

METADATA ENRICHMENT FOR DIGITAL HERITAGE: USERS AS CO-CREATORS

Getaneh Alemu (*PhD*)

Solent University, Southampton, U.K.

Getaneh.Alemu@solent.ac.uk

1. Introduction

Grounded on a constructivist (interpretivist) rather than objectivist philosophical framework (Alemu, Stevens, Ross, & Chandler, 2014; Charmaz, 2006), this paper argues that the “meta” in metadata, as well as the nature of information objects in digital heritage it describes, lends itself to diverse interpretations, making the case for a mixed metadata approach where both metadata experts as well as users enrich information objects. To do this, as argued in this paper, users need to be considered as proactive metadata creators rather than passive consumers. User co-creation does not make information professionals obsolete – instead, as Lankes argues, it expands their mission as change agents and facilitators of knowledge creation rather than mere custodians or stewards of books and cultural artefacts (Lankes, 2016). Lankes (2016) identifies key principles and concepts where librarians should be focusing in the 21st century. Lankes notes a shift from the current collecting and lending model of libraries to platforms for knowledge sharing, knowledge creation, enabling conversations, giving voice to minorities, allowing transformative social engagement and serving as agents for radical positive change. This shift requires social and political organisational changes and re-conceptualisation of existing models, tools and practices (Alemu, Stevens, Ross, & Chandler, 2012; Gartner, 2016).

Informed by several years of research on metadata (Alemu et al., 2012; Alemu, Stevens, & Ross, 2012; Alemu, Stevens, Ross, & Chandler, 2014; Alemu & Stevens, 2015), this paper presents the principle of metadata enrichment for digital heritage. It explores the potential benefits of a shift from passive to proactive user involvement in metadata co-creation (Alemu et al., 2012; Casey & Savastinuk, 2006; Lankes, 2016; Miller, 2005; O'Reilly, 2005; Tammaro, 2016). As digital information becomes ubiquitous and pervasive, the need for forging strategic partnership within and beyond libraries, archives and museums so as to devise new working practices, develop communities of learning and set quality assurance mechanisms become vital (Tammaro, 2016). For example, the Library of Congress, the largest library in the world, owns more than 164 million information objects (print and digital), followed by the British library which provides access to more than 150 million items. A more recent digital undertaking is Europeana.eu, which is a gateway to more than 51,533,591 artworks, artefacts, books, videos and sounds from across Europe. Other initiatives include the Digital Public Library of America (which provides access to more than 20,597,354 items from libraries, archives, and museums), Project Gutenberg (over 56,000 free and public domain e-books), World Digital Library (more than 19,147 items about 193 countries between 8000 BCE and 2000) and the Internet Archive (more than 15 petabytes of webpages) – are all facilitated, mediated and enabled by an ever changing set of computer and web technologies (Cameron & Kenderdine, 2007; Kalay, Kvan, & Affleck, 2008).

Cultural heritage institutions such as libraries systematically organise their collections using standards, systems and tools so as to enable users find, discover, explore, browse, access and use the data, information or knowledge contained in these information objects (Svenonius, 2000). Information organisation in general and metadata in particular is central to the findability and discoverability of information objects in cultural heritage institutions (Alemu & Stevens, 2015; Gartner, 2016; Svenonius, 2000; Wright, 2007; Zeng & Qin, 2016). The metadata currently in use in most institutions tends to be objectivist, top-down and expert-created (Alemu & Stevens, 2015). However, whilst these metadata

standards serve an important purpose by providing the underlying structure and required interoperability, they fail to rise to the challenges presented by changing technologies, size and diversity of collections as well as changing user expectations (Shirky, 2005; Tammara, 2016; Weinberger, 2007; Weinberger, 2014).

This paper espouses the concept of metadata enrichment through an expert and user-focused approach to metadata creation and management. To this end, it is argued the Web 2.0 paradigm enables users to be proactive metadata creators. As Shirky (2008, p.47) argues Web 2.0's social tools enable "action by loosely structured groups, operating without managerial direction and outside the profit motive". Lagoze (2010, p. 37) advises, "the participatory nature of Web 2.0 should not be dismissed as just a popular phenomenon [or fad]". Carletti (2016) proposes a participatory digital cultural heritage approach where Web 2.0 approaches such as crowdsourcing can be used to enrich digital cultural objects. It is argued that "heritage crowdsourcing, community-centred projects or other forms of public participation". On the other hand, the new collaborative approaches of Web 2.0 neither negate nor replace contemporary standards-based metadata approaches. Hence, this paper proposes a mixed metadata approach where user created metadata augments expert-created metadata and vice versa. The metadata creation process no longer remains to be the sole prerogative of the metadata expert. The Web 2.0 collaborative environment would now allow users to participate in both adding and re-using metadata. The case of expert-created (standards-based, top-down) and user-generated metadata (socially-constructed, bottom-up) approach to metadata are complementary rather than mutually-exclusive. The two approaches are often mistakenly considered as dichotomies, albeit incorrectly (Gruber, 2007; Wright, 2007) .

This paper espouses the importance of enriching digital information objects with descriptions pertaining the about-ness of information objects. Such richness and diversity of description, it is argued, could chiefly be achieved by involving users in the metadata creation process. This paper presents the importance of the paradigm of metadata enriching and metadata filtering for the cultural heritage domain. Metadata enriching states that a priori metadata that is instantiated and granularly structured by metadata experts is continually enriched through socially-constructed (post-hoc) metadata, whereby users are pro-actively engaged in co-creating metadata. The principle also states that metadata that is enriched is also contextually and semantically linked and openly accessible. In addition, metadata filtering states that metadata resulting from implementing the principle of enriching should be displayed for users in line with their needs and convenience. In both enriching and filtering, users should be considered as prosumers, resulting in what is called collective metadata intelligence.

Keywords: Cultural heritage, metadata, standards, Web 2.0, metadata enriching, metadata enrichment, metadata linking, metadata filtering

2. Metadata – does it matter?

Whilst the term metadata became very popular with the development of the web, especially with the creation of the Dublin Core metadata standard in 1995 in Dublin, Ohio, the notion of cataloguing goes back to the history of libraries itself (Wright, 2007). Library pioneers including Anthony Panizzi, Charles Cutter, Melville Dewey, S.R. Ranganathan and Seymour Lubetzky (Svenonius, 2000; Wright, 2007) played a major part in devising mechanisms to help users find books in a library. For these pioneers, the library catalogue has one essential function – to help users identify a particular book by title, author or subject and to do so by saving their time. Such individual contributions were further developed and became standardised by international initiatives such as the Anglo-American Cataloguing Rules (AACR), Paris Principles, Machine-Readable Cataloguing (MARC), Resource Description and Access (RDA), Library of Congress Subject Headings (LCSH) and a host of other

specialised agencies. These efforts led in the development of a plethora of metadata standards. For instance, with the development of digital collections, new descriptive metadata standards began to emerge – these include Metadata Object Description Schema (MODS), Dublin Core (DC), Metadata Encoding and Transmission Standard (METS) and Learning Object Metadata (LOM). Granted, these standards-based approaches have an enduring history and they have helped to guide resource description (cataloguing) and access (resource discovery) practice in libraries. However, the introduction of the collaborative web (Web 2.0) and recently the development of Linked Data (Web 3.0), pose scalability challenges for these standards to effectively describe the ever-increasing size of digital collections as well as incorporating the changing user requirements.

