

Preserving complex digital objects in the GLAM community through Digital Humanities: A study on Ancient Indian scripts

Ashwin Kumar Kushwaha

Department of Library and Information Science, Banaras Hindu University, Varanasi, India.

E-mail address: akushwaha272@gmail.com

Ajay Pratap Singh

Department of Library and Information Science, Banaras Hindu University, Varanasi, India.

E-mail address: apsingh_73@yahoo.co.in



Copyright © 2020 by **Ashwin Kumar Kushwaha & Ajay Pratap Singh**. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

GLAM community of any country plays an important role in the preservation and conservation of cultural heritage wealth of that nation. India – a country full of diversity incorporates rich cultural heritage in the form of manuscripts written in various ancient scripts. These scripts reveal a rich past of the country full of information and knowledge. But in present, it is not easy to retrieve this information and knowledge. Although Government of India has initiated a project- National Mission for Manuscript (NMM) for digitizing all Indian manuscripts. But this project is not able to ensure all round support regarding dissemination of thought content, present in such manuscripts. With these issues, we got an idea to present these manuscripts among their users in such a way, they may be found more informative, more easily accessible and retrievable through their presentation in the form of 'complex digital media'. These complex digital media counterparts of manuscripts have been presented using 'Juxta - an open-source tool', which provides a platform for preserving these manuscripts for the future.

In this paper, we have shared our experiences about the issues and challenges faced while applying Digital Humanities (DH) tool. The paper also discusses the sustainability of this DH tool for future which guarantees the preservation of the manuscripts. Manuscripts collection of Indian libraries has been deployed for performing experimentations with 'Juxta'. The 'Juxta' platform was customized according to our specific needs. Results obtained were surprising as after this customization, the deployed collection was less susceptible to its obsolescence. These results were achieved by comparing its technicality background before and after customization in Juxta and also in its absence. The impact of this research is very much positive as it has presented a new dimension for the ancient collection in Indian rather Asian libraries or libraries with similar collections worldwide.

Keywords: Digital preservation, Collation software, Brahmi script, Digital Humanities, Juxta.

Introduction

Preserving ancient knowledge is important for better and sustainable future, GLAM community is best suited for performing this task. According to UNESCO¹ “Digital preservation consists of the processes aimed at ensuring the continued accessibility of digital materials. To do this involves finding ways to re-present what was originally presented to users by a combination of software and hardware tools acting on data.” India is a country full of diversity with rich cultural heritage in the form of manuscripts and inscription written in Ancient Indian scripts. For promoting this national treasure, Government of India has initiated many projects, among them the most prominent project is National Mission for Manuscripts (NMM), started in the year 2009. But the outcomes of all digital initiatives of NMM are not as satisfactory as these were previously expected, and the resultant is very unlikely for the scholarly community. Therefore, with this research gap, we started working for making digital items/objects more usable in terms of research with the help of JUXTA – a collation software. Before heading towards the JUXTA let us understand the concept of following terms:

Complex Digital Objects: “Complex digital objects are discrete digital objects, made by combining a number of other digital objects, such as Web sites.”²

Digital Humanities: “The digital humanities (DH) cannot be easily defined. Many view DH as a movement within the traditional humanities and social science disciplines, which promises to bring digital technologies to bear on traditional research questions. The same questions that once required a lifetime of manual gathering and processing of data may now be answered within a few weeks, or even a couple days, with the aid of digitised information.”³

Background

Since antiquity scholars have been working in the important field of textual criticism and the foundation of this textual criticism is based on the preparation of its critical edition which can be presented in its improved version. This textual criticism is generally done by collecting textual witnesses and their comparison among themselves. In such cases collation is the main method of comparing witnesses.⁴ According to Merriam-Webster Dictionary ‘Collate’ means “to collect and compare carefully in order to verify, and often to integrate or arrange in order”.⁵

Either manuscripts or printed texts, both of them require precision for the preparation of collation, and these must be further accompanied by accurate transcription of images.⁴ Previously, this type of work was done manually, and scholars used to spend years to find out the ways to lighten such work as on one hand knowledge, precision and skill are needed and on the other hand it requires a bit of drudgery. The integration of computer for collating textual witnesses was not new.⁶ In 1960s first attempt was made,⁶ which was followed by the development of some computerized tools like ‘Collate’. Later on, many other tools have been developed, some of them are ‘CollateX’⁷, ‘The versioning Machine’⁸ and ‘JUXTA’⁹ etc.

JUXTA

Applied Research in Patacriticism group created Juxta at the University of Virginia. The group was led by Jerome McGann¹⁰, and Juxta was developed by Nick Laiacona⁴. Juxta is currently maintained by the NINES group at the University of Virginia.¹¹

According to Juxta User's Manual¹¹, "Juxta is an open-source cross-platform tool for comparing and collating multiple witnesses to a single textual work. The software allows you to set any of the witnesses as the base text, to add or remove witness texts, to switch the base text at will, and to annotate Juxta-revealed comparisons and save the results." It is a java-based application and can run on any personal computer. The installer of this application can be downloaded from <http://www.juxtaoftware.org/download.html>.

