Meta-data and Methodology : Standards in the digital archive

Priyanka Suresh and Navjyoti Singh

Center for Exact Humanities
International Institute of Information Technology, Hyderabad, India
E-mail: `priyanka.suresh@research.iiit.ac.in, navjyoti@iiit.ac.in`

Abstract

To support advancements in digital archival and curation processes, developing tools and capabilities that can accommodate cultural objects and their interpretations are necessary, in compliance with meta-data standards. This helps in establishing contexts within and among collections, along with pedagogical discourse and an extension of themes. In this paper, we present Curarium, a novel platform for curation and collaboration with large cultural data-sets. It explores collections-data through tools for in-depth object specific research, annotations, interpretive meta-data and data-visualizations. Moreover, it ventures into adopting methodology standards of knowledge preservation using data and lifecycle management in cultural understanding to retain authenticity and integrity.

1 Introduction

Cultural data-collections are inherently interpretive and susceptible to dynamic changes based on the curator, archival guidelines and available assets. Even though standardization of meta-data during import is useful to maintain regulated knowledge, a formal space for interpretation is required. This needs a setting for endurance of ideas and emergence of discussions withstanding technical and resource obsolescence. For this purpose we perform development and user studies of the platform Curarium [1] (https://curarium.com/), currently in its beta version. The users are a group of alpha and beta testers, who performed the iterative process in a classroom and academic setting.

For the purpose of this paper, we refer to a cultural artifact as an artifact, the digital representation of this artifact ("digital surrogate" [2]) as an object, and an associated digital surrogate to an object as a surrogate. We attempt a streamlined model to develop academic arguments with cultural data while aiding knowledge preservation. This process has cultural objects as the subject matter, and it establishes relations between objects and the collection, and among different collections in the archive. A systematic overview, necessary to retain the objective of the collections imported, formalized by Curarium has the following schema:

1. Import collections and structure initial meta-data

2. Filter and query collections to explore specific objects

3. Curate meta-data extensions for individual objects

4. Annotate individual objects for in-depth themes

5. Generate visualizations to represent objects and inter-relations

6. Create and categorize sub-collections with objects from across collections

7. Publish documentation and insights

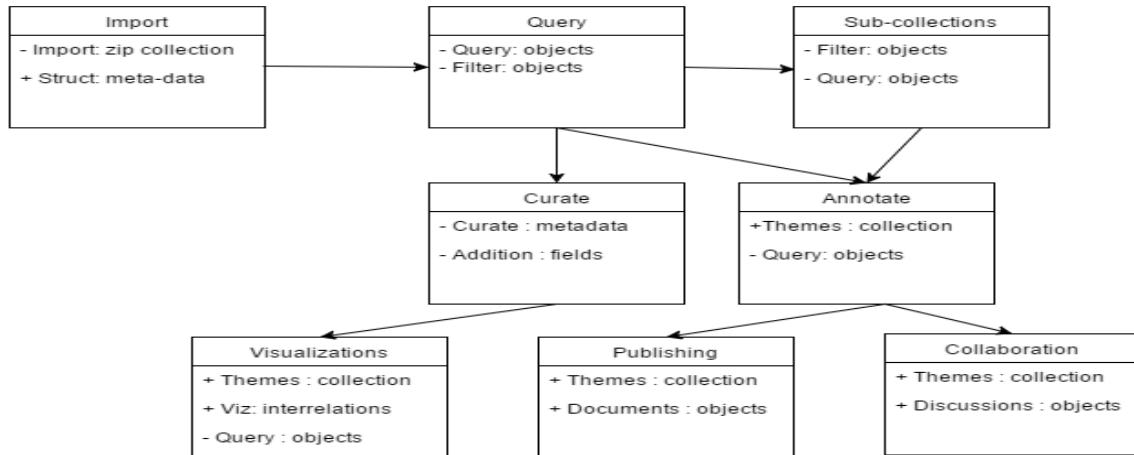8. Develop collaborative arguments over shared annotations

Figure 1: Curarium Process Schema

2 Related Work

There are many proprietary and open-source tools that support digital archival and curation processes on the web. Contentdm by OCLC [3] is a powerful and easy cataloging tool but expensive. Open-source tools such as Omeka [4] provide capabilities limited to meta-data analysis and indexing but have no support for collaboration, and others such as Collective Access [5] requires advanced technological capabilities for curation and collaboration. Pachyderm by New Media Consortium was highly interactive but did not provide extensive cataloguing capabilities and required download for individual hosting. In comparison, Curarium handles meta-data cataloguing and collaboration from a curator and user side, is free and easy to use, packaged without the requirement of browser downloads and highly interactive. For meta-data curation purposes, Curarium follows the Resource Description Framework [6] (RDF) model, which is a W3C recommendation for defining and sharing meta-data among communities.

