Master in Technical Communication and Localization (TCLoc)

A Meta-Analysis of Structured Documentation in Virtual Spaces

An exploration of ethics, practice, and documentation

A thesis submitted in partial
Fulfillment of the requirements for
the degree of Master of Arts

By

Michael Pell Blumenthal Martin

2024

Drew Eisenhauer: thesis director

Abstract

This exploratory research thesis challenges the conventional notion of audience in technical writing within virtual spaces. Focusing on the evolving landscape of authorship in digital environments, I put forward the question of the centrality of human audiences as primary criteria for authorship in technical communications. The shift towards machine intelligence and the prevalence of machine-generated content raise critical considerations for technical communicators. The article advocates a reevaluation of ethical responsibilities, controls, and publication processes in response to the transformation brought about by tools like XML and semantic segmentation.

I contend that documentation, viewed as data rather than a traditional document, undergoes a paradigm shift in the digital age. The proliferation of machine-authored content and automated management processes emphasizes the significance of semantic markup. The article explores the challenges of audience ambiguity, the non-linear temporality of documentation, and the complex nature of digital spaces. In Latourian terms, the article critiques the fictional mode of existence assigned to documentation for a specific audience, proposing a return to subject-based integrity and subject matter expertise for content delivery in the public domain.

A key argument centers around the necessity for a different ethical stance, acknowledging the pluralistic, boundary object nature of documentation in digital spaces. I advocate for a departure from the user-centric model, emphasizing the quality, integrity, and verifiability of information in the public domain. The ethical imperative extends to the societal impact of information dissemination, with a focus on maintaining information quality and integrity. The article proposes a unified effort among knowledge workers, including technical communicators, to establish stringent professional requirements, review processes, and recourse mechanisms for bad actors in information professions. The findings aim to inform industry practitioners and educators, facilitating informed decision-making and contributing to the ongoing evolution of effective technical communication practices and ethics.

Acknowledgements

This work would not have been possible without the support of my academic advisor, Professor Eisenhauer, and the support staff of the TCLoc program at University Strasbourg including Mme. Director Renate de la Paix, and Mme. Coordinator Valerie Ledermann. In addition to the academic and program support of the TCLoc program I would like to acknowledge the impact and tradition of academic discourse in the philosophical tradition which has continued to influence and be a beacon throughout my life. I am grateful for the support and guidance of my family in continuing to support my academic pursuits and especially to my partner for indulging my erudition.

Table of Contents

Methods

Theoretical Framework

The research begins with a literature review to establish the theoretical foundation and current practices in structured documentation. Key concepts, principles, and frameworks, such as DITA and DocBook, are critically examined to provide a comprehensive understanding of the theoretical underpinnings that guide the structured documentation approaches currently in use. A review of science and technology studies (STS) work of Star and Latour is reviewed, as well as sensemaking in virtual spaces to provide additional vocabulary and reframe the ethical scenario professionals work within.

Theoretical Frameworks and Literature Review:

DITA (Darwin Information Typing Architecture) Structured Documentation Framework

Origins and Development

DITA, initially developed by IBM, is a widely adopted structured documentation framework designed to facilitate the creation, management, and publication of modular and reusable content (IBM, n.d.). Emerging in the early 2000s, its development has been collaborative, involving contributions from various industry leaders. DITA's design is grounded in principles of information typing, modularity, and specialization.

According to the OASIS 2005 DITA standard publication,

> The Darwin Information Typing Architecture (DITA) specification defines both a) a set of document types for authoring and organizing topic-oriented information; and b) a set of mechanisms for combining and extending document types using a process called specialization. (Priestley & Hackos, 2005)

This serves to assist implementers of the DITA standard. The body of the standard consists largely of the distinctions of the elements used in DITA. This standard has evolved to include additional elements and specializations as the standard has grown. As seen in the distinctions between the opening pages of the 2005 standard above and the 2015 standard shown below.

> Darwin Information Typing Architecture (DITA) Part 3: All-Inclusive Edition (this document). This edition contains the base architecture, technical content, and the learning and training specializations. It is designed for implementers who want all OASIS-approved specializations, as well as users who develop learning and training materials. (DITA-v1.3-part3-all-inclusive, 2015)

In the ten years since the original standard release, we see an adaptation of the standard to the new needs and challenges associated with the standard being more widely adopted.

Acknowledgement that specific information is needed for developing learning and training materials has developed during the 10-year span as a community directed response to the dynamic digital economy which DITA is implemented in. This information specificity as a sign of advancement and adaptation is an important aspect of the development of DITA as a tool. It is also an important notion for the standard body itself as it is for the community driven response of developing knowledge management tools and frameworks. The rapid expansion of the scope of information management has created an increased demand for tools to organize and reuse content for specific purposes, reuse being a primary driver of structured authoring efforts. The scope of the effort required to implement structured authoring tools is an important point of consideration when an organization is considering adopting DITA as a tool for document management and authoring. Given the technical training required and the investment required to convert existing documentation and databases into DITA compliant forms, the weight and complexity of this effort can be costly and difficult to find consistent talent for, frequently this results in high-cost consultants being recruited for conversion and adoption efforts (Akinci, U., 2014). This is widely considered a critique of DITA and similar semantic markup single-source authoring and publishing approaches.

Information Typing

At the core of DITA's theoretical foundation is the concept of information typing. DITA classifies content into distinct information types, such as concepts, tasks, and references (OASIS-tcs, *Overview of DITA - Information Typing*). This typification allows for the creation of modular and reusable chunks of information, facilitating content reuse between different documents and contexts. DITA is rooted in the paradigm of topic-based authoring, where each piece of content corresponds to a self-contained topic. This approach promotes clarity, simplicity, and efficiency in content creation. Topics, encapsulating a single concept or task, can be easily interlinked and repurposed, contributing to the scalability and adaptability of documentation.