The collation in Juxta software can be created in two simple steps. The first step is to add witness documents to the Juxta software. To do this task, firstly open 'New Comparison Set' from the File menu or press 'Ctrl+N' button (Figure 1). In this 'New Comparison Set', we can import new witness document through the 'Add Document' button in the Toolbar. Witness document to be imported, should be either in XML or TXT format. While uploading XML file we have to select XML parsing template from the given 'Select Parse Template' dialog box, by default 'juxta-default' parsing template, is selected. Other parsing templates are TEI and TEI.2 and other customized templates can also be defined according to the need.

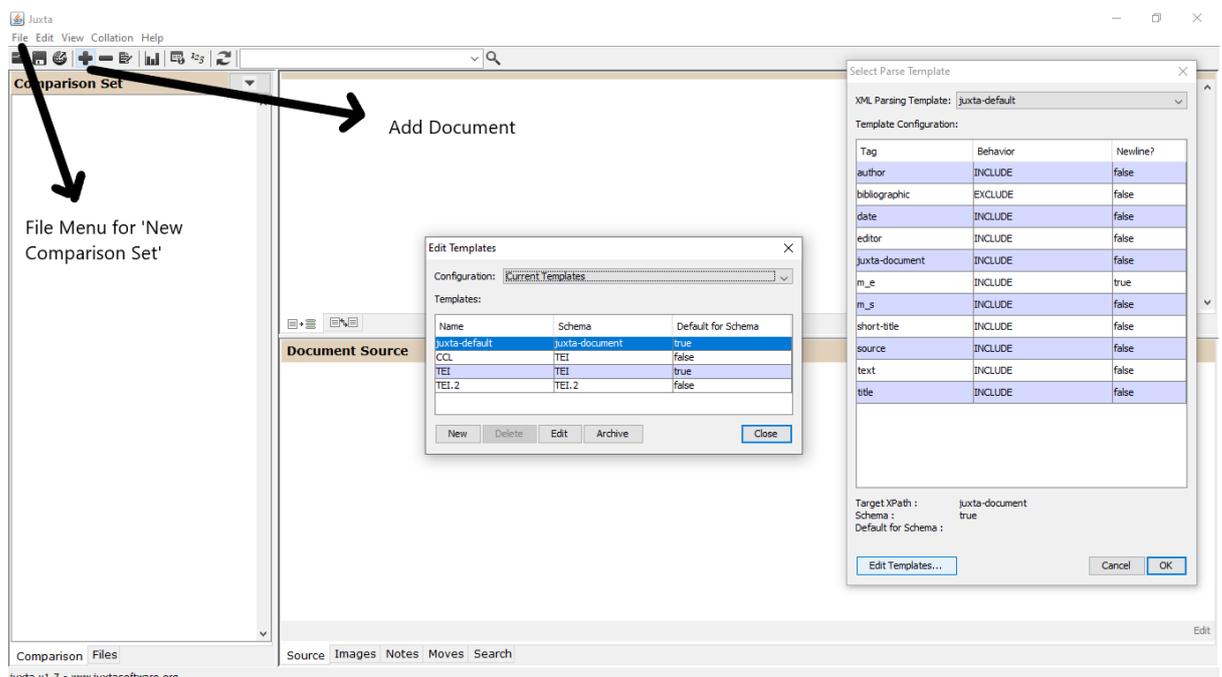


Figure 1: Adding document witness

The second step after importing the witness is to edit the source files. All the imported witnesses are displayed in the 'Comparison Explorer', situated on the left-hand side (Figure 2). The base document from which all the comparisons are made, is highlighted in the green colour and the meter on the right of each document indicates the level of variation with respect to base document. To the right of 'Comparison Explorer' is 'Document Panel', which displays a side-by-side comparison of two documents as well as collation results of multiple document witnesses. Collation view and Comparison view are the two modes present in the 'Document Panel' to facilitate side-by-side comparison and collation results. These two modes can be toggled through the buttons present at the bottom of the panel. 'Secondary Panel' is situated at the bottom of the screen which supports relevant information associated with the document visible in the 'Document Panel' at present. It has five modes: Source, Images, Notes, Moves and Search; which can be toggled by using buttons located at the bottom of the panel.