Metadata is ubiquitous - it is all around us. Your name, age, title, height, weight, hobbies, address, location, browsing history, comments, tags, likes, shares and ratings all constitute your personal metadata. Likewise, the attributes of a book such as title, author, format, genre, location and description are book metadata. The “meta” in “metadata” means “beyond”, hence metadata is often defined as data about data. It is the ‘about-ness’ of a thing. Gartner (2016) notes that “human knowledge is built on ‘about-ness’ and it is through our interpretation of what the world is ‘about’ that most of our intellectual endeavours are based. Without metadata we cannot have knowledge”. The US National Information Standards Organisation (NISO, 2004) defines metadata as “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”. By providing descriptive (author, title and subject), administrative (identity, provenance, rights, contextual and technical) and structural information (relations with other information objects), metadata plays an important role in cultural heritage institutions to support the findability and discoverability of cultural information objects by users. The solution to the information overload problem epitomised by the web, as Weinberger (2007, p. 13) notes, is to create yet more information about the information, i.e. more metadata. As Zeng and Qin (2008, p. 3) note, metadata is “the invisible hand” that enables effective information organisation. Information organisation is powered by metadata. As Weinberger opines “metadata liberates knowledge”. In recognition of this metadata importance, significant investments have been made to design metadata standards by a number of national, multinational and international initiatives in order to describe collections and to enhance the findability and discoverability of information objects (Alemneh, 2009; Chan & Zeng, 2006; Gartner, 2008; Gartner, 2016; Lagoze, 2001; Lagoze, 2010; Zeng & Qin, 2016).

Metadata is the *raison d'être* for libraries, archives and museums as it constitutes one of their core functions, i.e., ensuring print and electronic resources are findable, discoverable and usable by users – hence justifying the return on investment. Metadata is also central for the preservation and access of cultural heritage objects. Unless it is systematically organised through relevant metadata and indexed, it is impossible to access and determine relevancy of the multi-billion pages of information on the web (Boulton, 2014). Information discovery including search, relevancy ranking, and faceted refinement and grouping related resources all relies on metadata (Varnum, 2016).

The metadata creation and enriching process happens at various stages of the information resource life cycle. In theory, metadata creation and enhancement (metadata enriching) is a continuous process and it involves authors, publishers, suppliers, librarians and users. Unfortunately, in current practice, users are not allowed to add metadata. This is partly due to assumptions that user metadata is devoid of good quality, its management is problematic and there are no reliable platforms/tools to handle it. As the size of collections in libraries grows, cataloguers increasingly find it difficult to describe every information resource in a manner that reflects the terminologies as used by users so as to therefore achieve optimal discoverability. With the development of the Web 2.0 paradigm, new opportunities arise to involve users in the metadata creation process.

As regards to metadata, a painting in a museum essentially lends itself to various interpretations. People might objectively agree with such attributes as name of artist, type of painting, dimension, material

used, canvas, time of painting and current location. However, when it comes to the about-ness (subject) of a painting, there are probably as many subjective interpretations as the number of viewers. In contemporary metadata approaches, it is the librarian, archivist or the metadata expert who adds and maintains the metadata for these objects. The problem is that these experts might not anticipate the diverse interpretations inherent in users. Hence, they fail to adequately describe these information objects. The attempts to use controlled vocabularies solve only a part of the problem. There is an apparent disparity between controlled terminologies and terminologies used by users (Alemu & Stevens, 2015).

Furthermore, human beings by nature do not always agree on a single about-ness, interpretation and classification of things (Shirky, 2008; Weinberger, 2007). Classification and metadata are affected by socio-cultural, linguistic and political factors hence metadata (Bowker & Star, 1999). Whilst people, places, objects and events are real objective (verifiable) facts, the metadata that describes them is a social construct hence could be intensely subjective (Gartner, 2016). As Gartner contends metadata is an approximation to the things it represents, suggesting thus a social constructivist view point. Gartner adds that “the shape of metadata is designed by human beings for a particular purpose or to solve a particular problem, and the form it takes is indelibly stamped with its origins. There is nothing objective about metadata: it always makes a statement about the world, and this statement is subjective is what it includes, what it omits, where it draws its boundaries and in the terms it uses to describe it”.

In contrast to librarian/archivist-defined (standards-based) metadata, the Web 2.0 paradigm (O'Reilly, 2005) provides new opportunities for metadata creation and utilisation. For example, search for the “The School of Athens” in the Europeana.eu. As it stands now, the search result returns the object entitled “La escuela de Atenas – Cuadro with metadata details such as Creator: Raphael”; Classifications: Cuadro, Pintura, Escuela Española, Lienzo , Historia Antigua, Álvarez Enciso, Domingo, Raphael and Painting; Size:Sin marco: Altura = 62 cm, Anchura = 88 cm; Language:spa; Creation Date: 1701=1800; Rights: ©Ministerio de Educación, Cultura y Deporte, ©Museo de la Real Academia de Bellas Artes de San Fernando Fotografía, ©Museo de la Real Academia de Bellas Artes de San Fernando; Is Part Of: Museo de la Real Academia de Bellas Artes de San Fernand; First Published In Europeana: 2014-02-19; Last Updated In Europeana: 2016-04-21. In addition, it includes metadata elements such as Provenance, Identifier, Institution, Provider and Providing Country. Given Europeana aggregates metadata from multiple participating institutions, one wonders if this is the only instance of “The School of Athens” available? If not, could not an enriched metadata have resulted in more results? In FRBR lingo, are not there more Works, Expressions, Manifestations and Instances (WEMI) relating to this particular painting. Is standards-based metadata sufficient to enrich these objects? Could these objects be enriched through user-generated metadata? What about providing more details about the painting itself such as who are the people depicted in the painting? What if users are allowed to strike conversations about this painting and that content in turn could be used for educational purposes?