DITA promotes a modular approach to content creation, emphasizing the development of standalone topics. This modular structure compliments topic-based authoring and enhances content reusability, enabling authors to assemble documents from pre-existing components. The framework's specialization mechanism further supports customization, allowing organizations to

tailor DITA to their specific documentation needs without sacrificing interoperability. This allows for restriction of element use through the existing general elements while enabling flexibility when warranted (DITA-v1.3-part3-all-inclusive).

Structured Information Architecture

DITA introduces a structured information architecture that goes beyond traditional document-centric approaches. The framework emphasizes the creation of a map, or a hierarchical structure, to organize and link individual topics. This architecture enhances the overall coherence of documentation sets, making them more navigable for end-users and content managers.

Structured information architecture is supported by the use of elements which act as searchable reference points within the documentation and allow for the exploitation of tree data structures inherited from the eXtensible markup language (XML) markup language foundation. This hub or web of references is a key component of structured information, something we will consider further below in the exploration of XML.

Directions and Developments

DITA's development is dynamic, with ongoing evolution driven by community involvement and feedback. DITA's theoretical underpinnings include its status as an open standard, fostering interoperability among different tools and systems. Open standards promote collaboration and ease integration with various content management systems, authoring tools, and publishing platforms, aligning with the evolving landscape of technical communication (OASIS-tcs, n.d.).

In summary, the DITA structured documentation framework's theoretical underpinnings encompass information typing, modularity, topic-based authoring, structured information architecture, specialization, openness, and adaptability. Its design principles address many of the challenges of technical communication, offering a theoretical framework that aligns with the needs of organizations seeking scalable, modular, and interoperable solutions for content creation, delivery, and management.

The interiority of information demands has changed the corresponding demands on technological solutions to meet the use requirements of content and documentation. After exploring a similar tool for structured documentation, DocBook, we will explore the role of this shifting baseline on the uses and appropriateness of a methodology or framework of solutions. Secondly, we should keep in mind the transition seen in the 10-year period between standards presented. The shifting baseline between 2005 and 2015 usage of the DITA architecture acknowledges the need for training and instructional specialists to grow as the standard adjusts to the underlying trends of demand to address content management. This shifting baseline in the appropriate application of DITA speaks to a need for abstraction and theoretical development in technical communications as a divestment of training required into a more laterally and vertically transferable body of knowledge and professional skill set.

DocBook Structured Documentation Framework

Historical Context and Development

DocBook, an XML-based markup language for technical documentation, has its roots in the SGML (Standard Generalized Markup Language) era and was later adapted to XML (eXtensible Markup Language). Originally developed by HaL Computer Systems and O'Reilly & Associates, DocBook eventually broadened in scope and is now maintained by the OASIS consortium (Walsh, 2020).

DocBook has evolved over the years to become a widely adopted tool for creating and organizing technical documentation. Although a somewhat older XML tool, it still holds merit in the scope of this research as an approach to structured documentation which has successfully made use of open-source communities and tool chains to provide flexible solutions for structured digital authoring. The flexibility and accessibility of DocBook is of particular note as the ability to define the document type definition (DTD) leverages an existing tool kit around that specification. This contributes to the aim of single-source publishing and topic-based authoring efforts (Martinez-Ortiz, Ivan & Moreno Ger, 2006). The DTD approach established precedent in professional practice for digital authoring tools and capabilities in structured authoring. This is in part due to the validation tools of the XML parser and DTD in being able to vet the compliance

of structured documents with a particular set of requirements specified by the author or content manager.

Document Structure and Semantics

A framework of semantic elements form the core of DocBook’s flexibility. DocBook's theoretical framework focuses on defining a standardized structure for documents, encompassing elements for chapters, sections, paragraphs, tables, and various types of specialized content such as code listings, equations, and bibliographies. This structured approach enhances content consistency and facilitates automated processing. This processing is performed by a *parser* which segments the documents according to a set of rules. Often in DocBook, this is the DTD and the XML rule set which can check validity. Although not required, this aspect of markup document validity increases overall quality and integrity of documentation (O’Reilly & Associates, Inc., 1999). Semantic elements in the logical structure separates content from appearance and allows for the division of labor in the production of knowledge and the production or delivery of content.

Semantically structured documentation through semantic markup allows DocBook to promote content reuse by way of modularization. The framework allows the creation of document fragments that can be reused across multiple documents or multiple publication needs, with the principle of publication and appearance of information being handled with a different set of tools than the semantic categorization. The adaptability of the underlying SGML/XML markup structure ensures that the framework can accommodate diverse technical content requirements across different industries by making use of different semantic tags through parameter entities (O’Reilly & Associates, Inc., 1999).

Cross-Referencing and Linking

DocBook's foundation in structured documentation allows for a robust system of cross-referencing and linking between different parts of a document and even across multiple documents by relation of semantic elements. This linking mechanism enhances document and content navigability, ensuring that related content is easily accessible, fostering a coherent and interconnected knowledge base (O’Reilly & Associates, Inc., 1999). This is where we can begin

to see how the implementation of structured documentation has influenced expectations of single-source authoring and of information storage and retrieval.

This idea of linked knowledge builds on the foundational idea of markup in documentation to then have semantic or syntactic elements which are categorized and treated according to domain or application needs. This can be seen in current market implementations of middleware providers who exclusively facilitate this backend linking and reference dynamic of content repositories to delivery interfaces, one such example is PoolParty described as an intelligent content hub (daSilva, 2022). This concept of markup linkages also forms the backbone of Google's initial page rank algorithm, contemporary SEO backlink concepts, and early efforts at multidimensional knowledge representation through virtual spaces, such as the Xanadu project. This was also one of the early inspirations for Tim Burners Lee to implement a network of linked computer pages made available in the early academic model of the internet. Although the precise implementation has shifted over the years, the theoretical underpinning of tags and link reference via markup and expression matching via parsers has been a foundational development in the digitization of knowledge, the presentation of records and documentation, and the accessibility to information.