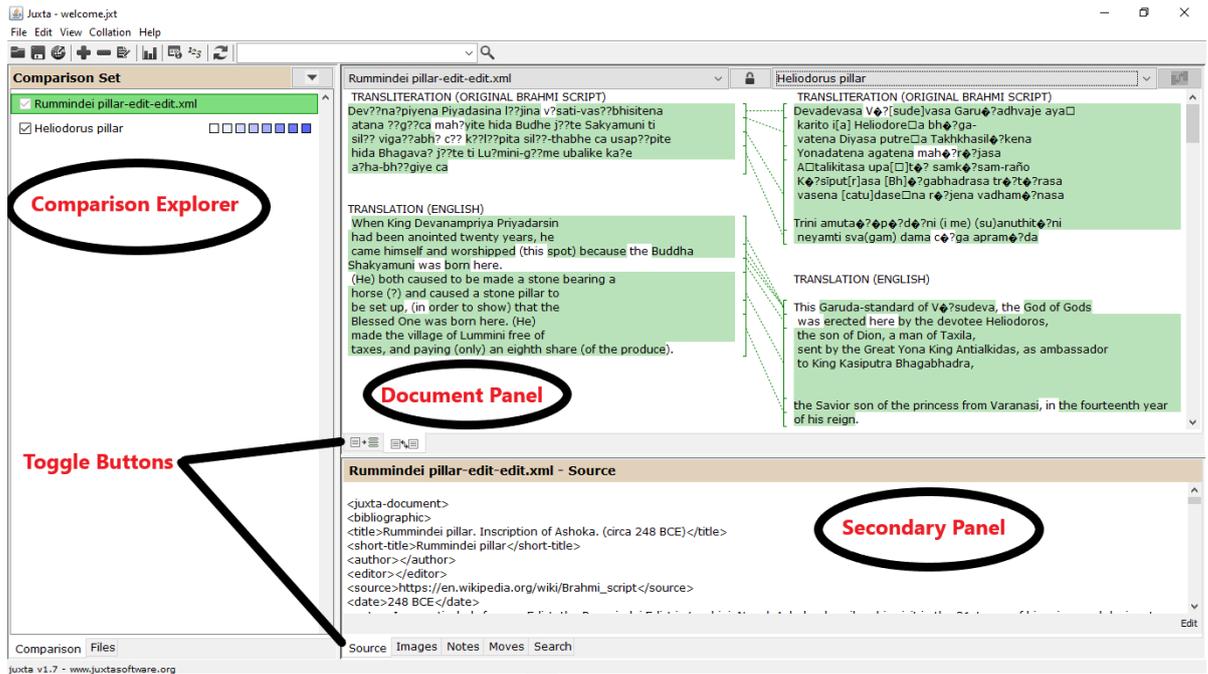


Figure 2: Editing Source document witness

Once the collation of the document is created, it provides a number of different ways for presenting the differences between the textual witnesses. In Juxta, it can be presented in three different ways:

1. Heat Map: Textual variant's heat map permit you to discover all the variations between the witness text and the base text at any level of textual unit.¹¹
2. The side-by-side view: This view permits you to inspect any two witnesses/texts at a time.⁴
3. The Histogram: This view visualizes the density of all variation with respect to the base text. Documents which consists of enormous amount of text, this histogram act as a handy finding tool for locating variants in these documents.¹¹

Further, Juxta has many advanced features which can be used for analysis and research in born-digital text and XML text.⁴ For advanced features and its uses, we can check them all in the Juxta blog.¹²

Sample

I will illustrate the potential use of Juxta with two samples. The first is the textual comparison of 'The Rummindei pillar edict in Lumbini' with its customized and non-customized text version in Juxta. The second is the similar comparison of the customized and non-customized text in Juxta of 'Heliodorus pillar rubbing (inverted colors)'.

Sample 1

For the experiment, I chose an image of 'The Rummindei pillar edict in Lumbini' (248 BCE) (Figure 3). This pillar edict describes as: The king Ashoka visits Lumbini, Nepal and designates it as the birthplace of Buddha. Further, it describes his visit to Lumbini. In the historical records

it was first time that “Sakyamuni (Sage of the Shakyas)” epithet was used to narrate the Buddha.^{13,14}

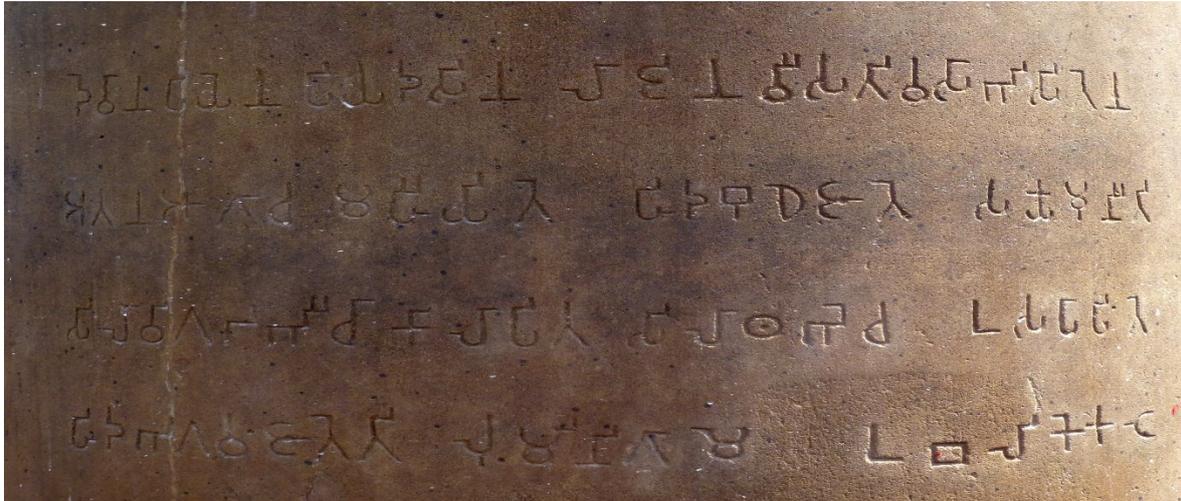


Figure 3: The Rummindei pillar edict in Lumbini¹⁵ (Document in the absence of Juxta)

Juxta software imports witness in TXT and XML file format. So, an XML file was created with the help of Notepad in which the transliteration of original Brahmi script and its English translation was included. The XML parser used for the purpose was ‘juxta-default’, XML version used was ‘XML 1.0’ and text encoding used was ‘UTF-8’. The output obtained after importing the above XML file as a witness is shown in Figure 4.