Cultural heritage institutions include galleries, libraries, archives and museums – often abbreviated as GLAM. For example, the European Commission dedicated 2018 as the year of cultural heritage which encompasses tangible, intangible and digital (EU, 2017). The British Museum, the British Library, the National Archives and the renowned British galleries all exert focus on ensuring the accessibility and preservability of their collections – both physical and digital. Howard (2015) critiques the re-convergence, skills required and roles of galleries, libraries, archives and museums (GLAMs) as epistemic infrastructures for today’s knowledge economy and offers recommendations for changing Howard argues “the knowledge economy can be directly linked to galleries, libraries, archives and museums and the role that they have played throughout history. These institutions – and the information professionals who work in them - by preserving human knowledge, have assisted in advancing human knowledge” (p.39). The services GLAMs provide depend on their information organisation functions which in turn relies on the creation, application, management and use of metadata (Lim & Liew, 2011;

Zeng & Qin, 2016). These institutions use a plethora of standards to describe, provide context and manage their collections. Some of the metadata standards in use include CIDOC Conceptual Reference Model (CRM), Dublin Core (web), VRA Core (museum and visual), EAD (archives), Marc21 (libraries), MODS (simplified MARC), IPTC Core (photo), CDWA (Categories for Description of Works of Art), OAI (Open Archival Information System), METS (structure), PREMIS (preservation), ISO2709 (MARC), XML, HTML5, microdata, RDFa, RDF/XML, JSON, RDA, AACR2, CCO, DACS, OAI, Linked Data, FRBR, LCSH (Library of Congress Subject Headings), AAT (Art & Architecture Thesaurus), TGN (Getty Thesaurus of Geographic Names), DDC, LCNAF and ISO 639-2 (ISO code for languages (Zeng & Qin, 2016).

This is because galleries, libraries, archives and museums (see Howard, 2015) assume a professional mandate to collect, curate, manage and provide access to an ever-growing collection of digital and physical cultural heritage content. Without complete, quality and enriched metadata, the information objects that reside in these institutions remain invisible to intended users – which has a direct bearing on usage and return on investment. The Europeana digital cultural initiative is one example of GLAM re-convergence where member countries commit significant investment in its realisation. Europeana amalgamates an extremely diverse collection of historical documents, arts, paintings, music, photography, fashion and maps on a plethora of subjects, events, places and language. One may ask how such diverse a collection as Europeana is described and enriched with metadata. Even more relevant is are users (citizens) involved in adding metadata and making sense of the content so that its findability, discoverability and usability is enhanced? The sheer size and diversity of digital cultural heritage content suggests the scale of the metadata challenge to adequately describe, enrich and add value to it.

The plethora of galleries, libraries, archives, and museums across the world are not mere collectors and storehouses of books, art work, paintings, music, video and manuscripts but they are places, physical and increasingly virtual) for scholarship, learning, innovation, creativity, entertainment and information (Lankes, 2016). Digital technology brought a significant shift in how cultural heritage information objects are created, represented, mediated, disseminated, preserved and accessed (Cameron & Kenderdine, 2007; Hedstrom et al., 2003; Kalay et al., 2008; Tammaro, 2016). New media and technologies provide an opportunity to preserving, expanding access and use of cultural heritage. It encompasses both born digital and digitised representations of physical information objects (Kalay et al., 2008).

Digital heritage now constitutes a central function of cultural heritage institutions. UNESCO's Charter on the Preservation of Digital Heritage, adopted on October 15th, 2003, acknowledges the unique nature, value, access and preservation requirements of digital resources which include text, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. The Charter also acknowledges the unique, dynamic, fluid and ephemeral nature of digital content which requires unique methods of production, maintenance and management (UNESCO, 2003). The Charter further states that "the purpose of preserving the digital heritage is to ensure that it remains accessible to the public. Accordingly, access to digital heritage materials, especially those in the public domain, should be free of unreasonable restrictions. At the same time, sensitive and personal information should be protected from any form of intrusion" (UNESCO, 2003). Cognisant of the importance of digital heritage, UNESCO calls on all stakeholders including "hardware and software developers, creators, publishers, producers and distributors of digital materials as well as national libraries, archives, museums and other public heritage organisations in preserving the digital heritage; develop training and research, and share experience and knowledge among the institutions and professional associations concerned, and encourage universities and other research organisations, both public and private, to ensure preservation of research data" (UNESCO, 2003). Digital heritage bridges the past to the present and into the future – thus the need to invest in its management, organisation and preservation. As Cameron and Kenderdine (2007) note digital technologies afford new

methods and tools to create and access the intellectual capital of the knowledge society. But the relationship between new media and cultural heritage crosses over multiple boarders - “philosophical, historical, social, artistic, biological, geographic, and the linguistic” (2007).

The CIDOC Conceptual Reference Model (CRM) is a widely adopted metadata ontology for the cultural heritage domain. The standard consists of 81 classes and 132 unique properties (Lourdi, Papatheodorou & Doerr, 2009). Other standards in use include Getty Thesaurus, Art and Architecture Thesaurus (AAT), Categories for the Description of Works of Art (CDWA), Encoded Archival Description (EAD) and Visual Resources Association (VRA). Europeana designed and developed the Europeana Semantic Elements. Metadata standards and protocols provide the underlying structure, granularity, semantics and interoperability (Alemu et al., 2012). It helps to ensure provenance (authority, source, change history and ownership), integrity and authenticity of cultural heritage information objects. In the presence of several metadata standards in the cultural heritage domain, metadata interoperability is a critical issue (Lim & Liew, 2010). In addition, standards help to bring about metadata quality and consistency (Lim & Liew, 2010). In addition, there is a growing challenge for metadata experts to adequately describe and enrich these information objects with about-ness metadata – calling for the involvement of several metadata actors to describe, ascertain provenance, provide context and make sense of a cultural information object.

It can be argued that whilst expert (librarian or archivist) defined and standards-based metadata captures some aspect of information objects, it lacks the richness that could be gained by involving users. Such metadata diversity increases the likelihood of conforming to the multitude of perspectives and interpretations which is inherent in users. Cultural heritage objects in its very nature lend themselves to various interpretations – hence making the case for metadata diversity and richness. Whether it is called social cataloguing, crowdsourcing, folksonomy, folksonology or democratic classification, as Gartner (2015) argues, the web 2.0 approach to metadata amounts to a paradigm shift. Lankes, Silverstein and Nicholson (2007) exclaims that “knowledge is created through conversation” suggesting the changing role of libraries as facilitators of conversation. Lankes, Silverstein and Nicholson propose the need for libraries to embrace Web 2.0 approaches not as a peripheral activity but as a core function.

Carletti (2016) links participatory digital heritage with the broad notions of participatory government, citizen engagement and a constructivist pedagogic approach where learners are encouraged to be actively engaged in the co-construction of their learning. Carletti (2016) notes that “crowdsourcing is only part of a wider context exploring innovative, inclusive and cohesive trajectories of engagement. [...] The collaboration between cultural and education institutions and the public can represent the platform to scaffold citizen scholarship, to assure engaged access to digital heritage, and to promote a culture of participatory heritage”. Metadata enrichment is therefore an important concern not just for cultural heritage experts but also for local, national and regional governments where significant investment is being done in digitisation projects. In the end, return on investment is measured on how much these collections add value to their users – for which metadata plays an essential function through enhancing access and discoverability.

