and notes, audio cassettes and files, and interviews with members of the group. One of the possible solutions in this process was the digitization of these notes and other ephemera into lossless formats, converting the analog files into digital formats. Work is ongoing in the form of a YouTube playlist that is already in the public domain (Parcham Song Squad 2023). This YouTube playlist is a significant resource, as it consists of thirty-two songs sung by the group (**Figure 5**). However, there are no lyrics, context, translations, or annotations available. Due to this, only people who have contextual knowledge of the songs and groups are the primary viewing audience for the playlist. During the process of addressing the gap, George started recording conversations with and between members of Parcham. George plans for these to be captioned, subtitled, and annotated.



The image shows the YouTube channel page for 'Parcham Song Squad'. At the top left is a circular profile picture featuring a red flag with the word 'PARCHAM' and a group of people. To the right of the profile picture, the channel name 'Parcham Song Squad' is displayed, along with '2.57K subscribers · 36 videos' and a red 'SUBSCRIBE' button. Below this, the section 'Latest from Parcham Song Squad' lists three videos. Each video thumbnail features the same red flag graphic and the text 'सूरत बदलनी चाहिये... SODRAY BADA IN CHAHYE' and 'The People's Song Squad'. The first video is 'Laal Jhanda Lekar Comrade' (3:31, 3.2K views, 1 year ago), the second is 'Yeh Jung Hai Jung-e-aazadi' (4:10, 1.7K views, 1 year ago), and the third is 'Hum Mehnatkash Jagwalon Se' (3:22, 1.5K views, 1 year ago). Each video entry includes the channel name and a brief description of the song's origin and composer.

Figure 5: Image from Parcham Song Squad playlist.

These songs of resistance will be soon lost if not transformed into accessible digital content. With an upsurge of new media, many of the songs are being rendered and adapted into versions that are completely in neglect of their original formats and performance styles. The Song Squad reported to George that they believe their soul exists not in their personal legacy but in their songs. George's vision was always to archive Parcham's legacy and history through the innovative use of digital media. In this task it was crucial to annotate while archiving in order to create the metadata necessary for critical digital repositories. As he described this process to the Parcham members, George benefitted from their innovative ideas on how to approach certain texts, and in what ways one can transform them into digital editions. This ongoing project serves as a digital humanities project model for creating critical digital repositories about art groups, people, movements, and more. It is, in a way, the map of the future for a historically significant, people-oriented body of work. This experiment and the

building of relationships with the Parcham allowed Steven George to think from the perspective of the form, content, and the medium. It is a work in progress, where there are multiple new possibilities to explore.

Project 6: Documenting and remediating Waris Shah's *Heer*

Vikalp Kumar and Abdul Rahman's collaborative project focuses on Waris Shah's late eighteenth-century *Heer* and its modern remediation in a variety of intermedial forms. These include popular forms, such as oral recitations and other kinds of public text like couplets from the text circulating on paper and non-paper material. Waris Shah's *Heer* is an eighteenth-century text of the qissa genre in Punjabi. The text circulates across the geographical boundary between India and Pakistan, pervading everyday realities of people on either side of the border. It is written in two scripts, in right-to-left Shahmukhi (شاه مکھی), especially in Pakistan's Punjab, and in the left-to-right Gurmukhi (ਗੁਰਮੁਖੀ) script in the east in India's Punjab. Recitations of *Heer* are a major site of community engagement, which take place on occasions ranging from weddings, mourning, and also other non-ritual social gatherings. In addition, it is the only text other than the Quran which many semi-literate Punjabi men and women learn to read. People from all walks of life attend performances of *Heer*-gayaks (professional *Heer* singers) organized in the form of contests.

Abdul Rahman documents the oral, print, and visual culture of Waris Shah's *Heer* in Pakistan's Punjab region that complements Vikalp Kumar's annotations of this material in India. Rahman documents the *Heer*-gayak's two forms of recitations, specifically melodic and chant styles, which have not been documented sufficiently before. Though a standard sequence of musical notes in the melodic mode Raag Bhairav is employed to recite the text, a limited amount of variation is allowed, and the ability of *Heer*-gayaks to make innovations within these constraints, in addition to their ability to connect with their audience via their body language and voice modulation in descriptive prose interludes, is a marker of their skill. The entire text is never recited in a single sitting. Different *Heer*-gayaks, in addition to their skills, are known for the parts of the text they specialize in. Working with a number of *Heer*-gayaks, the project documents the diversity in recitation styles. Documenting the chant style of recitation can involve multiple readers, each of whom usually has their favourite parts. Abdul Rahman recorded the chant-style recitations in and around Malka Hans and the melodic-style in and around Lahore, Kasur, Sheikhpura, and Sahiwal—where most *Heer*-gayaks live. As it is rare to find *Heer*-gayaks below the age of 60, this documentation is also an exercise in conserving the cultural heritage of this dying art.