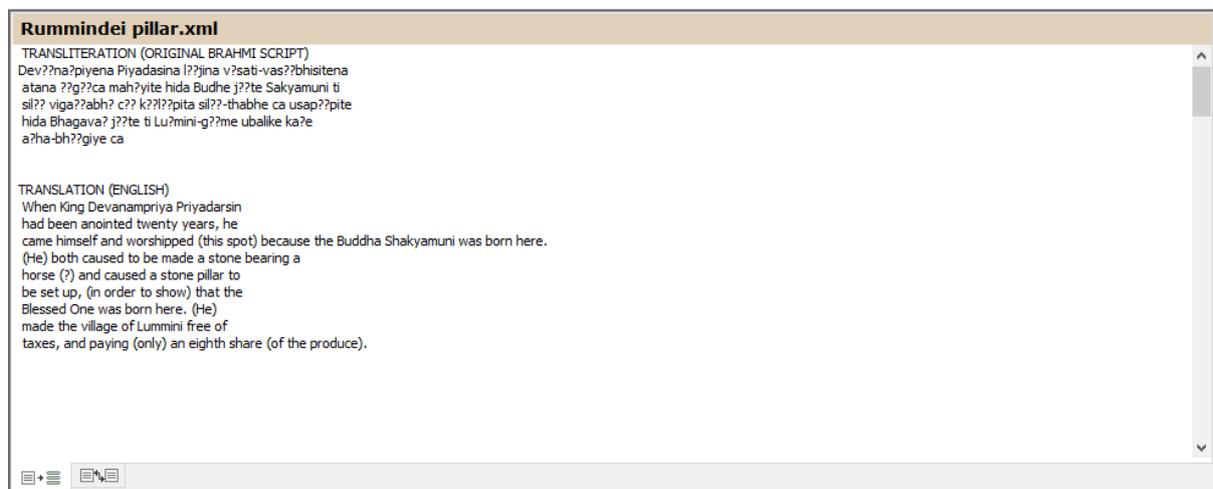


Figure 4: Witness imported to Juxta for Rummendei (Document before customization in Juxta)

Now to make witness more fruitful, an XML file was further edited to associate text images and milestones to the text. Here milestone refers to the association of an image with the desired text i.e. when the text is scrolled, the associated image is also moved and that part of image is displayed in the ‘Document Panel’. The output is displayed in Figure 5.

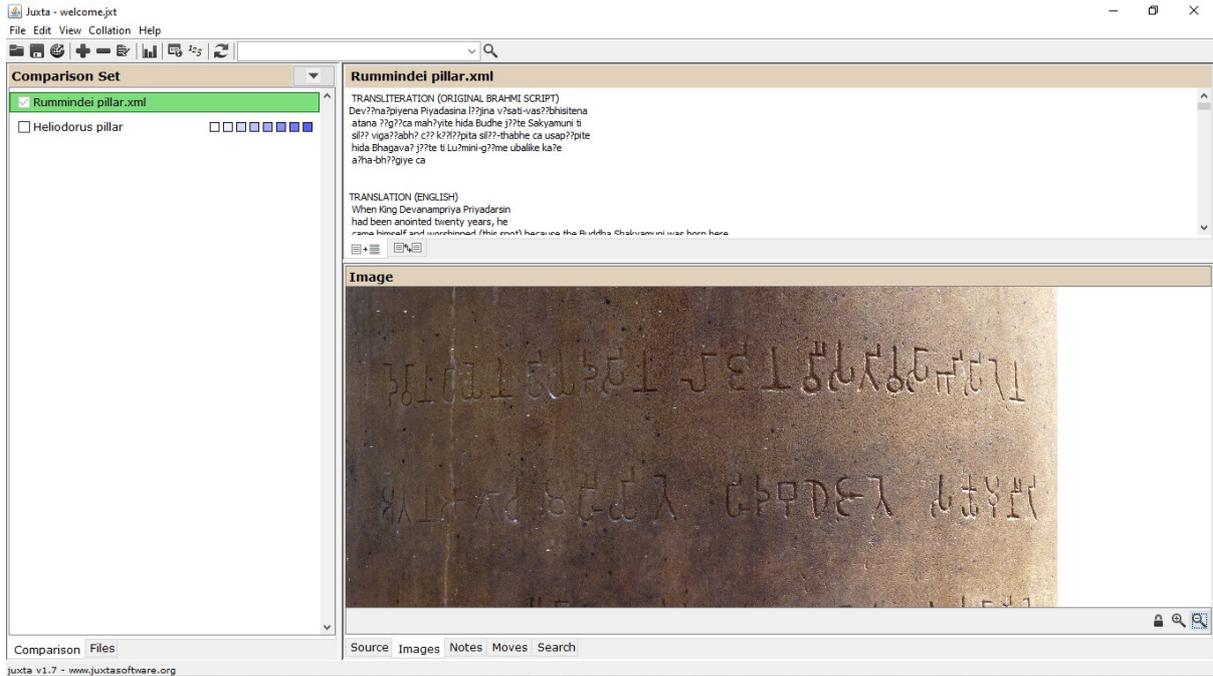


Figure 5: Association of a Rummendei image with text (Document after partial customization in Juxta)

