

Final Paper

Metadata is built on the concept of analyzing and organizing data about data and this perspective helps to set the standard for what users are intending to accomplish through the utilization of this practice. While metadata is valuable for means of description and access in all information systems, it can also be very convoluted and lacking integration as there are numerous metadata standards that often overlap in their intentions to better organize and structure the data systems and objects that they represent. One of the most common and prevalent metadata schemas due to its ease of use, is Dublin Core. This schema is popular and recognized as user-friendly because of its standardized vocabulary language which only includes fifteen elements. However, this ease of use does not make it unsusceptible to issues and long-term drawbacks to its greater utilization in organization of metadata. As Pomerantz describes in the text, *Metadata*, the elements in Dublin Core are made to be as “unambiguous as possible” which corresponds to, as he states, “[the] subtleties of meaning are flattened out”. This perspective in turn inspires us to ask the questions: 1) what we are losing through this schema and 2) how does that impact our data? Ironically, it can also be said, that in essence, the strength of this schema also carries its weaknesses. The significance of these questions is a common issue and great challenge for all metadata standards, as the dream of standardization being universally implemented across all domains may very well be impossible to achieve. However, by looking at this issue through the lens of the utilization of one standard, and in turn contrasting that analysis to another, users can identify how and when certain standards are needed and why the distinction between them is ultimately useful depending upon the subject material being recorded.

The necessity of metadata is only growing in our current technological age. As the advancement and dissemination of information further becomes an inherent and necessary right in modern societies, it is integral that we maintain systems that can promote the longevity of items, objects, artifacts, documents, images, and histories for the interminable future. Thus, the need for standardization of metadata is an ever-growing field as scholars continue to search for tactics that best incorporate information for future users from both a back and front-end perspective. In its conception, Dublin Core originated to describe and recognize varying items from both digital to physical resources. Today, this schema can be utilized for simple resource description to combining metadata vocabularies of different metadata standards. Further, Dublin Core can provide metadata interoperability for other vocabularies in both linked data cloud and Semantic Web applications. (“Metadata Basics”) This schema is built on the utilization of fifteen standardized elements in its description of data. As taken from the Dublin Core Metadata Initiative website, these include ‘contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, and type’. These areas are the foundational elements for this descriptive metadata schema.

Having established what the Dublin Core schema is, it is necessary to investigate the first argument of Pomerantz’s original statement regarding how the semantics being shared across domains is one of this standard’s inherent strengths. This can be recognized through the lens of the simplification of data and material sources. In this regard, it would be important to acknowledge that the purpose of metadata and the systems and infrastructures for preserving data is to make an object, item, and so forth accessible for the future. Naturally, one way to ensure that something (object, item, etc.) is accessible is to make it more simplistic. This position is something Dublin Core abides by, through having less categories and rules around the

description of an object the standard can enhance the user's comprehension when approaching objects and implementing this standard. At length, this foundation is precisely where Dublin Core's strengths prevail: through the process of establishing a limited vocabulary for the utilization of the standard, the barrier to entry is lower. This allows the standard to be user-friendly for both current and future users of an object, item, etc. for years to come as it is both consistent and more easily able to be classified and read for future analysis and investigation. (Ball)

Now, regarding the latter perspective that Pomerantz expressed in their initial statement, it is important to acknowledge that the simplification that is at the root of the Dublin Core schema is also its greatest weakness. While Dublin Core is valuable for promoting user accessibility, the negative cornerstone to that is the cost that comes from simplifying things down to a smaller pool of options. Here, the simplification results in the loss of material information and resources due to the minimizing of certain elements and options. Further, it is very possible to misinterpret data through this schema because the need to categorize in a specific way could create user confusion due to the limited categories that may not feel accurately applicable depending on what they are trying to describe. As stated in "Dublin Core: An Obituary", this "loss of specificity that occurs makes it difficult to then convert it into other systems or to transfer other data from different systems." (Beall) Beyond this sentiment, one could argue that metadata should be flexible to account for the various objects and artifacts it will come to be utilized for, and thus, through implementing a rigid system such as Dublin Core, we are losing space for broader identities and classifications. Therefore, it is easy to envision how this weakness in Dublin Core could potentially lead to difficulties with both retrieval and interoperability between standards.