The print history of Waris Shah's *Heer* also lacks a descriptive bibliography, a gap to which the other team member, Vikalp Kumar, attends. *Heer* was printed extensively in a variety of scripts, including Shahmukhi, Gurmukhi, and Devanagari, before critical editions finally appeared in the twentieth century in India and Pakistan and became the most commonly available ones. For the CoLab project, Kumar has prepared a descriptive bibliography of versions of Waris Shah's *Heer* printed between 1860 and 1950. Earlier editions and versions displayed a good diversity in visual representations before becoming standardized in the latter half of the twentieth century.

Abdul Rahman collects *Heer* as a public text that circulates through inscriptions/non-paper sources and paper sources. Firstly, he collects and documents calendars, posters, paintings, and stickers representing the text with its author and its character. These are especially popular across the rural eastern parts of Pakistan's Punjab. The collection is scanned, catalogued, and tagged with descriptive metadata and produced digitally using minimal computing principles. He also documents Waris Shah's *Heer* as a public text through inscriptions/non-paper sources by photographing selections from the text which are extensively inscribed on the walls of homes, institutions, and public signage of Malka Hans. The documentation of visual culture and public text constitutes the third part of the project.

The deliverables of the fellowship—the recordings of the text in two styles (in addition to integrating them with Vikalp Kumar's work); digitized cassettes of *Heer*-gayaks; collection of book covers, illustrations, stickers, posters, and paintings; and select digitized versions and editions of Waris Shah's *Heer*—will be curated as a web exhibition in addition to depositing this curated resource into an open-access repository. Two papers—an ethnography of *Heer*'s recitation tradition and an analysis of the visual culture of the text—are also future outcomes of this project.

Conclusion

We have argued for jugaad as an appropriate form for digital humanities practice involving Indian and South Asian studies. Jugaad in this context means the combination of multiple materialities, a focus on Indian and South Asian materials across national boundaries, innovation and creation that is collective as well as individual, flexible as well as makeshift, that works under constraints and yet holds open the possibility of the disruption of conventional orders of representation. We discussed the formation of the Jugaad CoLab, its research foci, and its initial set of projects: OCR of Historical South Asian Documents, Metadata for Bengali Periodicals, Noha and Other Urdu Poetic Keywords, Digitally Mapping the Urdu Marsiya, Songs of Resistance: Annotating and

Digitally Archiving the Parcham Song Squad, and Documenting and Remediating Waris Shah's *Heer*.

Jugaad CoLab projects benefit from the combination of multiple materialities associated with jugaad-as-method. This is most obvious in the text-based projects through digitization using OCR, which presents a workflow that uses free and open-source technologies, brought together from various sources. The potential of this project to increase the accessibility of the historical archive of Indian and South Asian language materials is enormous. The Bengali metadata project also directly contributes to the increase in access to historical materials. Bringing multiple materialities together, the Punjabi *Heer* project combines text, recitation, and public visual culture; the Urdu *noha* and *marsiya* projects join textual study with a consideration of digital forms of textuality and recitations; and the Parcham Song Squad project involves audio, visual, and textual materials, as well as ephemera.

These projects also demonstrate the need for a broad regionality when approaching Indian and South Asian materials. Urdu and Punjabi are both spoken languages in India and Pakistan. Bengali is spoken in both India and Bangladesh. A national model of digital humanities research, therefore, is clearly insufficient. However, we note that there are certain commonalities in the predicament of non-Anglophone cultural materials in the region such that, even as it is broad and capacious, grounded in position rather than identity, a regional understanding does have some utility. The OCR technologies developed for major languages and scripts, for example, can be adapted to the region's many underserved minority languages, as well. We also emphasize the possibilities of SAARC as an existing model for Indian and South Asian collaboration, while supplementing this post-colonial political alliance with the cultural history of the diaspora and, in considering our scholarly community, the global community of academics and citizen scholars engaged with this part of the human cultural record.