Despite this customization, it is revealed that some issues related to the text encoding, still exist. Although the Brahmi script character is ‘UTF-8’ compatible (Unicode range U+11000–U+1107F), but some of its characters were read as garbage by the Juxta software. So, for this purpose manual editing was done in its source code to incorporate such characters in the Juxta software. The output was displayed in Figure 6

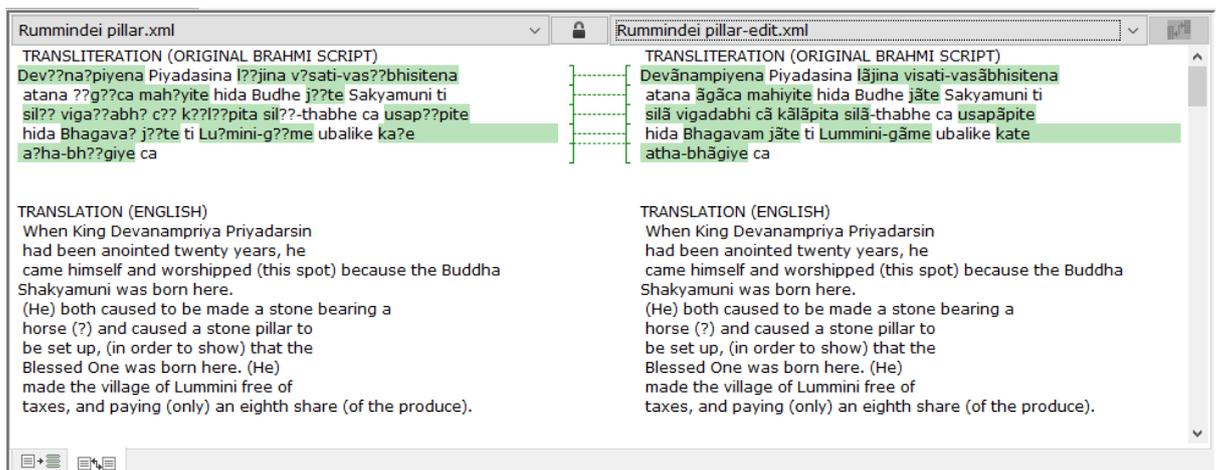


Figure 6: Manual text encoding in Rummendei (Document after customization in Juxta)

Sample 2

Another sample for the study was a stone column named as ‘The Heliodorus pillar’ (Figure 7) in central India (Vidisha, India) build around 113 BCE.¹⁶ It was erected by Heliodorus - an ambassador of an Indo-Greek king Antialcidas in Taxila¹⁷, who was sent for the court of the Shunga king Bhagabhadra. With reference to Vaishnavism in India, it was the earliest known inscription of Indian history.^{14,16,18,19}

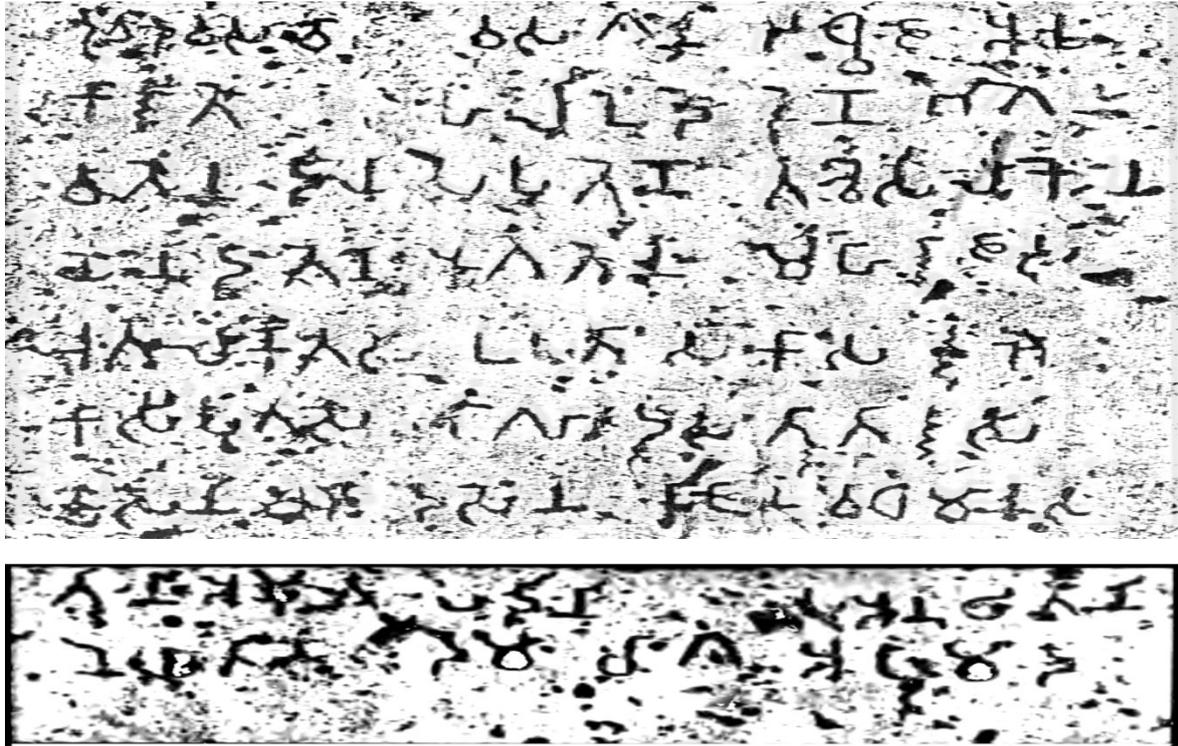


Figure 7: Heliodorus pillar rubbing (inverted colors)²⁰
(Document in the absence of Juxta)

Similar to sample 1, witness was imported with the same file format, XML parser and text encoding in the Juxta software and its output is displayed in Figure 8. Similar to Sample 1, after partial customization, results are displayed in Figure 9 and customization after the text encoding was displayed in Figure 10.