The argument for socially-constructed, user-generated metadata or metadata co-creation is not a new phenomenon (Alemu et al., 2012; Casey & Savastinuk, 2006; Maness, 2006; Miller, 2005). For example, Smith-Yoshimura and Shein (2011) wrote that “traditionally, staff at libraries, archives, and museums (LAMs) create metadata for the content they manage. However, social metadata—content contributed by users—is evolving to both augment and contextualise the content and metadata created by LAMs. This paper focuses on the principle of metadata enriching and its application for digital heritage.

3. The Web 2.0 paradigm

In 2007, Farkas offered a book length exploration of the role of social media for libraries. The focus then was on Web 2.0 technologies such as social networking, wikis, online communities, blogging, social bookmarking, podcasting, screen-casting, vodcasting, Second Life and video games to support existing library functions such as cataloguing, circulation, referencing, services marketing and outreach (Farkas, 2007). As Vander Wal (2007) defines folksonomy as a collaborative approach to tag information resources with keywords which can aid the tagger for later retrieval. As Andrew Lee argues the Web 2.0 offered a substantial change in the way information is created, edited and disseminated. “Wikipedia initiated something new and unprecedented - it demonstrated collaborative accumulation of knowledge was not only feasible but desirable. Its neutrality policy, combined with a global team of volunteers, helped make Wikipedia not just a clone of existing encyclopaedias, but an encyclopaedia that made recording human history a revolutionary, collaborative art”.

As the dust from the dot-com hype and bubble began to settle after the year 2000, Dale Dougherty and Tim O'Reilly examined those common factors that uniquely enabled some web-based businesses to sustain themselves and prosper, whilst many others had gone out of business (O'Reilly, 2005). O'Reilly (2005) noted that “far from having ‘crashed’, the Web was more important than ever (O'Reilly, 2005). The success of the Web 2.0 paradigm, Floridi notes, is its focus on people (participation) and metadata. He points to the importance of differentiating between semantic intelligence (truth and understanding) and metadata (identifiable information). For him, “humans are the only semantic engines available”. It is also indicated that the Web 2.0 paradigm is predominately a socio-technical phenomenon (Lagoze, 2010; Miller, 2005; O'Reilly, 2005), thus as O'Reilly (2005) and Miller (2005) contend, it has more to do with attitude and culture than technology. The participatory nature of Web 2.0 (O'Reilly, 2005) and its focus on metadata (Floridi, 2009) has implications for education in general (Alexander, 2006; Anderson, 2007; Nielsen, 2012) and libraries in particular (Alexander, 2006; Casey & Savastinuk, 2006; Evans, 2009; Kroski, 2008; Miller, 2005). O'Reilly identified the following concepts underpinning the Web 2.0 paradigm. These include active two-way collaboration, users as co-creators, the Wisdom of Crowds, variable participation and openness (O'Reilly, 2005). These concepts were further elaborated and discussed by several other authors (Alexander, 2006; Anderson, 2007; Casey & Savastinuk, 2006; Evans, 2009; Floridi, 2009; Kroski, 2008; Maness, 2006; Miller, 2005; O'Reilly, 2005; O'Reilly & Battelle, 2009).

The role of the user community is emphasised (Lankes, 2016). According to Lankes (2016) User engagement happens when libraries provide a platform co-owned by the community whereby creativity and co-creation can be harnessed. Weinberger (2012) is of the view that society as a whole is better off by embracing “information [knowledge] overload” generated by a network of experts without necessarily having the traditional pre-publication filters – incorporating thus varied, diverse, possibly unsettled but continually enhancing discourses (conversations). This however does not entail anything that goes to the network is true, credible and trustworthy. What is suggested here is a post-filtering mechanism where the contributions are exposed to an open scrutiny where ideas are transparently debated and argued. In the end truth and facts do matter – however they come after publication. “The authority of a work of knowledge is no longer a badge granted by its publication, but is continuously negotiated within systems of editing, reading, reviewing, discussing, and revising” (Weinberger, 2012). With the emergence of Web 2.0, not only is our notion of information overload questioned but as Weinberger (2012, p. 185) contends the traditional DIKW (data, information, knowledge and wisdom) hierarchy itself is no longer tenable. Hence, as Weinberger notes “the solution to the information overload problem is to create more information: metadata”. He adds that “there is no hiding from knowledge overload anymore”. Such metadata, Weinberger adds, helps users to ascertain the veracity of sources and credibility of the information. “A net richer in metadata is richer in more usable and useful knowledge”. Lankes (2016) concurs that knowledge and its neat categories are increasingly under

scrutiny. Put another way, Lankes contends that the McDonald's approach to libraries (and of knowledge) fails to accommodate the diversity inherent in nations, peoples and cultures – hence noting the role of libraries is not to filter out and reduce what they would consider “noise” but to serve as a platform for conversations and re-constructions of knowledge. Lankes admits librarians are not necessarily supposed to be neutral but principled enablers of conversations and knowledge creation.

Coveted knowledge locked up in people's minds and institutional siloes is of limited value – calling therefore for a networked, open collaborative systems where experts share their solutions to problems (Nielsen, 2012). However, this is easier said than done. As Nielsen notes there are still many scientists who “guard their data jealously”. If openness and sharing become a norm, society as a whole would benefit. Of course, openness does not come cost free. Wikipedia adopts a bottom-up, volunteer-based, open collective model where any page can be created about anything and it gets edited afterwards. Malicious entries are removed and tagged as contentious. The barriers-to-entry are minimal, and it has an easy interface. It is open to create new article as well as free to edit, share use and share. It has a linking mechanism to cross reference content (Anderson, 2006). There are of course all kinds of people in Wikipedia. Christakis and Fowler (2011, p. 279) categorised them as co-operators, free riders and punishers. Punishers act the community police of Wikipedia: they patrol and remove malicious content. Wikipedia is a result Web 2.0. Put another way, without the new collaborative model, we do not have Wikipedia. It challenged the status quo, i.e. the traditional model of Encyclopaedia. Encyclopedias such as Britannica served their purposes during the read-only printing age. Shirky (2008, p.21) argues that “by making it easier for groups to assemble and for individuals to contribute to group effort without requiring formal management, [Web 2.0] tools have radically altered the old limits on the size, sophistication, and scope of unsupervised effort.” Other successful projects include Galaxy Zoo, LibraryThing, National Library of Australia's OCR correction project, Ushahidi and Flickr. As Nielsen (2012) indicates Galaxy Zoo alone has more than 200,000 volunteers who take their passion to spot, classify and discuss about galaxies. If Web 2.0 has the above to offer for science, there a lot more that can do for digital heritage.