Instead of merely makeshift and individual, the projects demonstrate that jugaad can be collective and flexible as well. The *Heer* project demonstrates the benefits of transnational collaboration, addressing Punjabi culture in both India and Pakistan. The keywords project brings multiple scholars into collaboration in documenting Urdu's poetics, including in its various digital forms. The Bengali metadata involves students as credited data editors while also empowering their own research. The OCR project suggests and requires collaborative effort, especially as it is extended to other languages. The methods and workflows are flexible by design and can be repurposed for other projects.

That flexibility derives from these projects' use of tools that are free and preferably open source, with attention paid to developing for them an interface accessible without extensive technical expertise. There is a consideration of the constraints faced by the larger research community. These projects consider how to make digital humanities tools available without access to commercial products, expensive laptops, or high-performance computing. While we have embraced minimal computing, we have nothing against high performance computing and use it when necessary, as in the OCR project. However, we consider our research community to include the typical humanities graduate student in South Asia, who surely has a cellphone, perhaps an older laptop, and maybe but not necessarily some command of English. Designing for this audience, as in the Bengali metadata project, requires reducing the technical stack necessary to participate in digital humanities research while also providing enough training to allow meaningful participation.

The final component of *jugaad* in digital humanities that we have identified, the disruption of orders of representation often through juxtaposition, involves a combination of the affordances offered through remediation. The project on the Urdu *noha* elucidates this process, but so too do the remediations of *Heer*, the Parcham Song Squad, and the Urdu *marasi*. Those projects also reflect a larger movement within Indian digital humanities related to literary cultures to move beyond the traditional genre of “the book” to consider digital forms of textuality (cf. Zaidi and Pue 2023). These projects, as well as the Bangla metadata project and the OCR project more generally, hold the potential, through search, of offering new insights into the literary and cultural past that destabilize accepted forms of representation, including the ordering of genres and languages, as in the Parcham Sound Squad project. What the Bengali metadata project, as well as the OCR project more generally, opens up is the possibility of comparative analysis across languages, something that remains quite rare in Indian and South Asian studies, despite the interconnections of language communities; instead, languages appear as isolated silos, their literary histories being told as individual and separate tales of the rise and fall of language communities. By contrast, digital remediation allows for the sort of interlingual and intraregional juxtapositions that hold open the possibility of destabilizing conventional narratives and understandings.

Our emphasis on adapting free and open-source tools to address both computational analysis and digital preservation provides a model that can be adapted by future researchers. Indeed, the research areas of this affinity group—decolonial poetics, minimal computing, and pirate care—attract wide-ranging support, including academics based at public universities and communities of citizen scholars. We have

settled on a university-based lab as the current infrastructure for the Jugaad CoLab, in part to counter the difficulties of collaboration across South Asian nations. We still face serious challenges: how to facilitate multi-institutional and international collaboration; how to arrange for sustained funding to support further research; how to arrange for equipment sharing; and so on. Future jugaad will be necessary to address these issues. The devastation wrought by the pandemic has deepened our own resolve to preserve and analyze the cultural record of India and South Asia and its diaspora. The outcomes of these summer research projects, as well as our ongoing collaborations, we hope, will inspire further initiatives in research and pedagogy, and we look forward to further collaborations with emerging scholars and other collectives in the future.

Competing interests

The authors have no competing interests to declare.

Contributions

Authorial

Authorship in the byline is by contribution level. Author contributions, described using the NISO (National Information Standards Organization) CrediT taxonomy, are as follows:

Author name and initials:

Rohan Chauhan: RC
 Agniv Ghosh: AG
 Zahra Rizvi: ZR
 Saniya Irfan: SI
 Steven S George: SSG
 Abdul Rehman: AR
 Vikalp Kumar: VK
 A. Sean Pue: ASP

Authors are listed in descending order by significance of contribution. The corresponding author is ASP.