Heliodorus pillar

TRANSLITERATION (ORIGINAL BRAHMI SCRIPT)

Devadevasa V[sude]vasa Garu[adhvaje ayam]
karito [a] Heliodoreṇa bh[ga-
vatena Diyasa putreṇa Takḥkhasil[kena
Yonadatenā agatena mah[ra]jasa
Amṭalkitasa upa[m]t[sa]m[ra]ṇo
K[asiput]ra[Bh]gabhadrāsa tr[ir]asa
vasena [catu]daseṃna r[aj]ena vadham[na]sa

Trini amuta[ra]d[ra]ni (me) (su)anuthit[ra]ni
neyamti sva(gam) dama c[ra]ga apram[ra]da

TRANSLATION (ENGLISH)

This Garuda-standard of V[sude]va, the God of Gods
was erected here by the devotee Heliodoros,
the son of Dion, a man of Taxila,
sent by the Great Yona King Antialkidas, as ambassador
to King Kasiputra Bhagabhadrā,

the Savior son of the princess from Varanasi, in the fourteenth year of his reign.

Figure 8 Witness imported to Juxta for Heliodorus
(Document before customization in Juxta)

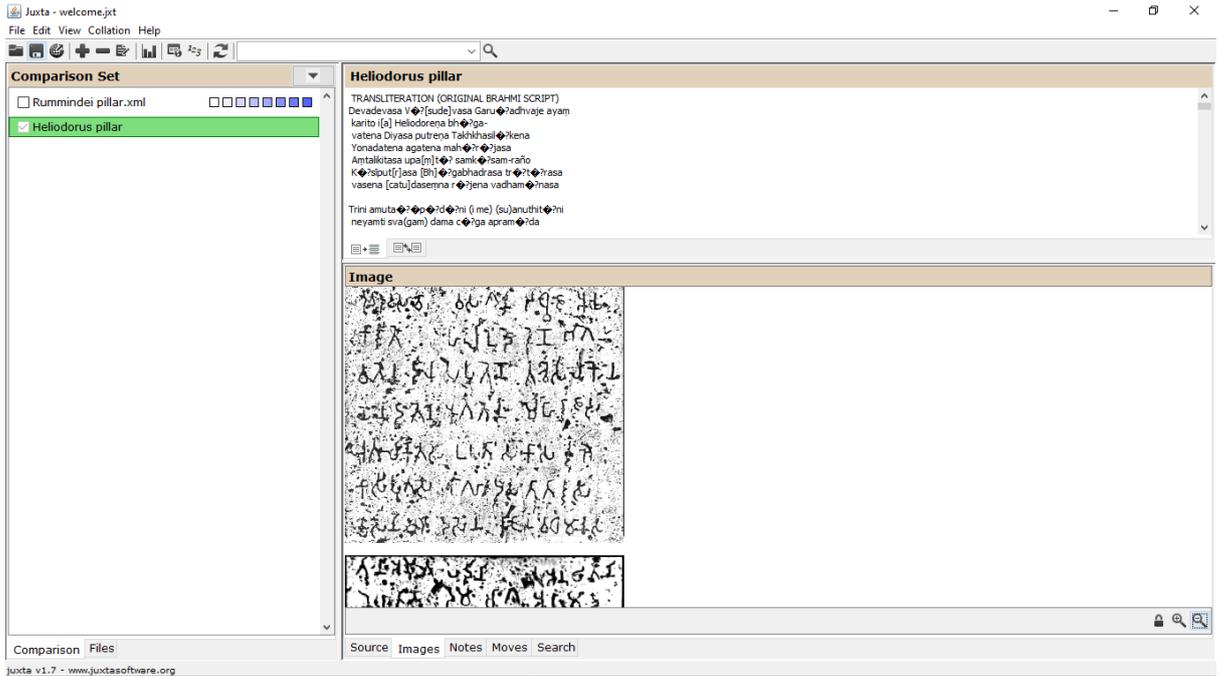


Figure 9 Association of a Heliodorus image with text (Document after partial customization in Juxta)