Socio-technical Implications of Structured Authoring

Recognizing the global nature of technical communication, DocBook incorporates features to support internationalization and localization. Through the use of attributes and specialized elements, DocBook documents can be adapted for different languages and cultural contexts, enhancing the framework's applicability in multinational settings (O'Reilly & Associates, Inc., 1999). In light of these attributes of DocBook, we can begin to interpret the socio-technical context in which structured authoring takes place. To introduce some concepts from Science and Technologies Studies (STS), Susan Leigh Star's concept of a boundary object as an artifact which is interpreted in a variety of ways by different actors is useful. She broadly classifies boundary objects as either attempts at intelligence or attempts at intelligibility (Star, 2016). DocBooks is an example of an attempt at intelligibility where the socio-technical context of the object generated is generally an attempt at conveying a concept to another human intelligence. This definition of DocBooks as a tool for the production of boundary objects carries with it a

valuable type of wayfinding in the cultural geography of structured documentation. The XML basis of element typing that DocBooks builds upon, as already stated above, is a set of abstracted element tags which can perform semantic roles across a variety of domains. In the context of Star's work and definition of boundary objects, this linking and tagging aspect becomes a mapping mechanism of what Star states as the Durkheim test, a contrast to the more widely known Turing test:

> the Durkheim test would be a real-time design, acceptance, use, and modification of a system by a community. Its intelligence would be the direct measure of usefulness applied to the work of the community—in other words, its ability to change and adapt, and to encompass multiple points of view while increasing communication across viewpoints or parts of an organization. (Star, 2016. Pg 247)

In light of this broader socio-cultural context of the role an artifact plays, an ontological motive of structured authoring is highlighted. The development and delivery of structured documentation becomes the work of multiples, not the work of a single author. In this manner we can reframe the development of structured authoring not toward a singular audience or set of personas but as establishing group epistemic basis for the conceptualization of document content as a pluralistic ontological status. Stated differently, the community under question is not only the authors and professionals but also the users of the products of that profession over time and the consequents of those who use those products. A document then has a lifecycle which is the sum of all input references for authorship and stakeholders with professional interest as well as all those parties who view, make decisions, and are affected by the availability, content, and quality of information in that document. This comprises the group epistemic socio-cultural context of documentation as a pluralist ontology.

In the case of structured authoring, rather than having a document centric view of a boundary object we can objectify the modular components of authorship and the issuant delivery channels for the lifespan of that information module. Such a view significantly expands the consequence and responsibility of authorship ethically as we consider a larger web of actors, actants, and objects which relate to the boundary object of documentation. Entailments of this position on the role of professional authors in digital spaces will be explored further below.

Community Involvement, Standardization, and Scope

DocBook's ongoing success is dependent on a commitment to community involvement and consensus-based standardization (O'Reilly & Associates, Inc., 1999). As an open standard maintained by a collaborative community, DocBook benefits from many of the aspects identified by Star in the Durkheim test. The "acceptance, use, and modification of a system by a community" results in updates and refinements of DocBook and of the realms of application for the tools developed by the community. This ensures that the framework remains relevant and adaptable to emerging technical communication needs in an open system of multiple actors with a diverse range of purposes and motivations using the tool. This is an important historical development in the use of technology. However, community in this sense is still limited in the view that only authors are the developing members of the community, there is limited input from public governance and parties impacted by structured authoring tools in public information access, availability, and quality. The extent and impact of these digital authorship tools and their best practices are not adequately represented in contemporary decision or ethical systems even though they are open access.

DocBook's approach recognizes the importance of integration with publication and information storage and retrieval or database toolchains (O'Reilly & Associates, Inc., 1999). The foundation in XML technology has allowed for the integration of DocBooks with a large variety of production level publishing and tracking tools. Notably among these is the concept of using version control in relation to documentation development and the idea of forking development efforts to adapt to changing business needs, knowledge base updates, or data backup in the event of corruption. This concept of agile development is testament to the boundary object concept. A functional requirement in knowledge or content development using structured documentation approaches is now that the audience, or needs of a publishing audience, may change in a manner that is independent from the semantic knowledge base itself. Therefore, viewing semantic markup as a tool for the development of boundary objects in a knowledge base which is selectively delivered to audiences as the need is specified is not only a tool for the delivery of business and consumer products but also a tool in the development of knowledge and ontology more broadly when presented to the public domain. Content and knowledge production is shifted from an approach to production that emphasizes audience or user, to a matter of syntactic

delivery of established modules. Stated separately, because of the division of labor that structured authoring allows for, the audience becomes a concern of syntax delivery not of semantic quality, this changes the frame for authoring but has not yet been recognized or adopted by the technical writing community more broadly. The integrity and accuracy of public domain semantics and ontologies is decoupled from the audience in this view.

Consider that the abstraction at play in structured documentation approaches to documentation has allowed for greater utility and reuse of knowledge products because of their semantic structure being rooted in cross-domain boundary object formats. These formats have led to the integration of structured documents into other tool chains as needed, extensions of artifact supports the concept of boundary object that Star developed as each delivery of the same content is tailored for a separate audience who is likely to have a different interpretation and context for the same content. Secondly, we can view the role of structured documentation as being both an *attempt at intelligibility* and an *attempt at intelligence,* structured documentation is both machine and human readable, documentation is both trying to relay information and store information as concisely as possible without information loss. In concert, documentation efforts through digital environments must by definition of their context occupy a pluralistic ontological status.

From the pluralist view of the ontology of documentation, there becomes a need to readdress the ontological status of the *work* of a technical communicator or team in the production of documentation because the audience will always be multiple, a nominal development audience and the real audience exposed to the documentation throughout its life time. This redress is called for in the act or process of how work is regarded and conducted by professionals. All documentation artifacts are now boundary objects with extended scope of lifecycle because either the semantic structure or the syntactic structure of the same artifact can be interpreted and utilized by multiple actors, for different reasons, across different temporal periods, by people at varying levels of education and development (Star, 2016). This now also includes non-human actors of automated machine and statistical learning tools. This concept is particularly important as we move forward in the development of structured documentation through the early 21st century and see the development away from DocBook and similar tools toward more abstracted tools and more interoperable formats of these concepts, such as the middleware concepts found

in content management systems and more recently in intelligent content hubs (daSilva, 2022). This is especially apparent in the development of transformer tools, such as ChatGPT, and in the context of pervasive web crawlers and scrapers.