3 Data preparation

Curarium is a web application built on Ruby on Rails framework. Based on the model-view-controller standard, Curarium uses PostgreSQL as the database server and JSON for data transfer. Curarium also heavily utilises JavaScript and CoffeeScript, thereby making JSON the ideal choice for a lightweight data-interchange format [7]. During import, the collection provider follows an organized protocol for meta-data by filling out the fields for general information - name, description, provider detail, a sample of the collection and a zipped file containing all the records in JSON format. This is further parsed to provide a list of name/value objects making the information highly human-readable. The collection provider is then able to drag-and-drop required values from the parsed JSON data into fields of interest (Examples include "Unique identifier", "Title", "Thumbnail" etc). There is a provision for the collection provider to add custom fields, adding flexibility to curation for the meta-data to be very collection-specific.

4 User study

Iterative alpha and beta testing was carried out for observing User Experience among users with different levels of software and curative expertise. A quick guide is presented within the tool to

aid users in understanding the usage and associated vocabulary. The tool has been tested for use following the schema in Figure 1 with use cases for (1) academicians and curators in research settings, (2) students in classroom settings with the course instructor as moderator and (3) workshop participants with organizers as demonstrators for the purpose of ideation. In these use cases, users were provided special access to the website with generated accounts. The platform currently hosts around 8 collections, sourced and imported successfully with a large focus on curation through import. Villa I Tatti's Homeless Paintings of the Italian Renaissance [8] collection was the first to be imported into Curarium, followed by collections made available and hosted by Harvard Art Museum, Japan Digital Archive [9] and others. Within the classroom setting, the available collections were useful for students in developing academic arguments through publishing and discourse on message boards promoting engagement within established contexts. Within the workshop setting, the collections are provided for presentation in broader contexts for pedagogical dialogue through visualizations, documentation and useful content. Through the agile model of development and version control on Github [10], many illustrations and iterations were incorporated for release. Curarium is now available with user profiles, a provision of adding users into research-specific circles and login credentials authenticated by Mozilla's Persona [11]. The platform is linked to Redmine [12], a project management tool for reporting issues and feedback.

5 Standardization design

5.1 Technical standards for materials

The qualification of an artifact (and thereby an object) as possessing cultural value requires categorization in order to fit it into archival themes. The first step to this is determining cultural artifacts on the basis of their utilitarian value [Prown 13] and the second step as identifying materials and their digital representations [Portela 14]. This allows for thematic extensions as well as a seamless transition and interaction between objects.

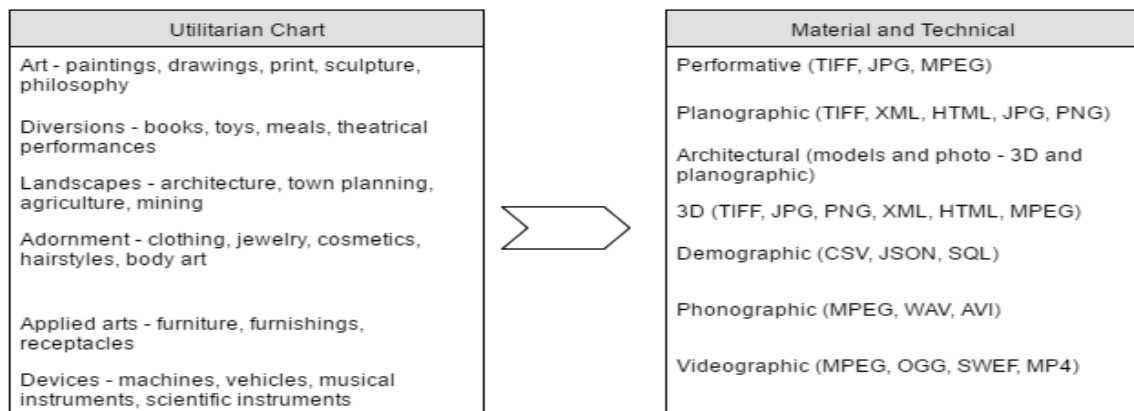| Utilitarian Chart	Material and Technical
Art - paintings, drawings, print, sculpture, philosophy	Performative (TIFF, JPG, MPEG)
Diversions - books, toys, meals, theatrical performances	Planographic (TIFF, XML, HTML, JPG, PNG)
Landscapes - architecture, town planning, agriculture, mining	Architectural (models and photo - 3D and planographic)
Adornment - clothing, jewelry, cosmetics, hairstyles, body art	3D (TIFF, JPG, PNG, XML, HTML, MPEG)
	Demographic (CSV, JSON, SQL)
Applied arts - furniture, furnishings, receptacles	Phonographic (MPEG, WAV, AVI)
Devices - machines, vehicles, musical instruments, scientific instruments	Videographic (MPEG, OGG, SWEF, MP4)