4. Users as co-creators

One of the central concepts of the Web 2.0 paradigm is the notion of involving users as co-creators of content and metadata (O'Reilly, 2005). This conceptual underpinning assumes two-way collaborations between users and institutions, such as libraries, and has a strategic advantage. This emerging strategy has helped to realise what Toffler (1980), in his book “The Third Wave”, envisioned as the shift from passive consumerism to proactive prosumerism. In it, Toffler saw that two-way collaborative approaches, fostered by the social and political will, reoriented the traditional ways of doing business and offering services. As Kroski (2008) notes the Web 2.0 paradigm considers users as major stakeholders whose proactive participation is sought as a central strategic competitive advantage. Kroski adds that Web 2.0 is inclusive by design where users can “participate, organise, read, write and play online”. It is thus important to explore the implication of the concept of “users as prosumers” in library metadata functions. One of the emerging concepts in light of users as prosumers is the notion of the “Wisdom of Crowds”, where emphasis is placed on the collective and aggregate value that can be harnessed from users rather than the individual contributions.

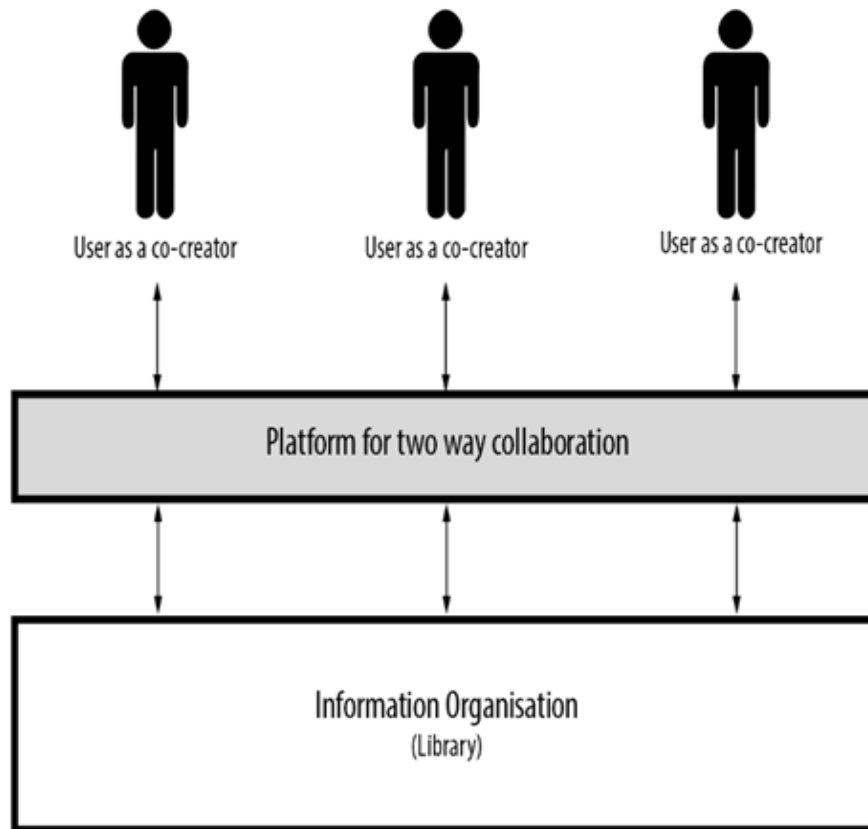


Figure 1: Users as metadata co-creators (Alemu & Stevens, 2015)

5. The Wisdom of Crowds

Web 2.0 enables collective action and collaborative production where no one single person takes the credit (Shirky, 2008). One of the concepts of the Web 2.0 paradigm is the “Wisdom of Crowds” (Surowiecki, 2004). Employing the Web as a platform, O’Reilly argues, the small contributions made by individual users can be collectively re-mixed, aggregated and harnessed, providing thus what he refers to as collective intelligence (O’Reilly, 2005). Principles that help in harnessing collective intelligence include long-tail distribution of products and services, including talent across the network; low marginal costs of cooperation and coordination; granular addressability of content; and the network effect, the architecture of participation and the concept of users as co-creators are central pillars of collective intelligence (Anderson, 2006; O’Reilly, 2005; Shirky, 2008).

The phrase “Wisdom of Crowds” was popularised by James Surowiecki, in his oft-cited book “The Wisdom of Crowds: Why the Many are Smarter than the Few” (2004). Surowiecki describes how small, but unique, contributions aggregate and become greater than the sum of the constituent parts. Surowiecki (2004, p.28) emphasises the importance of diversity, independence of thinking, and activity-as-a group for adding value in the form of collective wisdom. Surowiecki also notes the importance of decentralisation and specialisation (diversified knowledge base) in enhancing the value of contributions. Surowiecki (2004, p.30) argues that “cognitive diversity needs to be actively selected, and it’s important to do so because, in small groups, it is possible for a few, biased individuals to exert undue influence and skew the group’s collective decision”. The goal is not to strive for consensus, homogeneity, and group-thinking in knowledge creation or decision making, Surowiecki argues. Most importantly, the theory of the “Wisdom of Crowds” posits that the idea of gathering diverse groups of people does not preclude the inclusion of the smartest person. Surowiecki admits that, just as democracy has its own limitations, emanating from the fallibility of its enactments, so does the “Wisdom of Crowds”.

Harnessing the collective intelligence of groups of people involves well-thought out coordination, cooperation as well as aggregation costs.

The wisdom of crowds (Surowiecki, 2004) does not imply the wisdom of the naïve or the uninformed. The crowd includes those who are high level experts in their subject areas. As Gartner (2015) demonstrates the benefits of Web 2.0 proved its significance in projects such as galaxyZoo, Ancient Lives, Old Weather and LibraryThing. In the cultural digital heritage domain, a mixed metadata approach involving crowdsourcing (nichsourcing) is demonstrated by (de Boer et al., 2017). In the absence of a semantically structured, enriched and linked ontology, researchers in the digital humanities had to manually triangulate information from heterogeneous media collections to inform their research (de Boer et al., 2017). To address these challenge, de Boer et al. (2017) designed a Linked Data model with entities such as sem:Event, sem:Place, sem:Actor, sem:Time, oa:Annotation, dive:MediaObject and skos:Concept. As a result, a knowledge graph is created linking media objects, people and places with historical events which in turn lends itself for navigation and serendipitous browsing. Collections are enriched with structured metadata which also creates semantic links across the various digital objects. Firstly, a semantic data model is designed. Secondly, through a crowdsourcing service, users (scholars) are allowed to add metadata and continuously enrich collections. “Based on extensive requirements analysis done with historians and media scholars, we present a methodology to publish, represent, enrich, and link heritage collections so that they can be explored by domain expert users” (de Boer et al., 2017).

Crowdsourcing, a phrase coined by Jeff Howe in 2006, refers to a new model of outsourcing in which an open call is made using Web 2.0 platforms and a set of tasks that are accomplished over the network by crowds of people (Adams & Ramos, 2010). Crowdsourcing is defined as “a massive collaboration technique that enables individuals, working as a virtual group, to collectively accomplish a shared, large and significant goal” (Calhoun, 2014, p.xx). Crowdsourcing could be either creating new, novel, resources or augmenting existing ones with additional annotation or metadata (Carletti, 2016). The latter is appropriate for the metadata enrichment in the digital cultural heritage domain. Carletti (2016) identifies key tasks of crowdsourcing applications including curating (image tagging; crowd-curated exhibitions), revising (transcribing; optical character recognition), locating (mapping content; geo-located storytelling), documenting history (sharing personal memorabilia, family stories, and photographs), documenting personal life (audio-recording of intimate conversations; filming private moments) and augmenting locations (recording sounds of the environment; referencing locations mentioned in literary works).