In summary, DocBook's theoretical underpinnings encompass document structure, content semantics, modularity, topic-oriented authoring, cross-referencing, internationalization, output format versatility, community involvement, and integration with toolchains. This theoretical framework positions DocBook as a comprehensive and adaptable tool for the development of boundary objects that technical communications makes use of for selective and emergent audiences. Meeting the diverse needs of organizations engaged in creating and managing technical content, this boundary object single-source authoring approach has been influential on the development and delivery of technical communication artifacts and the development of the profession overall.

We must consider the abstraction away from a temporal notion of publication. The temporal relevance of a document is called into question. Semantic modules may occupy different temporal spaces for separate documents and occupy different epistemic status for the same document at different times. Stated differently, the boundary object nature of structured authoring artifacts takes into question the ontological status of the audience as the primary consideration of authorship. Audience in this view can be considered an intractable phenomena which does not guide the work of the technical writer in a meaningful direction precisely because the author cannot know a priori who the actual audience will be. In this view the elements and the abstraction principles behind semantic differentiation rather than audience needs to become the principle guiding force in the development of documentation. This shift in focus for the technical writer allows for alternate epistemic and ethical considerations which will be explored further below.

eXtensible Markup Language (XML), The Internet, and Socio-technical Pluralism

At this point in the exploration of digital authoring tools and structured authoring framework the pervasive and total impact of the Internet and web technologies must be acknowledged.

Although beyond the scope of this thesis, the impact of the web and digital multimedia tools on publishing requirements, and therefore on authoring, is undeniable. As mentioned, the backbone on which the above tools of DITA and DocBook run is XML, an extension of the SGML standard which draws on the same theory underpinning page-rank and search infrastructures (O'Reilly & Associates, Inc., 1999).

XML makes use of Document Type Definitions (DTD) to specify and describe data, specifically text data. Because of this we must regard the technical foundation for these authoring tools as descriptions of data and by extension as methods of organizing data within a document. This influences the storage, organization, and delivery of information and documents to an audience. Audience here being an extended definition that includes all stakeholders and professionals who interact with that information for the entirety of the information lifecycle. By extension, we must consider the epistemic impact on the conceptualization and understanding of documentation content by the expanded concept of audience that the work of technical communicators has. Computational tools must make use of the technical machinery of computers and software theory as transportation and distribution networks have made use of the technical and theoretical basis of combustion engines. In this manner, the medium or technology of communication influences the delivery and organization of information across that medium. This in turn affects those who work through and with that medium. Concatenative associations of influence and their affluence upon the profession of technical communication must be taken into account when considering the theoretical foundations of structured authoring. Current language for discourse regarding these topics in technical communication has been found lacking in my opinion, with an emphasis largely on user-centricity, audience, and culpability or conformance to regulatory requirements; discourse in technical communications is largely uni-directional, delivery based, and reactive. The historical sciences, library science, cognitive sciences, and the humanities have more adequate language to reflect the multi-modal, non-linear, and multi-agent systems that this paper frames technical documentation within. Because of this, language from science and technology studies will be introduced to clarify some of the epistemic and phenomenological issues at play.

Bruno Latour's work in science and technology studies has closely explored the tie between work, environment, artifact, and how existence is a modal plural condition. Latour's work in *An*

Inquiry into Modes of Existence, establishes the concept that the mode of existence influences the rationale and methods of veridiction for phenomena, agents, and artifacts of that mode (Latour, B. 2018). For present purposes we can consider that there is a heavy technological and referential component to the veridiction or truth value of a work of structured authoring in this environment purely because of the technical environment that a technical communicator's work takes place in. In Latourian terms, the modes of existence, at minimum, have a cross of reference, legal, technology, and fiction with methods and values of veridiction that are unique to politics and economy. Such modes are identified by spans of discontinuities that entities occupy. We can in this way view "audience" members as existential entities that span a discontinuity in the purveyance of documentation through digital media. The technological status of documentation today primarily obfuscates these discontinuities to the lay reader by methods which are indicative of beings of technology in Latour's work. The continual renewal of requirements and the event triggering processes which follow those renewals are indicative of beings of politics (Latour, B. 2018). Ontological status in Latourian terms is a pluralist status which embraces a relativist metaphysics, for our purposes we can view the ontological status of a document or of a body of content as being the summation of ontologies that interact with the boundary object. By way of Star's notion of boundary object applied to the totality of technical content produced in a structured authoring scenario we are considering a high-density network of ontological relations, actors, actants, and non-linear relationships of authorship and readership. To clarify this system of relations under consideration let us turn to the context of digital authorship which technical documentation now takes place in.

Sociocultural Context of Digital Authorship

This subtle influence on the profession of technical writing and technical communication that XML tools have can be explored partly through the Web interface tool which is used by browser environments to manipulate XML data and JSON files, the DOM. "DOM (Document Object Model) is a standard set of function calls for manipulating XML files from a programming language." (Sumathi, S., Esakkirajan, S., 2007). This is accomplished through an interaction with the rules for an XML document as defined in the schema or DTD as a means to standardize data exchange. Understanding the context of XML documents and data objects, from the perspective of Star, as a boundary object with multiple interpretations and permutations in terms of meaning

and work for different groups or, from the perspective of Latour as a referent artifact which spans multiple discontinuities and thus satisfies multiple modes of existence; allows us to appreciate that the specification or typification of the schema for an XML document to have an internal veracity is an ontological fact. Stated differently, the communicative act of defining a technical specification for information exchange is an ontological or taxonomic distinction which alters the perception of that information for all relata. This places the technical authorship act of semantic differentiation into a public rather than private stakeholder sphere. The choice of segmentation in XML, the manner in which information exchange via XML documents takes place because of those segmentation principles, and the downstream effect of how tertiary documents are interpreted because of how the initial authorship was classified take on very different roles with the above pluralist concepts of ontological status.