Conceptualization: RC, AG, ZR, SI, SSG, AR, VK, ASP
 Data Curation: RC, AG, ZR, SI, SSG, AR, VK, ASP
 Formal Analysis: RC, AG, ZR, SI, SSG, AR, VK, ASP
 Funding Acquisition: ASP
 Investigation: RC, AG, ZR, SI, SSG, AR, VK, ASP
 Methodology: RC, AG, ZR, SI, SSG, AR, VK, ASP
 Writing – Original Draft: RC, AG, ZR, SI, SSG, AR, VK, ASP
 Writing – Review & Editing: ASP, RC, AG, ZR, SI, SSG, AR, VK

Editorial

Special Issue Editors

Roopika Risam, Dartmouth College, United States
 Barbara Bordalejo, University of Lethbridge, Canada
 Emmanuel Château-Dutier, University of Montreal, Canada

Copy Editors

Morgan Pearce, The Journal Incubator, University of Lethbridge, Canada
 Christa Avram, The Journal Incubator, University of Lethbridge, Canada

Layout Editor

A K M Iftekhhar Khalid, The Journal Incubator, University of Lethbridge, Canada

References

- Bichitra. 2023. "Bichitra: The Online Tagore Variorum." *School of Cultural Texts and Records*. Jadavpur University. Accessed August 14. <http://bichitra.jdvu.ac.in>.
- Birtchnell, Thomas. 2011. "Jugaad as Systemic Risk and Disruptive Innovation in India." *Contemporary South Asia* 19(4), 357–372. DOI: <https://doi.org/10.1080/09584935.2011.569702>.
- Bode, Katherine. 2018. *A World of Fiction: Digital Collections and the Future of Literary History*. Ann Arbor: University of Michigan Press. DOI: <https://doi.org/10.3998/mpub.8784777>.
- Chaudhuri, Sukanta, ed. 2015. *Bichitra: The Making of an Online Tagore Variorum*. Heidelberg: Springer. DOI: <http://doi.org/10.1007/978-3-319-23678-0>.
- Clérice, Thibault. 2022. "You Actually Look Twice At it (YALTAi): Using an Object Detection Approach Instead of Region Segmentation within the Kraken Engine." *arXiv*. arXiv:2207.11230. DOI: <https://doi.org/10.48550/arXiv.2207.11230>.
- Cordell, Ryan. 2016. "How Not to Teach Digital Humanities." In *Debates in the Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein. Minneapolis: University of Minnesota Press. DOI: <https://doi.org/10.5749/9781452963761>.
- Department of Information Technology & Electronics. 2023. "Home." Government of West Bengal, India. Accessed August 14. <https://itewb.gov.in>.
- Deshpande, Sudhanva, ed. 2007. *Theatre of the Streets: The Jana Natya Manch Experience*. New Delhi: Janam.
- Dodd, Maya. 2023. "Digital Culture in India: Digitality and Its Discontents." In *Literary Cultures and Digital Humanities in India*, edited by Nishat Zaidi and A. Sean Pue, 19–37. New York: Routledge. DOI: <https://doi.org/10.4324/9781003354246>.
- Gabay, Simon, Jean-Baptiste Camps, Ariane Pinche, and Nicola Carboni. 2021. "SegmOnto [version 0.9]." *GitHub*. Accessed July 4, 2023. <https://github.com/SegmOnto>.
- Ghosh, Arjun. 2012. *A History of the Jan Natya Manch: Plays for the People*. New Delhi: Sage.
- Grallert, Till. 2022. "Open Arabic Periodical Editions: A Framework for Bootstrapped Scholarly Editions Outside the Global North." *Digital Humanities Quarterly* 16(2). Accessed August 4, 2023. <https://digitalhumanities.org/dhq/vol/16/2/000593/000593.html>.
- Graziano, Valeria, Marcell Mars, and Tomislav Medak. 2019. "The Pirate Care Project." *Piratecare*. Centre for Postdigital Cultures, Coventry University. Accessed July 1, 2023. <https://pirate.care/pages/concept/>.
- Hall, Stuart, ed. 1997. *Representation: Cultural Representations and Signifying Practices*. London: Sage.
- Iqbal Academy. 2023. "Iqbal Cyber Library." Accessed August 14. <http://iqbalcyberlibrary.net>.
- Janes, Juliette, Ariane Pinche, Claire Jahan, and Simon Gabay. 2021. "Towards Automatic TEI Encoding via Layout Analysis." Paper presented at Fantastic Future '21: 3rd International Conference on Artificial Intelligence for Libraries, Archives, and Museums (ai4lam), Paris, France. Accessed July 1, 2023. <https://hal.science/hal-03527287>.