Through further investigation of these constraints presented in Dublin Core, it is necessary to endeavor to foresee a disaster scenario due to the weaknesses found in this metadata schema. As this standard only allows for fifteen elements, more obscure and articulate information is at risk for long term obsolescence due to these potentially restricting standards. Take for example if a museum were needing to create a Dublin Core record for a photograph, due to the elements being so generalized we could in turn create a record that is inherently unspecific. Due to this, the prospect of losing data about the photograph is one potential disaster scenario. Beyond this, a second issue stems from the lack of consistency that could exist within a record, which, unfortunately, could revert the preservation efforts of this archived photograph null.

While the weaknesses in Dublin Core are prominent and should necessarily be considered, the Dublin Core Metadata Initiative (DCMI) has already begun to foresee a progressive future in this schema. According to the DCMI, the current focus at the organization is furthering the implementation of linked open data and the role that Dublin Core will enact in this future. With this initiative, the goal is to connect data from a variety of open sources—including libraries, health organizations, and government data—by utilizing this common standard. The hope through this process of establishing linked open data, is to develop the semantic web to where all data is both linked and easily analyzed by computers. The Dublin Core Metadata Initiative believes that through implementing Dublin Core metadata in conjunction with web content that a business or website can then become a part of the semantic web. At length, the data that is being stored in the semantic web is only increasing with the progression of time due to the rapid developments in cloud computing. Therefore, the implementation of a standardized metadata schema will allow further maintenance of this data and create ease of accessibility for all.

After examining Dublin Core, it is useful to compare it with another schema to analyze what other metadata practices can offer for long term preservation, description, and access. While Dublin Core can be utilized in various areas for resource management, certain other standards can prove just as (if not more) useful depending on the circumstances of what is being standardized. As an example, regarding the management of literature metadata, Dublin Core in conjunction with TEI are the two standards that are most implemented in this subject area, which in turn makes them a natural comparison as each provides effective tools in the formation and standardization of metadata. The Text Encoding Initiative, or TEI, was designed for marking up humanities documents and focuses on structure using coding. Further, it can be customized to include any variation of metadata. This ease of use is what has garnered TEI the reputation as a popular standard, and, as stated from the Introduction to the TEI Guidelines, “[it] provide[s] a means of making explicit certain features of a text in such a way as to aid the processing of that text by computer software running on different machines. This process of making explicit we call markup or encoding.” In comparing the two standards, TEI and Dublin Core overlap in their respective intentions towards accessibility. However, TEI is constructed with more flexibility than what is found in Dublin Core due to its limited set of fifteen elements. In TEI, the following, broader elements must be followed, which include “common core identifiers, and textual features are easily shared and accessible; easy modification features for additions and removals, multiple textual encodings of the same textual resource should be possible, mark-ups should be user defined, following certain small requirements, and the sufficient need for documentation to be written alongside what is encoding.” These features are the core module for TEI and highlight the differences in flexibility and fluidity it presents as a metadata standard in comparison to the Dublin Core Metadata Initiative.

TEI offers useful benefits for users as a metadata standard. As the TEI website states, the first strength TEI offers is that it is easily verifiable via an online checked validator which establishes ease of use for all current and future users of this system and enhances accessibility. In addition, TEI is sustainable and open source and is regularly updated. The schema also supports text analyses efforts and adheres to a common standard that can operate on a variety of systems. Further it addresses the problem of varying texts in being designed to be adaptable to these different texts. These factors allow the initiative to be continuously progressive rather than stagnant and abiding by a predetermined catalyst which makes it vulnerable to becoming outdated and insufficient for future user needs in the face of the rapid ascension of technology in the current day and age. In comparison to this perspective, TEI does, as all standards inevitably do, come with its' own inherent disadvantages. In this case, these weaknesses include: 1) the process of encoding can be time-consuming due to the efforts of marking up documents. Additionally, 2) the way documents can be marked up also has the potential to be inconsistent as these markups are dependent on the contributor, and thus could make descriptions different not only across but also within varying projects.