This is an important point of emphasis for the theoretical foundation of structured authoring because of the role that structure has in the context in which communication or speech acts take place and are renewed or disseminated in contemporary digital infrastructures. Here we can consider the work of Kuksa and Childs in considering the cultural impact and interpretation of virtual spaces and virtual mediums. The role of virtual spaces in the socio-technical context of a document must be considered when looking at the demands of structured authoring from both an industry perspective of what the requirements for a technical document are, what the psychological affectations of digital media and spaces are, and from an author's perspective of how the audience interprets a document as an artifact; all are *attempts at intelligibility* and *attempts at intelligence,* placing them in the context of Star's criteria for a boundary object.

Sensemaking in Virtual Spaces

Kuksa and Childs make particular note of the peculiarity of virtual spaces as being what they define as a 'fourth space'. "In fourth places, the situated experience of the space is characterized by an engagement of belief, and an experience of metaxis, a splitting of awareness into apparent and actual intent" (Kuksa, 2014). These fourth spaces are defined as being virtual spaces of engagement which have predefined sociocultural identity and inclusion for participation. This is distinct from the obligation of a public space where a variety of beliefs and values may be

encountered, participation itself in a fourth space is a participatory filter of values toward like precepts and beliefs (Kuksa, 2014).

Sensemaking in this context of virtual spaces and of technical documentation is a different mode of existence, and by extension method of veridiction, than that of a purely referential mode of existence which has classically been associated with documentation and reference material (Latour, 2018). The network effect of participation in virtual spaces writ large by contemporary cultures presents a shift in the veracity of digital media, by virtue of sensemaking in virtual spaces for social and personal values, or contrariwise, by virtue of sensemaking in virtual spaces for professional and legal values, there is an extension of the validity of the mode of existence and value system into other artifacts which occupy that space (Star, 2016; Latour, 2018; Kuksa, 2014). This conflation of modes of existence into like objects by what Latour defined as *categorical errors* has far reaching implications for authorship in digital media on a theoretical level. Stated differently, because targeted audiences are consuming content primarily in fourth spaces digitally, rhetorical evaluation and adjudication of content by audience has different psychological priming than traditional reference material–generally considered a second space in these terms. This has deep ethical implications for the uptake of information in virtual spaces and becomes a chief concern of speech act consequents with increased potential for error between intended comprehension and actual comprehension of consumed content.

Authorship, and technical writing in particular, is focused on informing, instructing, or persuading an audience toward a particular outcome of a procedure, belief, or speech act. From this vantage we can view the impact of the technological modes of existence, the technical constraints and context of structured authoring via digital media; these tools become restricted subsets of language subject to different criteria for validation and veridiction. Structured authoring in fourth places and virtual media also then become artifacts of strong rhetoric because of the position and authority documentation occupies in fourth places. This becomes a point of concern when the rate of change and rate of error for professional publication, loose or absent standards for quality, and high potential for unethical profiteering are taken into account through the economic conditions that contemporary documentation and content production take place in. Stated differently, the interpretation of reference data in a virtual space is not the same as

consideration of reference data in a non-virtual context by all audiences and therefore the artifact being interpreted occupies a pluralistic ontology that spans multiple modes of veridiction. This is contrary to the prevailing notion of audience and persona-based writing in technical writing.

The use of restricted technical vocabulary on audience interpretation is not a new aspect of technical communication, it could be viewed as a cornerstone of the profession to maintain communication efforts within a standardized or restricted technical vocabulary. However, what differs here is the structure, representation, and medium of information which is being leveraged to facilitate sensemaking in the audience. Sensemaking in the professional context of technical communication deserves a more empirical treatment than can be considered in this paper. Sensemaking in digital media settings is not a pure set of comprehension and linguistic values but is influenced by the pluralistic ontological status identified in this paper, this includes the concept that professional artifacts, such as online help documentation, are primarily valued by their ability to reduce further cost interactions with customers, not necessarily on the customer's enhanced comprehension of the system or product in question. There are additional considerations for the ethical integrity of attention economies and the evaluation of digital content. Subsequent microeconomic interpretations and studies of digital authorship are needed to consider the extent to which ontological status is intentionally obfuscated or selectively defined rather than considered in total. This is a more abstract notion of *where* sensemaking is taking place and of what is being communicated, with the notion of who here being recognized as largely out of the author's control. This fact of audience as a fictitious guideline for authorship is one of the primary distinctions in this paper. Authorship in this new ontological context should have an emphasis of parity value of information and integrity or quality to the subject matter itself. I am of the opinion that an archivist ethos should take precedence when the audience of documentation cannot be explicitly determined. In short, if one cannot control the audience then the default audience should be considered as the totality of human knowledge writ large and that contributions of any size to the public domain should be scrutinized as strong influences on the future of humanity's knowledge.

The interpretation and delivery of information and documentation through a digital medium is partly driven by the desire for reuse of technical writing (Oliboni, B., & Pozzani, G. 2010). This

echoes the premise of single-source publishing found in the development of DITA and DocBook as tools of structured authoring. However, what stands in contrast to the current presentation of technical writing and authorship is that the ontological status of a document is no longer determined by its author. The open status of content delivery in the public domain, from middleware and automation tools, results in influences on structure and delivery of content that cannot be accounted for during content design, authoring, or even management stages. This is perhaps less true for private and internal documentation but nonetheless plays a role in undermining referential validity of information due to the shared environment of digital content delivery in auxiliary aspects of audience interaction with unrelated content. With the purported value system and value-metric system of documentation as largely being driven by business metrics, the profession has allowed economically viable but informatically diluted circumstances to become the norm. This is an unacceptable position for documentarians to occupy, a similar stance for finance professionals, engineers, or medical professionals would be met with disgust and in many cases legal consequences; why is information and knowledge production any different given its impact on human behavior?