Figure 2: Utilitarian and Technical extensions of artifacts

In Figure 2, the utilitarian chart helps distinguish the various artifacts in a collection based on their usage. This not only helps establish an overall utilitarian theme for the collection itself but also individual objects in the collection; and then recognise the diverse materials of the artifacts and their technical representations supported by the web platforms. Example : The Lighting Devices [15] collection currently hosted on Curarium would belong to the "Devices" section and contains "Photographic materials" of Lighting Devices.

5.2 Interpretive spaces

We create an extension space for exploration of themes where individual objects are subject to supplementary meta-data creation. This provides contextual detail to the existing meta-data structures. On Curarium, upon clicking on an object, a custom client-side extension is present over the preferred set of primary fields that were selected during import. By making additions through this interface, the user with approval of the curator, is able to build relations between objects in a collection, make sub-collections and find overlaps in cross-collection contexts. The dimensions of contexts [Beaudoin 16] for object-specific curation are diagnosed for further investigation and collaboration.
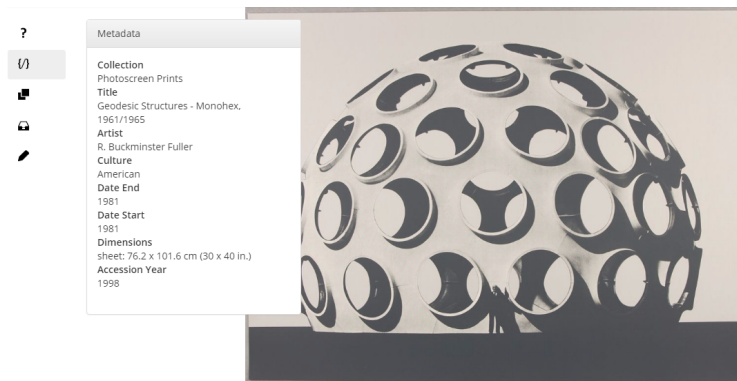

Figure 3: Object-specific meta-data

Figure 3 displays a record from the "Photoscreen Prints" [17] collection currently hosted on Curarium. On clicking the record, a panel on the left presents a tool-kit to view (and edit with authorization) the associated meta-data.

There is an intention space, where annotations are developed for an interactive method of delving deeper into object specific research. The object is administered on a Javascript canvas designed to be malleable for indexing and focussed interpretations within an image [See Figure 4 example]. Detailed and critical findings are then published and associated content is saved as a surrogate of the object. This contributes to the increased utility of objects and fertile evolution of the collection and discourse around it.

Additionally there is an analytical space in the form of data-visualizations to observe how objects in a collection and sub-collection interplay with each other [See Figure 5 example]. In-depth queries for words or phrases relating to a collection, a color representation of objects, and scaled versions (using thumbnails, object maps, tree maps) are used to visually structure complex relations within structured meta-data.

Figure 4 (on the next page) displays a record from the "Photoscreen Prints" [17] collection currently hosted on Curarium. On clicking the record, a panel on the left presents a tool-kit to annotate a portion of the image and transcribe associated information. Similarly, Figure 5 (on the next page) displays the collection "Photoscreen Prints" [17] currently hosted on Curarium in visualization mode [18] of the object map. The panel on the left displays the available types of visualizations, properties of viewing, and filters associated to the visualization.
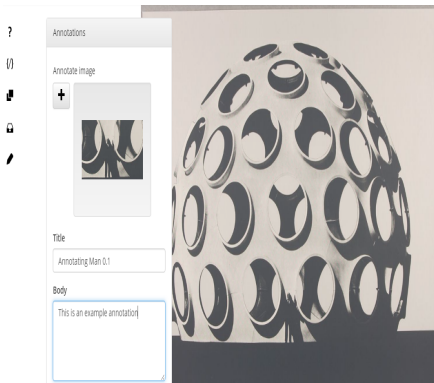
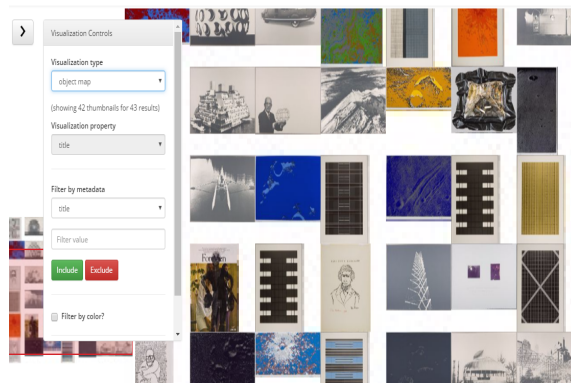Figure 4: Object-specific annotations