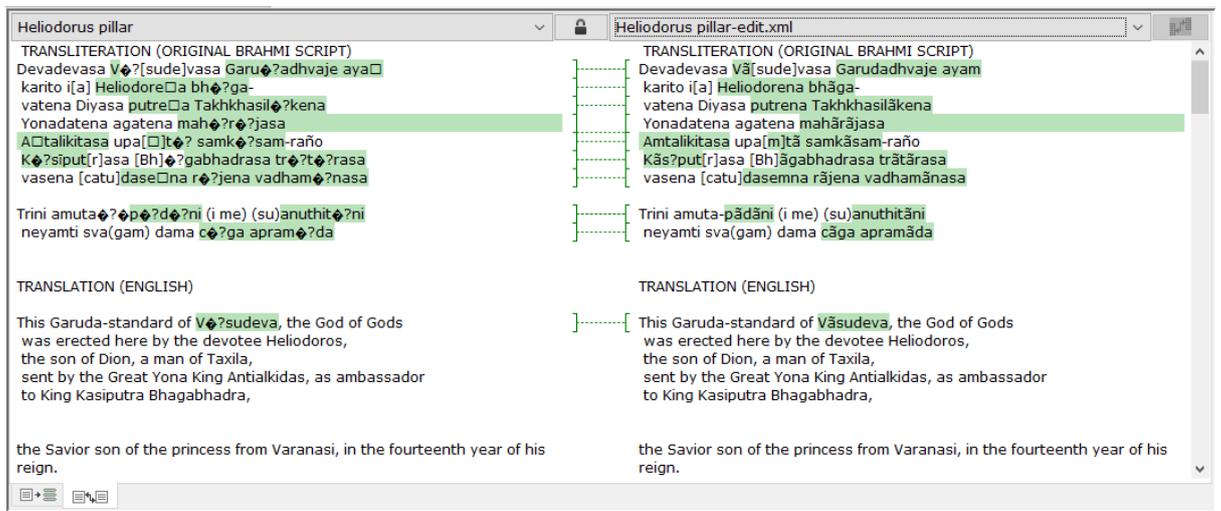


Figure 10 Manual text encoding in Heliodorus (Document after customization in Juxta)

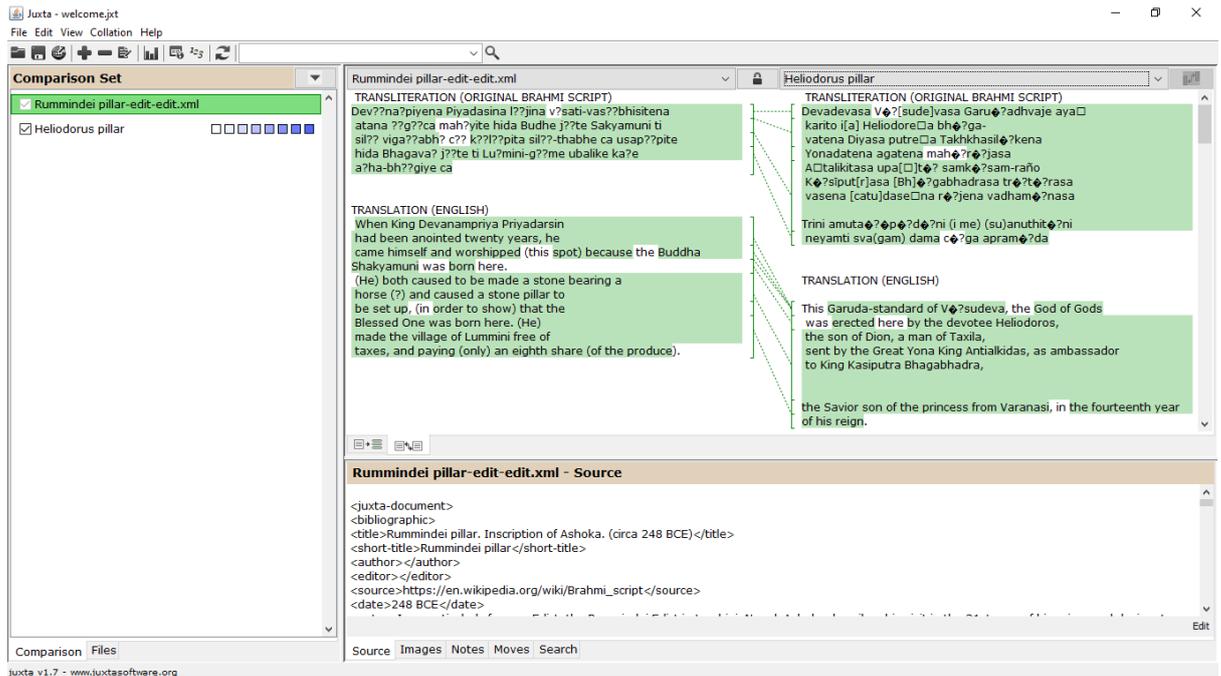


Figure 11: Comparison view of Sample 1 and Sample 2 in Juxta software

Discussion and Conclusion

According to Library of Congress, XML (eXtensible Markup Language) format is an international standard since 1998 and is an appropriate option for long-term preservation. XML is both a file format and a text-based, self-describing, human-readable markup language that is independent of hardware and operating systems.²¹ In Juxta, XML file is an input for the collation of the witness documents; and due to the structured nature of XML, it is suitable for automation, which empower us to deal with various new file formats in the future. In this way, Juxta supports the preservation of documents. Now talking about the document presentation paradigm of witnesses in Juxta as complex digital objects, the representation of the image is especially important to preserve the unique phenomena of manuscripts. XML is platform independent so it is one of the finest options for sharing the re-purposed material of a document with others, and this is what exactly done by collating the witnesses in the Juxta software as depicted in Figure 2, 5, 6, 9, 10 and 11.