Drawing from these weaknesses, it is important to imagine what a potential disaster scenario could encompass due to the flaws in this specific standard. As discussed, the inherent likelihood of experiencing inconsistencies so easily within the process of this standard is a major issue as fluctuations and differences could lead to long-standing difficulties. However, while TEI can substantiate these differences, it can also, through creating well-formed XML documents, be easily validated against standard TEI schema which allows modifications to be correctly documented across platforms quite easily due to the use of software and online validator. In essence, while any metadata standard would inherently possess its own positives and negatives,

and although the barrier to entry for the user to interface with TEI is greater than that of Dublin Core, the positives in TEI inherently solve the negatives. Further, the weaknesses present in Dublin Core are more significant, as the threat of loss of specific information that exists in Dublin Core is something TEI is wholly qualified and designed to prevent. In the end, TEI utilizes expansive space to encourage proper description for long term accessibility needs that can be implemented as successful digital preservation efforts at museums, libraries, and other archival institutions.

As we've discussed, all metadata standards are susceptible to the loss of information. However, TEI does prioritize the improvement of its guidelines in the future to address these concerns. In a study conducted on TEI's user base, it was discovered that while TEI addresses many of the needs for its users, there are some areas and users that TEI does not address properly and these individuals' concerns are now being recognized and acknowledged for the improvement of the standard in whole. In "The Present and Future of the TEI Community for Manuscript Encoding" the authors address how it appears that the issue most users addressed was the amount of effort it took to simply use TEI, due to the technicalities inherent in the standard such as TEI schemas and further personalization and customization that is available. In the article, it is hypothesized that this issue could initiate from the directions in the guidelines provided on the use of TEI rather than the negatives of TEI itself. However, it is still critical to have these concerns addressed properly for the further ease of description and accessibility for all. Further, it appears some drawbacks come from the specificity utilized in the schema and the desire for more permissible guidelines that allow users to better handle uncertainties. In turn, the Text Encoding Initiative through the configuration of a broad and diverse range of scholars' efforts is reconfiguring its priorities to strengthen the guidelines to address these grievances

within the standard in the hopes of implementing long-term solutions in the future of this initiative. (Burghart and Rehbein)

At length, both Dublin Core and TEI offer simplistic and useful ways to organize resource materials into metadata schemas which in turn is necessary for further preservation, description, and access. The foundation of this comparison between standards stems from the notion of Dublin Core's inherent strengths and weaknesses as acknowledged in Pomerantz's original statement. However, after thoroughly comparing the Dublin Core Metadata Initiative to TEI, it is made clear that despite its weaknesses the standard will remain for its ease of entry and implementation as a metadata standard. While Pomerantz's argument was right to be addressed as standards should be critically and conscientiously analyzed rather than accepted as they are, wholly unchangeable and immutable. Yet, this challenge and analysis is vital and necessary for greater understanding, development, and continuous improvement and most keenly a representation of the evolution inherent in our current world. As our technologic and preservation needs progress with the influx of data and the necessity for long term accessibility, the potential loss of information only increases, which is in turn why these systems must be reexamined to ensure longevity. Although Dublin Core does in fact permit for a level of flattening out of information that could very well continue to occur, the long-term positives outweigh the arguable negatives that are inherent in this metadata standard. In the end, both Dublin Core and TEI offer positive and negative aspects and these must be acknowledged and accepted before the utilization of one over another on any given classification that could arise in the future.

Resources

1. Ball, A. "Review of the State of the Art of the Digital Curation of Research Data (version 1.1)." *ERIM Project Document erim1rep091103ab11*. University of Bath, 2010., Accessed June 5, 2022. <http://opus.bath.ac.uk/18774/2/erim1rep091103ab11.pdf>

2. Beall, J. "Dublin Core: An Obituary." *Library Hi Tech News*, 21(8), 2004, pp. 40-41.,
Accessed June 1, 2022., <http://taddeo.emeraldinsight.com/>
3. Burghart, Marjorie, and Malte Rehbein. "The Present and Future of the TEI Community for Manuscript Encoding." *Journal of the Text Encoding Initiative*, no. 2, Feb. 2012, pp. 1–27., doi:<https://journals.openedition.org/jtei/372#tocto2n5>.
4. "DCMI Metadata Terms" n.d. Accessed June 5, 2022.,
<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
5. "Introducing the Guidelines – TEI: Text Encoding Initiative." n.d. Accessed June 1, 2022.,
<https://tei-c.org/support/learn/introducing-the-guidelines/>.
6. "Metadata Basics" n.d. Accessed June 1, 2022.
<https://www.dublincore.org/resources/metadata-basics/>
7. Pomerantz, Jeffrey. *Metadata*. MIT Press, 2015, pp. 1-64.