To summarize the theoretical underpinnings of structured authoring presented, the role of markup languages as tools for the interface and exchange of text-based information digitally has had a substantial impact on the development of tools utilized by technical communicators in digital spaces. The context of sensemaking required by audiences of technical authors in virtual spaces via computational tools and digital media has different modes of existence and therefore different methods of veridiction than previous publishing paradigms. In concert, these aspects of structured documentation play into the sociotechnical network within which the work of technical communications takes place, shifting the ontological classification of technical documentation from that of a well-defined singular analytic document to a pluralist ontological status of boundary objects, wherein the clarity or fuzziness of definition becomes audience and context dependent. These three theoretical aspects in consideration of the sociotechnical environment of The Internet frames a very different motif of what aspect of documentation is being structured and what reach standards bodies actually have in the profession. This network of considerations has strong implications for the direction of technical communications and structured authoring due to the appreciation of variables and the value which is purported or

realized in documentarian practice. The efforts for the categorization and management of data in an increasingly documented world of virtual spaces made real must have more robust awareness of its impact, integrity, and actual consequence for actors that are determined to have violated public trust in this sphere.

Deterritorializing Audience in Technical Writing

When looking at the reality of authorship in virtual spaces and the relationship between information publication, information availability, and information access through virtual spaces via internet browsers, I call into question the validity of the audience as the primary criteria of authorship. When the ontological status of documentation is viewed in plural as an artifact that will both be read by human intelligence and by machine intelligence then we must consider, by extension, the role of that document as input for machine authoring processes and the subsequent readership of those secondary media. The reality of machine generated content and of automated content delivery and management, casts artifacts in the public domain in a more consequential light, that being that documents are no longer primarily read by initially intended human audiences but will overwhelmingly be read, interpreted, used and reused more by machine author audiences. Documentation that will reach the public domain must be viewed as data, not as document. Contemporary use of XML and semantic segmentation of documentation in technical communications, content development, and middleware tools supports this view. The division of syntax and semantics in documents via elements or tags have made possible rapid machine interpretation of vast amounts of data across a wide variety of virtual spaces. I am not advocating a return to pre-computational methods. These tools have simply reconfigured the ethical and professional consequences of technical writing and of publishing information in the public domain to such an extent that a revision of responsibilities, controls, and publication processes must be developed.

Scope of Impact for Documents in the Public Domain

When we view the public domain as a vast corpora of semantic markup, the role of single-source authoring as the formation of contributions to hyper-boundary objects or collections which will be read many times over to form like associations or, even as the basis to form new artifacts completely outside of the original authors awareness, we absolutely must consider a different

ethical domain than contemporary concepts of copyright and culpability account for. This idea of digital media as having a hyperpluralist ontological status demands a different depth of ethical considerations because the reuse of the content is no longer in the control of the author, ownership in the public domain has become an oxymoronic concept. Once published there is no guarantee that something may be completely removed, there is no control mechanism for complete updates. We now face the inverse issue to reference and plagiarism; we face a plague of publications with no strong claim of quality or veracity due to a historically low barrier to entry and no assignment of consistent quality control or evaluation. Although legally there may be ways to pursue and alter publication after the fact, the rate of reproduction and the extent of sampling cannot be known due to the current status of web crawler access to public domains for site indexing in search engines and scope of potential authors with access to publication. Our digital infrastructure fundamentally relies on access to the semantic structure and content of published digital documentation to purvey access to audiences from content providers. Authors in this sense can have an audience in mind for the development and organization of documentation but it is an idealized nominal audience, the actual real audience will be substantially different, not might be but will be. It is because of this that I feel the attention toward a particular audience is a vestigial approach to structured documentation and technical communications as a profession.

Documentation and Modes of Existence in Virtual Spaces

In Latourian terms, documentation for the audience relies on a fictional mode of existence, the narrative of the artifact must appeal to the narratives of a particular audience–marketing as a discipline successfully leverages this concept to this day. However, when considering the veracity of a virtual space and the environment in which documentation is experienced by the audience, the pluralistic boundary object nature of documentation delivered over digital spaces, and the non-linear temporality of documentation, we cannot take documentation as occupying merely a fictional mode of existence. There will always be secondary audience members or casualty audiences who were not intended recipients but who nonetheless are subjected to the content in some manner. At minimum, we must consider a technological dimension, a referential dimension, and a political dimension because the interactions are predominately taking place in a fourth space where involvement and occupation of space is primed by belief activation and

identity inclusion. Realistically, an author cannot feasibly write for such an ambiguous audience, consider so many conflicting factors, and still produce economically viable goods in a reasonable timeline. Therefore, a detailed investigation of each mode of existence is not necessary as such a methodology would not aid the technical writer in the development of the profession. What is required is an acknowledgement by the profession that the audience is an idealization which is reliant on simplifications of context toward a generalized reader. This has important consequences philosophically. In Latourian terms, this is a categorical error, the method of veridiction for technical documentation has now departed from the modes of existence which it occupies. This means that the basis for the measurement of quality in technical documentation has become contrived. We, as technical communicators must then return to the basis of the ontological status of our work and of documentation in this digital era.

To reassess the ontological status of documentation we must suspend certain contexts. To begin, let us maintain the ontological pluralism of the non-linear notion that the total audience of a document cannot be known by an author or writer of a document. The artifact will have a life of its own well outside the author or publisher's control after the publication of a document to any kind of public domain. Likewise, the interpretation of this document as a boundary object also cannot be controlled by the author. These are aspects of exteriority which do not inherently alter the document's properties. What then are the aspects of interiority that can be evaluated as a means of determining quality? If the aim of the author is to inform, and the aim of the document is to be published, what then are we informing? Taking into consideration that we are essentially no longer regarding the audience as a viable filter for what is to be informed we can reinterpret the speech act and the audience. The audience in this sense is instilled with the responsibility of sorting through and determining necessary information themselves. If we consider the tool of semantic markup and empower the user to make use of such a tool then the author is freed to focus on the internal quality of their documentation. In this frame of consideration the technical writer returns to becoming a subject matter expert. If the audience fades from view and is granted the elevated position of an informed audience, or super-user to use user interaction terms, then we can conceptualize a different ontological status of documentation as representative of knowledge. Rather than a determinate for the delivery of knowledge, the representation of knowledge becomes the point of focus.