- Kiessling, Benjamin. 2019. "Kraken—A Universal Text Recognizer for the Humanities." Paper presented at Digital Humanities Conference (DH2019), Utrecht, Netherlands. DOI: <https://doi.org/10.34894/Z9G2EX>.
- Knapczyk, Peter Andrew. 2014. "Crafting the Cosmopolitan Elegy in North India: Poets, Patrons, and the Urdu Marsiyah, 1707–1857." PhD diss. University of Texas at Austin. Accessed August 22, 2023. <http://hdl.handle.net/2152/44099>.
- Lassner, David, Clemens Neudecker, Julius Coburger, and Anne Baillot. 2021. "Publishing an OCR Ground Truth Data Set for Reuse in an Unclear Copyright Setting. Two Case Studies with Legal and Technical Solutions to Enable a Collective OCR Ground Truth Data Set Effort." *Zeitschrift für digitale Geisteswissenschaften* (Sonderbände 5). DOI: https://doi.org/10.17175/sb005_006.
- McKinney, Wes. 2010. "Data Structures for Statistical Computing in Python." In *Proceedings of the 9th Python in Science Conference (SciPy)*, 445(1), 51–56. DOI: <https://doi.org/10.25080/Majora-92bf1922-00a>.
- Mukherjee, Souvik. 2020. "Digital Humanities, or What You Will: Bringing DH to Indian Classrooms." In *Exploring Digital Humanities in India: Pedagogies, Practices, and Institutional Possibilities*, edited by Maya Dodd and Nidhi Kalra, 105–123. New Delhi: Routledge. DOI: <https://doi.org/10.4324/9781003052302>.
- Nicholson, Reynold Alleyne. (1907) 1969. *A Literary History of the Arabs*. New York: Charles Scribner's Sons. Reprint, London: Syndics of the Cambridge University Press.
- Parcham Song Squad. 2023. "Playlists." YouTube. Accessed August 22. <https://www.youtube.com/@parchamsongsquad6694/playlists>.
- Patel, Geeta. 2017. *Risky Bodies & Techno-Intimacy: Reflections on Sexuality, Media, Science, Finance*. Seattle: University of Washington.
- Pinault, David. 1997. "Shiism in South Asia." *The Muslim World* 87(3–4): 234–257. DOI: <https://doi.org/10.1111/j.1478-1913.1997.tb03638.x>.
- Pinche, Ariane, Kelly Christensen, and Simon Gabay. 2022. "Between Automatic and Manual Encoding: Towards a Generic TEI Model for Historical Prints and Manuscripts." Paper presented at Text Encoding Initiative (TEI 2022), Newcastle, United Kingdom. DOI: <https://doi.org/10.5281/zenodo.7092214>.
- Pritchett, Frances. 1994. *Nets of Awareness: Urdu Poetry and Its Critics*. Berkeley: University of California Press.
- . 2006. "Urdu: Some Thoughts about the Script and Grammar, and Other General Notes for Students." *Urdu Language-Learning Resources*. Columbia University. Accessed July 4, 2023. <http://www.columbia.edu/itc/mealac/pritchett/00Urdu/urduascript/>.
- Pue, A. Sean. 2017. "Acoustic Traces of Poetry in South Asia." *South Asian Review* 40(3), 221–236. DOI: <https://doi.org/10.1080/02759527.2019.1599561>.
- Pue, A. Sean, Nishat Zaidi, and Mohd Aqib. 2022. *Dreaming of the Digital Divan: Digital Apprehensions of Poetry in Indian Languages*. New Delhi: Bloomsbury.
- Rajan, Vinodh (virtualvinodh). 2021. "Aksharamukha [software]." *GitHub*. Accessed July 4, 2023. <https://github.com/virtualvinodh/aksharamukha>.