Figure 5: Collection-specific visualizations

6 Strengths and Limitations

6.1 Strengths of the Platform

Curarium makes the process of curation and collaboration highly intuitive and interactive, aiding pedagogy and enhancing expertise in the inter-disciplinary domain of Digital Humanities. Moreover, it enables data-visualization of large and complex cultural data-sets which enables recognising patterns, building relations and grasping sophisticated concepts. Additionally, it comes with a wide-range of customizable in-house tools for archiving and open to public feedback for improvement of the products. Lastly, the image and thumbnails are in the form of links provided, so that the object of the collection can be mirrored into the platform from its respective home-site rather than domestically hosted. This makes Curarium highly scalable in terms of importing and working with large cultural data-sets.

6.2 Limitations of the Platform

For the sake of converting other data-formats (CSV, XML) into JSON, Curarium uses an in-house third party API. Integration of this API into the platform would make it more versatile in handling a variety of data-sets. Also, Curarium currently archives only photographic material. Providing the capacity for audiographic and videographic material would enrich the archive and enable diverse relations among objects and collections.

7 Acknowledgment

Curarium is supported by metaLAB(at)Harvard associated to the Berkman Klein Center for Internet and Society. We thank our colleagues who provided the support, insight and expertise that greatly assisted the research.

8 Conclusion

In conclusion, a space for interpretive meta-data is required to preserve the multi-dimensional knowledge contained in a cultural object. The biggest challenge occurs in the form of defining boundaries for cultural objects as the features are reliant on each other for meaning, and provide a means for an abundance of hybrid information. Through the technical affordances of Curarium, we have developed a mechanism for capturing structure using tools for the curator, while allowing flexibility for interactive interpretation of the objects during collaboration for the user.

References

[1] Curarium (2005) MetaLAB(at)Harvard. (URL : https://curarium.com/)

[2] Rabinowitz, Adam (2015), A Work of Archeology in the Age of Digital Surrogacy. In *Visions of Substance : 3D Imaging in Mediterranean Archaeology*, The Digital Press at The University of North Dakota pp. 27-43

[3] Contentdm (2016) OCLC Online Computer Library Center, Inc. (URL : http://www.oclc.org/contentdm.html/)

[4] Omeka (2007) Roy Rosenzweig Center for History and New Media, George MAson University. (URL : http://omeka.org/)

[5] CollectiveAccess v1.6 (2016) (URL : http://collectiveaccess.org/)

[6] Resource Description Framework, W3 Consortium (URL : https://www.w3.org/RDF/)

[7] JSON (2013), *Standard ECMA-404*, 1st Edition, ECMA International

[8] Berenson, Bernard (1969) Homeless Paintings of the Renaissance, London, Thames & Hudson

[9] Japan Disasters Digital Archive, Reischauer Institute of Japanese Studies, Harvard University

[10] Curarium on Github, Inc. (2016) (URL : https://github.com/berkmancenter/curarium)

[11] Mozilla Persona (2011), Mozilla Foundation (URL : https://developer.mozilla.org/Persona)

[12] Redmine v. 3.3 (2016), Redmine (URL : http://www.redmine.org/)

[13] Prown, Jules (1982), Mind in Matter: An Introduction to Material Culture Theory and Method. In *Winterthur Portfolio*, Vol. 17, No. 1

[14] Portela, Manuel (2014), Multimodal Editing and Archival Performance: A Diagrammatic Essay on Transcoding Experimental Literature. In *Digital Humanities Quarterly*, Vol. 8, No. 1

[15] Lighting Devices (2016), Curarium (URL : https://curarium.com/collections/3)

[16] Beaudoin, Joan (2012), Context and Its Role in the Digital Preservation of Cultural Objects. In *The Magazine of Digital Library Research*, Vol. 18, No. 11/12

[17] Photoscreen Prints (2016), Curarium (URL : https://curarium.com/collections/9)

[18] Photoscreen Prints visuailzation (2016), Curarium (URL :: https://curarium.com/collections/9/works)