A Sociotechnical Constructivist View of Cognition and Documentation

There are several practical issues with the above presentation which may seem historically ignorant of the development of technical authorship away from subject matter experts toward the user-centric school of technical communication. Indeed, I acknowledge that the presentation of the above position is in direct opposition to the school of user-centric documentation. I feel that such an opposition is important in light of the pluralism and non-linearity which artifacts now have in digital contexts. User-centric documentation has the fallacy of audience at play, although the delivery of documentation toward a specific audience for a specific purpose is a model for delivery that has been made economically viable, it does not take into account the full context or life-cycle of that documentation. I feel that the user-centric model is an abuse of semantic and structured authorship. Targeted delivery makes use of repositories of semantically marked elements which can then be filtered and delivered for specific aims, however, the scope of information in this regard suffers from specificity. This specificity of information shifts the dynamic of information to a very limited use case. Reuse of information in this regard, an established aim of structured authoring in technical documentation, does not imply any regard for the quality of information delivered, the quality of documentation published, or the efficacy of that documentation in informing the audience. A life cycle view of documentation and of the audience gives a different value system from which to judge the efficacy of documentation in the goal of informing. Stated differently, to consider the developmental life cycle and cognition of the reader as an aspect of the documentation life cycle reframes ethical considerations involving documentation in the public domain.

Considerations of scale for this distinction are important. In isolation, or even in consideration of an industry or of a business quarter, just-in-time delivery of user-centric documentation make sense and can be ethically executed. However, when we interpret the dual role of documentation as publicly available data in a pluralistic audience, side-effects from lack of information quality control, inappropriate or inaccurate semantic markup of documentation, and temporal relevance of documentation take on a more complicated nature. This delay of documentation use by third parties, human or non-human, introduces more general concerns about the ethics of the technical communicators' work in a broader context. Particularly, from a historiographic context, the sum effort of technical communicators in the production of external documentation and of the

production of public domain documentation as digital media, carry implications for the societal or meso-level processing and ontological classifications of documentation content. In this regard, there is an ethical obligation of documentarians to produce works that maintain a high quality of documentation not only by an internal standard but also an obligation that public facing information with an intent to inform be of a quality standard akin to educational material and of academic rigor. This is because of the role that web crawlers, scrapers, and semantic markup plays in the infrastructure of information delivery in 21st century contexts. Not only the dissemination but also the formulation of documentation or, the analyses required to instantiate the production of documentation, are affected by the purveyance of information via digital infrastructures. The semantic markup of these documents becomes the basis for analysis and delivery of information in the public domain, this guides the formation of student opinion, and crucially affects the veracity and scope of people's weltanschauung based on their experience and selective exposure to information through these mediums.

These environmental changes in the work of the technical writer are arguably ancillary factors outside of the control of the technical communicator and are tertiary concerns for stakeholders surrounding the production of documentation in business settings. I am of the opinion that it is precisely this laissez faire approach to responsibility for quality of information in the public domain that has led to some of the more tenuous social realities that are the hallmark of 21st century public discourse and politics. It is through a commitment to quality in all areas of interaction that the profession of technical communications can contribute to a higher quality information standard in the public domain. I feel that the appropriate ethical response to uncertainty in the information quality of the public domain is a return to subject based integrity and subject matter expertise. Although competitive advantage does exist within the concepts of time-to-market and customer or user-centric approaches to documentation and information needs, the ethics of motive are questionable at best and generally follow a tragedy of the commons whereby the resource of many is depleted for the benefit of few actors who currently have advantage to exploit the collective resource. The resource being exploited is that of the domain of public understanding or the scope of public knowledge and quality of public knowledge. With lower barriers to production and lower barriers to entry for publishing in the rise of the Internet and digital tools there has been a mass proliferation of unverified and

unvalidated information regarding nearly everything that can be searched. Although this historical breakthrough in access to both information and publishing tools has given voice to previously suppressed peoples and views, it has also diluted quality of information, ontological integrity, and verifiability of information. This has led to a substantial reduction of trust in public discourse, in academic discourse, and in the validity of many bodies of knowledge which are integral to our digital and social infrastructures.

I feel what is called for is a return to professional ethics and standards which require extensive proofing of documentation that will be available in the public domain as an assurance to the quality of information available to future generations. Because of the scope of technical communications, localization, and documentarian labor, the technical communicator profession stands as a crucial profession to instill stricter requirements of information quality and integrity as they are often the last in-line for the publication of documentation. I see the future of technical writers and documentarians as participating in an emerging ecosystem of knowledge workers which maintain and develop information domains which will then shape future worldviews. The consequences of this domain cannot be overstated because they occupy an intractable scenario rooted in the wicked problem of future education outcomes (Rittel & Webber, 1973). Given this high-risk aspect of public information, a rallying call for knowledge workers at large is needed. This should include library science professionals, social scientists, data professionals, technical communicators, project managers, academics, and independent contributors to the public domain as well as dedicated auditors and review professionals.