- Rajdou, Navi, Jaideep Prabhu, and Simone Ahuja. 2012. *Jugaad Innovation: Think Frugal, Be Flexible, Generate Breakthrough Growth*. London: Jossey-Bass.
- Ray Murray, Padmini, and Chris Hand. 2015. "Making Culture: Locating the Digital Humanities in India." In *Critical Making: Design and the Digital Humanities* [special issue], edited by Jessica Barne and Amy Papaelias, *Visible Language* 49(3): 140–155. Accessed August 22, 2023. <https://journals.uc.edu/index.php/vl/article/view/5910>.
- Reul, Christian, Dennis Christ, Alexander Hartelt, Nico Balbach, Maximilian Wehner, Uwe Springmann, Christoph Wick, Christine Grundig, Andreas Büttner, and Frank Puppe. 2019. "OCR4all—An Open-Source Tool Providing a (Semi-)Automatic OCR Workflow for Historical Printings." *Applied Sciences* 9(22): 4853. DOI: <https://doi.org/10.3390/app9224853>.
- Risam, Roopika. 2019. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Evanston, IL: Northwestern University Press. DOI: <https://doi.org/10.2307/j.ctv7tq4hg>.
- Risam, Roopika, and Alex Gil. 2022. "Introduction: The Questions of Minimal Computing." *Digital Humanities Quarterly* 16(2). Accessed July 4, 2023. <http://www.digitalhumanities.org/dhq/vol/16/2/000646/000646.html>.
- Rizvi, Zahra, Rohan Chauhan, A. Sean Pue, and Nishat Zaidi. 2022. "Minimal Computing for Exploring Indian Poetics." *Digital Humanities Quarterly* 16(2). Accessed July 4, 2023. <http://www.digitalhumanities.org/dhq/vol/16/2/000589/000589.html>.
- Roy, Jibendra Sinha. 1960. *Kalloler Kal: Kallol, Kali-Kalam, Pragatir Din*. Kolkota: Dey's Publishing.
- Sagar, Rahul. 2018. "The Periodicals." *Ideas of India*. Accessed July 4, 2023. <https://www.ideasofindia.org/>.
- Said, Edward W. 1979. *Orientalism*. New York: Vintage.
- Sehksaria, Pankaj. 2013. "The Making of an Indigenous Scanning Tunneling Microscope." *Current Science* 104(9), 1152–1158. Accessed July 4, 2022. <http://www.jstor.org/stable/24092394>.
- Shibli, Maulana Nomani. (1907) 1936. *Mawazina-e-Anees-o-Dabeer*. Allahabad: Lala Ram Narayan Lal Bookseller.
- Smith, David A., and Ryan Cordell. 2018. "A Research Agenda for Historical and Multilingual Optical Character Recognition." *DRS (Digital Repository Service)*. Northeastern University. Accessed July 4, 2022. <http://hdl.handle.net/2047/D20297452>.
- SPARC (Scheme for Promotion of Academic and Research Collaboration). 2023. "Home." Indian Institute of Technology Kharagpur. Accessed July 7. <https://sparc.iitkgp.ac.in/index.php>.
- Spivak, Gayatri. 2008. *Other Asias*. London: Blackwell.
- Springman, Uwe, and Anke Lüdeling. 2017. "OCR of Historical Printings with an Application to Building Diachronic Corpora: A Case Study Using the RIDGES Herbal Corpus." *Digital Humanities Quarterly* 11(2). Accessed July 4, 2023. <http://www.digitalhumanities.org/dhq/vol/11/2/000288/000288.html>.
- Stokes, Peter A., Benjamin Kiessling, Daniel Stökel Ben Ezra, Robin Tissot, and El Hassne Gargem. 2021. "The eScriptorium VRE for Manuscript Cultures." *Classics@* 18(1). Accessed July 4, 2023. <https://classics-at.chs.harvard.edu/classics18-stokes-kiessling-stokl-ben-ezra-tissot-garge>.

Tanvir, Kuhu, and Ramna Walia. 2021. "Remix." *BioScope: South Asian Screen Studies* 12(1-2). DOI: <https://doi.org/10.1177/09749276211026438>.

Thangavel, Shanmugapriya, and Nirmala Menon. 2020. "Infrastructure and Social Interaction: Situated Research Practices in Digital Humanities in India." *Digital Humanities Quarterly* 14(3). Accessed July 4, 2023. <http://digitalhumanities.org:8081/dhq/vol/14/3/000471/000471.html>.

Underwood, Ted. 2014. "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago." *Representations*, 127(1), 64–72. DOI: <https://doi.org/10.1525/rep.2014.127.1.64>.

Wikisource. 2023. "Main Page." Wikimedia Foundation. Accessed August 14. <https://wikisource.org>.

Zaidi, Nishat, and A. Sean Pue, eds. 2023. *Literary Cultures and Digital Humanities in India*. New York: Routledge. DOI: <https://doi.org/10.4324/9781003354246>.