In light of metadata enrichment, crowdsourcing has huge implications for digital heritage. It can be used to solicit metadata for specialised and niche collections that cannot be described by hired archivists or librarians. This may include adding metadata in different languages, labelling old maps with place names, tagging photographs, deciphering obscure letters, and so on. As Gartner (2016, p.109) argues “in a different way the growth of crowdsourcing as a way of creating metadata also says much of human ingenuity and its altruistic side. So many people spending so many hours helping to push forward knowledge by describing the world, looking for patterns in it and sharing what they find, even when the work required is mundane and repetitive, testifies to the inspiration that knowledge for knowledge’s sake remains even in a world of neoliberal market ideologies. That the creation of metadata is seen by those giving their time in this way as a valid way to move knowledge forward is a sign of its centrality to human thought. It remains as ever a potent gateway which, by putting shape to the amorphous cloud of knowledge, allows us to make sense of it. By generating metadata, we assert our desire to understand and grow our understanding”.

Kefalidou, Georgiadis, Coles and Anand (2012) indicate that “crowdsourcing cultural heritage could potentially open up new ways for experiencing, interpreting and analyse cultural environments and data through the incorporation of personal records and observations from visitors’ perspective while

navigating through ancient pathways”. The authors added that crowdsourcing helps to enrich user experience and enhances learning. Kärberg and Saarevet (2016) discuss the role of crowdsourcing for digital cultural heritage collections. Old historical photographs (of people, places and events) could be tagged by the community. “These contributions are aggregated using an algorithm that considers users' trustworthiness (based on the accuracy of their previous contributions) and then averages the geo-coordinates. A photo is considered geotagged when enough trustworthy users tag the photo into approximately the same location (Kärberg & Saarevet, 2016). In addition to automatic, algorithm-based crowdsourcing applications, a curated crowdsourcing is also suggested where metadata experts could approve or reject metadata entries. According to Kärberg and Saarevet (2016). The workflow covers:

- Select what to enrich
- Send the selection to crowdsourcing
- Provide crowdsourcing functionality
- Return results
- Perform quality assurance
- Store results

The Boston Public Library launched a crowdsourcing project asking the public to transcribe anti-slavery letters handwritten from 1830s to 1870s (Open Culture, 2018). The project call reads as: “We need your help to turn our collection of handwritten correspondence between anti-slavery activists in the 19th century into texts that can be more easily read and researched by students, teachers, historians, and big data applications.”



Figure 2: Boston Public Library crowdsourcing project (Source: <https://www.antislaverymanuscripts.org/classify>)

Similar projects listed by <https://www.zooniverse.org> include Zooniverse Talk, Parochial Archive Project in Rome, Snapshot APNR, Weather Rescue, Planetary Response Network and Rescue Global: Caribbean Storms 2017, Identify New Zealand animals, Cheetahs of Central Namibia, Scribes of the Cairo Geniza, Planet 9 and others. The British Library crowdsourcing project transcribing, enhancing and discussing about 100,000 historic playbills, the National Library of Australia OCR text correction project and Galaxy Zoo are notable initiatives which indicate the possibilities afforded by Web 2.0 collaborative projects.

An integral component of the concept of the “Wisdom of Crowds” is the phenomenon known as ‘variable participation’ (Shirky, 2008), whereby no contribution is considered too small, but is aggregated to form collective intelligence (O’Reilly, 2005). In other words, it is not the case that every contributor adds or edits a complete entry, but that the numerous ‘tiny’ contributions slowly accrue to bring about a network effect, the phenomenon wherein value appreciates as the number of users grows. Most proponents of the Web 2.0 paradigm are of the same opinion, stating that, unlike the conventional peer-review model, where individual contributions make a significant mark, the Web 2.0 paradigm functions in a bee-hive or ant-colony fashion, where the contribution of a single bee is insignificant without taking the sum of contributions of the whole hive (Shirky, 2008). Wright (2007, p.14) notes that “no individual bee possesses the intelligence to make such a decision, but as a group, the bees generate a collective ‘mind’ far cleverer than the sum of its tiny-brained parts.”

The GLAM sector has a wealth of content and expert-curated metadata across millennia. Web 2.0 has a built-in architecture for collaboration, openness and metadata sharing (O’Reilly, 2005). Various novel open licensing schemes are being developed and used across the web (e.g. The Creative Commons License). The benefits of embracing openness are discussed by various authors (Alexander, 2006; Anderson, 2006, 2010; Shirky, 2005; Tapscott & Williams, 2010; Udell, 2004; Weinberger, 2005, 2007). Metadata openness facilitates knowledge sharing, re-usability, mash-ability and re-mixability of information and metadata created by various cultural heritage institutions.

6. Web 2.0 technologies and applications

Whilst Web 2.0 technologies (such as tagging, social bookmarking, reviews, blogging, crowd sourcing and recommendations) are instantiations of the concepts that underpin the Web 2.0 paradigm, applications are the specific tools developed to implement the instantiations (Anderson, 2009a, 2009b; Anderson, 2007; Floridi, 2009; Maness, 2006; Miller, 2005; Morville, 2005; O’Reilly, 2005; Weinberger, 2005, 2007). Web 2.0 has various layers and applications (see Figure 3).