Engineering and accounting professions offer a roadmap toward the level of accountability which needs to be required of future information or knowledge work professions which develop publications for the public domain. Although this is a broader concept and conversation than appropriate for this article, structured authoring is one of the cornerstone technologies which enables both the development and auditability of the public domain. The work of developing, mapping, assessing, and disseminating information in the public domain should be a chief concern of the public and of governance bodies. This is a high abuse potential area similar to financial reporting. History has numerous examples of controlled information networks affecting the decision making of social groups. Particularly, the history of propaganda and the efficacy of

marketing in the 20th and 21st centuries speak to the danger and concern for unchecked publication of information in a global society moving forward. There must be a concerted effort toward information integrity and specifically toward ontological verification in the public domain moving forward. In the contexts of conflated and conflicting interests where public domain publications have the potential for monetary gain, there is a strong need not only for review and oversight after the fact but a strong need for the development of more stringent professional requirements, review, and recourse for bad actors in information professions. It is for this reason that I put forward an initiative to technical communicators and information sciences professionals to organize toward a more fiduciary stance of information quality with recourse in line with professional engineer standards and consequence of licensure loss and civic penalties.

What may seem drastic or unrealistic at this stage of public governance in open market economies may be viewed in the future as the bare minimum. The history of public health initiatives, building codes and civil engineering, and safety regulation in many industries is indicative of this. Given the conflicts that are arising with intellectual property, academic integrity, and the value of data in 21st century economics, a concerted effort is needed to unify against the deterritorializing market forces that currently dilute information quality and integrity. I feel that the tools of documentarians and technical communicators have a formative role to play in fielding, validating, and auditing information quality to deliver not only a profitable economic product but deliver the public a quality ontological product which contributes to the integrity, longevity, and value of human knowledge for generations to come. The short sighted exploitation of public access to information for short term gains should be seen as having similar consequences to commercial trawl fishing, strip mining, and hazardous waste dumping. If information informs behavior, which the ongoing economic success of marketing attests to, then the consequence of information availability on human behavior should be viewed as affecting human health. The same scrutiny that we give medical alteration of human environments should be given to the information landscapes that people are exposed to full stop. This mediation must begin with the professions which purvey information acknowledging its role in human behavior, cognition, and worldview, and move toward an acceptance of the immense responsibility for the public domain of information which they profit from. The public domain and information

availability is a public resource which is currently under exploitation without regard for its longevity or quality. Part of a sustainable economy is a public domain of information that has been cared for, attended, and curated for integrity and quality. The receiving public should not be regarded as the lowest common denominator but instead given the highest quality information available and given the responsibility to determine for themselves. Information should educate, inform, and elucidate the receiver toward the frontier of human understanding.

References

Akinci, U. (2014) 3 Disadvantages of Writing Documents with DITA and Structured Authoring, Technical Communications Center, accessed: 01/10/24. https://technicalcommunicationcenter.com/2014/10/03/some-facts-about-shifting-to-dita-and-structured-authoring-platform/

daSilva, A. (2022) From CMS to Intelligent Content Hub: Why the need for a change?, Semantics Vienna 2022, accessed: 01/10/24. https://2022-eu.semantics.cc/cms-intelligent-content-hub-why-need-change

IBM. (n.d.). *Dita 101 guide - Leveraging the DITA standard for content creation, control and delivery*. DITA 101 Guide Enterprise Content Management Solutions Better decisions, faster. https://public.dhe.ibm.com/software/info/television/filenet/tmp/IBM14042USEN.PDF

Priestley, M., & Hackos, J. (Eds.). (2005, May 5). Oasis darwin information typing architecture (DITA) language. https://docs.oasis-open.org/dita/v1.0/dita-v1.0-spec-os-LanguageSpecification.pdf

[DITA-v1.3-part3-all-inclusive]
Darwin Information Typing Architecture (DITA) Version 1.3 Part 3: All-Inclusive Edition. Edited by: Robert D. Anderson and Kristen James Eberlein. 17 December 2015. OASIS Standard. Latest version: http://docs.oasis-open.org/dita/dita/v1.3/dita-v1.3-part3-all-inclusive.html.

Kuksa, I. (2014). *Making sense of space: The design and experience of virtual spaces as a tool for communication* (Ser. Information Professional Series). Chandos Publishing.

Latour, B. (2018). *An inquiry into modes of existence: An anthropology of the Moderns*. Harvard University Press.

Martinez-Ortiz, Ivan & Moreno Ger, Pablo & Sierra-Rodríguez, José & Fernández-Manjón, Baltasar. (2006). Using DocBook and XML Technologies to Create Adaptive Learning Content in Technical Domains.. International Journal of Computer Science & Applications. 3.https://www.researchgate.net/publication/26621368_Using_DocBook_and_XML_Technologies_to_Create_Adaptive_Learning_Content_in_Technical_Domains

OASIS-tcs. (n.d.). *Overview of DITA - Information Typing*. Information typing. https://dita-lang.org/dita/archspec/base/information-typing

Oliboni, B., & Pozzani, G. (2010). Part 1: Uncertainty in XML. In Z. Ma & L. Yan (Eds.), *Soft computing in XML data management: Intelligent Systems from decision making to data mining, web intelligence and Computer Vision* (Vol. 255, Ser. Studies in Fuzziness and Soft Computing, pp. 1–106). essay, Springer-Verlag.

O'Reilly & Associates, Inc. (1999). *Chapter 1. getting started with SGML/XML*. Getting Started with SGML/XML. https://www.oasis-open.org/docbook/documentation/reference/html/ch01.html

O'Reilly & Associates, Inc. (1999). *Chapter 5. Customizing DocBook*. Customizing DocBook. https://www.oasis-open.org/docbook/documentation/reference/html/ch05.html

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. Policy Sciences, 4(2), 155–169. https://doi.org/10.1007/bf01405730

Susan Leigh Star, 2016. "The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving", Boundary Objects and Beyond: Working with Leigh Star, Geoffrey C. Bowker, Stefan Timmermans, Adele E. Clarke, Ellen Balka

Sumathi, S., Esakkirajan, S., (2007) 13.5 XML, *Fundamentals of Relational Database Management Systems*, Studies in Computational Intelligence 47, Springer-Verlag, Berlin Heidelberg,

Walsh, Norman (2020) What is DocBook?, DocBook.org, accessed: 12/12/23, https://docbook.org/whatis