Another beneficial factor is the sustainability of Juxta software. Juxta is open-source software and it is maintained by the NINES (Networked Infrastructure for Nineteenth - Century Electronic Scholarship), a scholarly organization devoted to forging links between the material archive of the nineteenth century and the digital research environment of the twenty-first.²² Sustainable open-source projects are those that are capable of supporting themselves. NINES affiliates consist of reputed Scholarly organizations (like Association for Computers and the Humanities) and Digital Humanities Centres (like DHO: Digital Humanities Observatory), so its sustainability feature is quite assured for near future.

While doing this research, some issues and challenges were also faced like scarcity of images of Ancient Indian text, its translation and transliteration work due to Covid-19. Challenges were also faced in text encoding due to the non-availability of some character in the encoding scheme of this software. Some other issue was related with the lack of syntactic and semantic

knowledge of ancient Indian script, therefore only those witnesses are taken whose semantic and syntactic information are provided by the subject experts. After observing all features and issues relating to JUXTA software including preservation and presentation of the documents, sustainability and challenges of software, this software inspires information professionals to initiate such activities for satisfying user's need while fulfilling the requirements for preserving complex digital object counterparts of ancient manuscripts.

References

1. Concept of Digital Preservation. Accessed August 25, 2020. <https://en.unesco.org/themes/information-preservation/digital-heritage/concept-digital-preservation>
2. Oliver G, Harvey R. Introduction. In: *Digital Curation*. second. ALA Neal-Schuman; 2016:3-18. Accessed August 25, 2020. http://digitalcuration.umaine.edu/resources/harvey_introduction_to_digital_curation.pdf
3. Holm P, Jarrick A, Scott D, Holm P, Jarrick A, Scott D. The Digital Humanities. In: *Humanities World Report 2015*. Palgrave Macmillan UK; 2015:64-83. doi:10.1057/9781137500281_4
4. Prebor G. New Technologies for the Collation of Hebrew Texts. *Zutot*. 2013;10(1):53-64. doi:10.1163/18750214-12341254
5. Collating | Definition of Collating by Merriam-Webster. Accessed August 26, 2020. <https://www.merriam-webster.com/dictionary/collating>
6. Peter M. W. Robinson. Collation, Textual Criticism, Publication, and the Computer. *Text*. 1994;7:77-94. Accessed August 26, 2020. https://www.jstor.org/stable/30227694?seq=1#metadata_info_tab_contents
7. CollateX. Accessed August 26, 2020. <https://collatex.net/>
8. Schreibman S, Kumar A, McDonald J. The Versioning Machine. *Lit Linguist Comput*. 2003;18(1):101-107. doi:10.1093/lc/18.1.101
9. Juxta | Collation Software for Scholars. Accessed August 26, 2020. <https://www.juxtasoftware.org/>
10. DAVIS ME. Juxta. Open-source tool and web service. *Renaiss Reform / Renaiss Réforme*. 2018;41(4):181-185. doi:10.2307/26644071
11. *Juxta User's Manual*.; 2011. www.juxtasoftware.org/
12. Using Juxta in the Classroom: Scholar's Lab Presentation | Juxta. Accessed August 26, 2020. <https://www.juxtasoftware.org/using-juxta-in-the-classroom-scholars-lab-presentation/>
13. Hultzsch E. *Inscriptions of Asoka*. Clarendon Press; 1925. Accessed August 26, 2020. <https://archive.org/details/InscriptionsOfAsoka.NewEditionByE.Hultzsch>
14. Brahmi script - Wikipedia. Accessed August 25, 2020. https://en.wikipedia.org/wiki/Brahmi_script
15. Dharma P. 008 Pillar Edict in Brahmi Script. <https://commons.wikimedia.org/w/index.php?curid=37363661>
16. Avari B. *India: The Ancient Past: A History of the Indian Subcontinent from c. 7000 BCE to CE 1200*. Routledge; 2016.
17. Wallace S. GREEK CULTURE in AFGHANISTAN and INDIA: OLD EVIDENCE and NEW DISCOVERIES. *Greece Rome*. 2016;63(2):205. doi:10.1017/S0017383516000073
18. Boppearachchi O. Emergence of Viṣṇu and Śiva Images in India: Numismatic and

- Sculptural Evidence. In: *From Alexander the Great to Kanishka: Numismatic Evidence in Constructing Early Central Asian and Indian History With Osmund Bopearachchi.* ; 2016. Accessed August 26, 2020. https://www.academia.edu/25807197/Emergence_of_Viṣṇu_and_Siva_Images_in_India_a_Numismatic_and_Sculptural_Evidence
19. Thapar R. *Early India: From the Origins to AD 1300* . University of California Press; 204AD. Accessed August 26, 2020. https://books.google.co.in/books?id=-5irrXX0apQC&pg=PA216&redir_esc=y#v=onepage&q&f=false
 20. Rapson EJ. Ancient India [1]. Published online 1914. <https://commons.wikimedia.org/w/index.php?curid=73465261>
 21. XML (Extensible Markup Language). Accessed August 25, 2020. <https://www.loc.gov/preservation/digital/formats/fdd/fdd000075.shtml>
 22. NINES: About. Accessed August 26, 2020. <https://nines.org/about/>