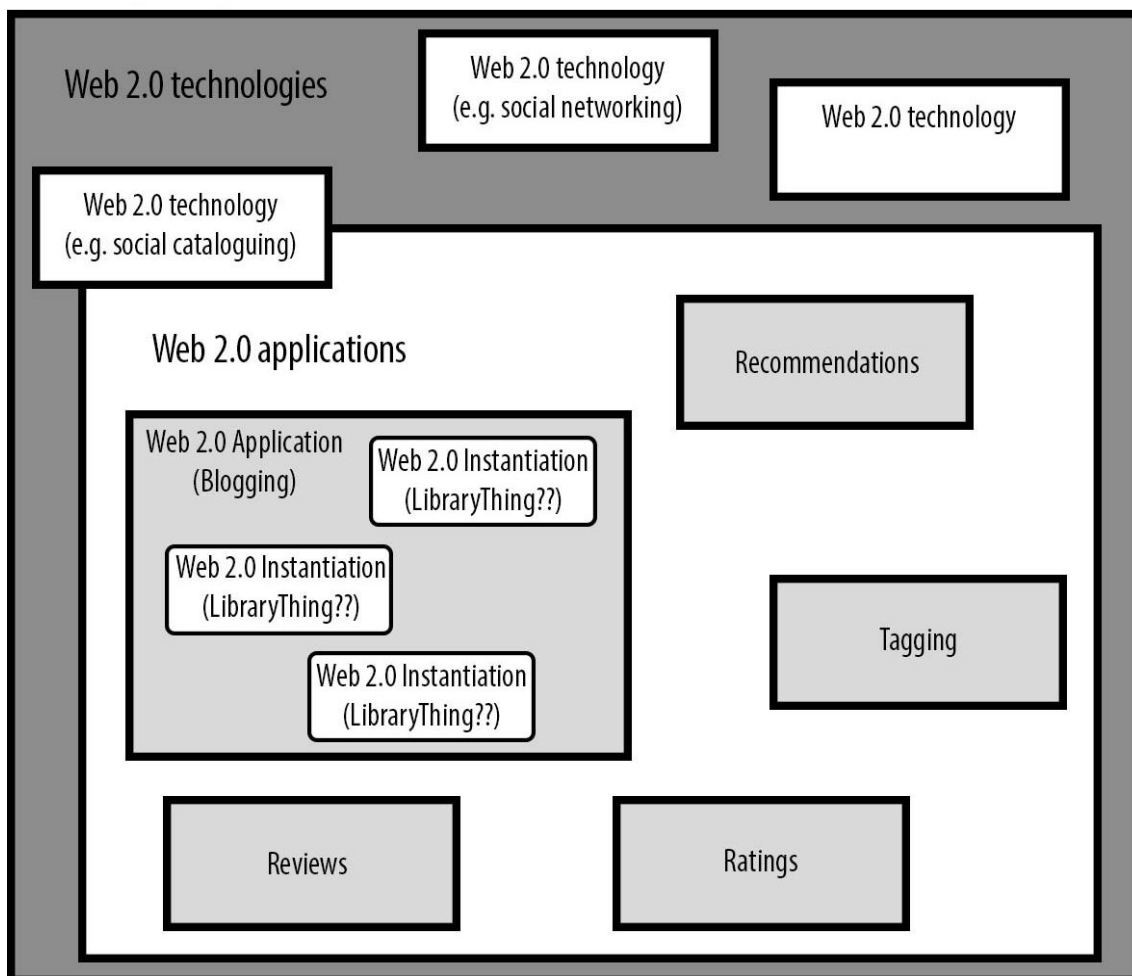


Figure 3: Web 2.0 paradigm (Alemu & Stevens, 2015)

Tagging, user reviews, ratings, crowdsourcing and recommendations are some of the applications that can be integrated in digital cultural heritage systems.

7. Motivation metadata co-creators

As Howe (2009) notes crowdsourcing projects are not motivated by money. Hence, other non-financial incentivise need to be sought. Howe notes that whilst technology becomes “the great equalizer”, the notion of collaborative projects goes back to the “invisible college”, to journal publishing (17th century), open source software (1990s) and plethora of examples where people were able to co-create, collaborate and share ideas and knowledge for the common good. In most of these cases, the financial incentivise are either little or non-existent. Howe notes Linus Torvald’s style of open source software development, i.e. release early, and often, with open licences. “Coordinated in a Bazaar style where everything is coordinated below rather than from above as in a Cathedral”. According to Howe, for crowdsourcing to succeed, it needs to re-orient its incentive mechanisms, in particular there is a need to democratising the means of production (faster, cheaper, smarter and easier tools), democratising education, democratising distribution, making friends, influencing people and turning readers into writers.

Nielsen (2012) argues that open and collaborative approaches such as blogging, if implemented and managed properly for example with appropriate motivation and incentives to scientists, would amplify the “collective intelligence” of humans. Getting users motivated enough to create metadata for digital cultural heritage institutions is a big challenge. If there is not enough motivation, there are not good number of users to start a network effect that is seen in Wikipedia and LibraryThing. LibraryThing announces it has more than 2,000,000,000 users and over 100,000,000 books catalogued – with a rich

set of user tags, ratings, comments and recommendations. For the digital cultural heritage sector to attract the attention of users in metadata co-creation, it needs to re-think user motivation factors. Some of these include:

- Reducing barriers to contribution
- Simplicity, interestingness and fun of interface
- Personalisation and re-findability
- Altruism and reputation,
- Sense of ownership
- Engaging with users

As Howe (2009) notes, “ask not what the crowd can do for you, but what you can do for the crowd” this is because people are drawn to participate because some psychological, social, emotional need is being met. And when the need isn’t met, they don’t participate” (p.287-288).

8. Limitations of the Web 2.0 paradigm

Web 2.0 is an open, democratic and bottom-up system for content creation, social networking and conversation. And democracy has its own limitations. The failure to acknowledge such limitations leads to the failure of the system as a whole. The scientific method to ensuring credibility, authenticity, authority, reliability and facet-checking is still essential. Citizen science projects, crowdsourcing, social networking, blogging, tagging and rating do not in any way indicate the obsolescence of peer-review and other quality checks. Shirky (2008) and Nielsen (2012) acknowledge that when a network of collaboration is open for everyone, it is likely that good as well as bad content will be added to the system. The concerns include control, quality, trust and integrity (Liew, 2015). Whilst the notions of user empowerment are embraced, some still consider Web 2.0 as a threat to established practices and control (Liew, 2015). The Web 2.0 paradigm is, in general, criticised for lack of editorial control, which used to be a main characteristic of the conventional, peer-reviewed, edited, and filtered-before-publication model. Andrew Keen argues that Web 2.0 compromises established editorial and quality assurance mechanisms hence making it difficult to ascertain authenticity, authority and veracity (Keen, 2007, p.27). He notes the absence of fact-checking and editorial controls in Wikipedia and similar Web 2.0 systems. In addition, content that is added through socially-constructed metadata approaches (such as tagging, rating and crowdsourcing) might be spammed. The GLAM sector prides itself for being trustworthy that resulted from its selection, curation and expert-vetted procedures of content and metadata inclusion. In addition, metadata that is added through Web 2.0 lacks semantic and syntactic structure.

The Wikipedia model is neither devoid of problems nor without its share of critics. It is often asserted that encyclopaedias created by the conventional model are more credible, as entries are pre-filtered, passed through rigorous edits and are peer-reviewed (Keen, 2007). Keen strongly criticises the democratisation of media and the amateurism of expertise. According to Keen, it is gambling at humanity’s real peril to stand aside and watch Web 2.0 demolishing what has for so long been a credible and authoritative media, one that informs and educates. However, it can be counter-argued that even the supposedly peer-reviewed and pre-filtered sources, such as Encyclopaedia Britannica, are not always accurate.

It is therefore important that provenance metadata is created and utilised to ascertain credibility and authority of information. As in any democratic system, there need to be check and balance among the various actors. Web 2.0 approaches need to have embedded quality and fact checking mechanisms where malicious and idiosyncratic metadata is identified. However, such quality assurance mechanisms need as well be communicated and approved by the community. Tamaro (2016) also supports the need to ensuring quality of in GLAM crowdsourcing projects. The author indicates the need to triangulate various principles and techniques from across domains including psychology (behavioural

norms), quality metrics, technical systems and human interventions where erroneous information is detected and corrected or removed so as to enhance veracity of such projects – lest such projects may lack credibility and fail to accrue desired benefits.

9. Metadata enrichment using the platform of metadata co-creation

In contrast to standards-based and expert created metadata, user generated metadata espouses the co-creation of metadata by library users. There are various ways of applying these approaches to the cultural heritage domain, some of these include developing, integrating or embedding applications that support user involvement. These systems should include algorithms to analyse user-generated metadata so as to produce tag clouds, frequently used terminologies and ranking mechanisms – providing thus collective metadata intelligence. These applications should be designed as platforms that allow for new applications to be linked and developed on top of it. Some of the instantiations of these platform could be tagging systems, crowdsourcing, rating and recommendations. Figure 4 shows the various interacting components for metadata enrichment and filtering.

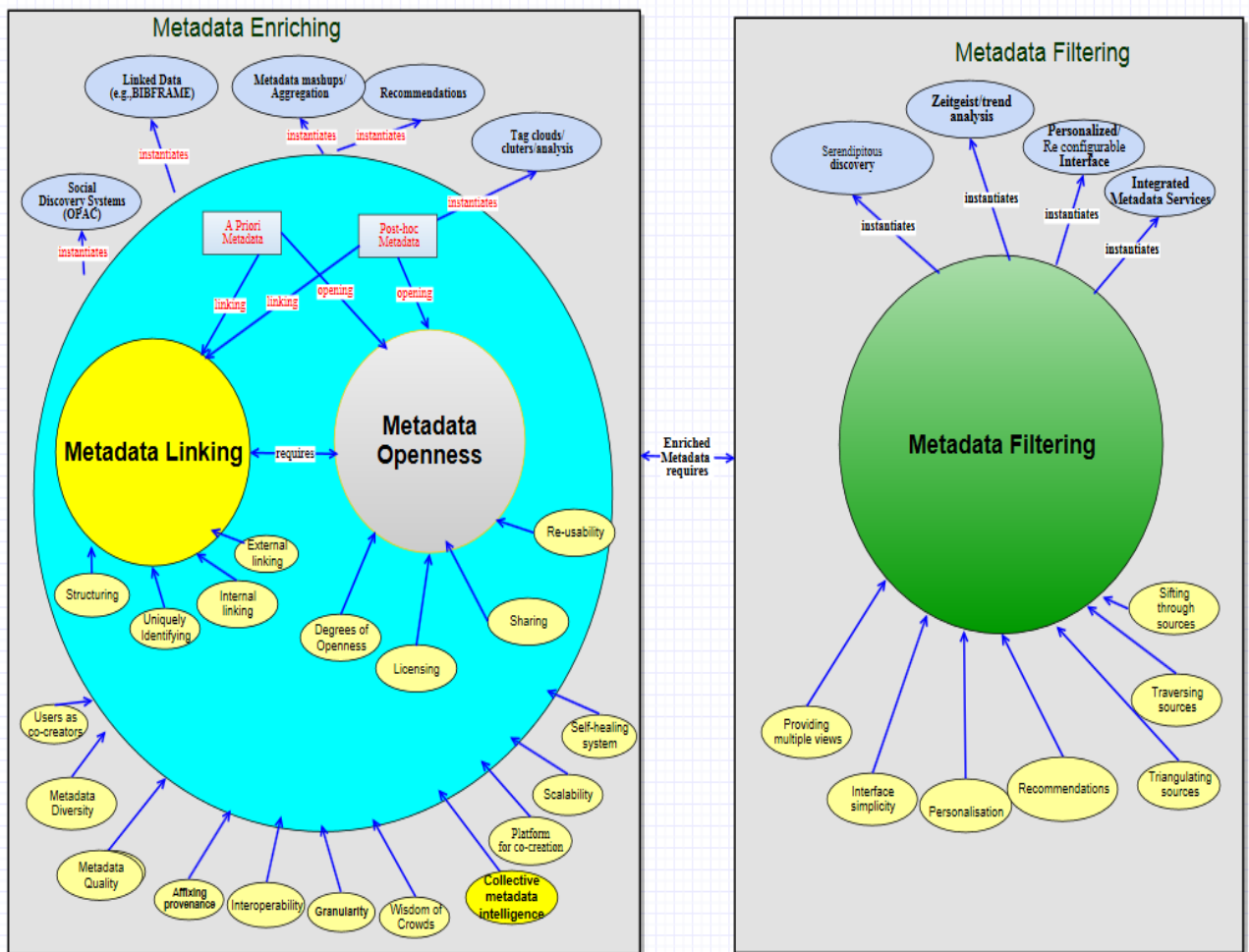


Figure 4: Platform of metadata co-creation and components (Alemu, 2014)

With the development of Web 2.0, there is a paradigm shift in the way content and metadata are created in GLAMs. If this shift is considered, metadata creation is no more the sole responsibility of the

metadata experts. Through the platform of metadata co-creation, digital cultural heritage collections can be enriched through crowdsourcing, tagging, rating and recommendations applications.

The Web 2.0 paradigm can therefore be a collaborative environment for users to co-create metadata. This can be implemented through user tagging, ratings, review and recommendations. Over the years there have been attempts to implement socially-constructed metadata approaches in libraries. Getting enough motivation to get users to add such metadata is considered the primary challenge. However, the potential benefit of these approaches is demonstrated by some online applications such as LibraryThing.com.

In addition to the librarian (expert) supplied basic descriptive metadata, users can be able to add metadata of their own including tags, ratings and reviews. Once such metadata is aggregated, it can be used to re-find books previously viewed and to organise own (personalised) reading lists. In addition, through further data analysis (such as tag clouds), new insights into the popularity of the information object and new ways of navigation (filtering) can emerge. Not only is the book "Too big to know" related to "Everything is miscellaneous" by the same author but it is also related to "Glut" by Alex Wright. Clicking on "Everything is miscellaneous" retrieves related titles among which is Andrew Keen's "The cult of the amateur", which happens to have contending views with that of David Weinberger's.

Very often the challenges to post-hoc (user-generated) metadata is to obtain enough traction to get users motivated to co-create metadata. Some of the incentives to consider include reducing barriers to contribution, simplicity, interestingness and fun, re-findability, sense of ownership, own vocabularies, altruism and reputation and engaging with users.

Through Web 2.0 platforms, socially-constructed metadata approaches encourage users to add metadata post-hoc. In addition to the basic (a priori) metadata, users can tag, comment, review, rate or share the information object – this can happen post-hoc, hence the introduction of the concept of post-hoc metadata in this research. However, post-hoc metadata creation had its own limitations. Some of these limitations include lack of semantic structure (absence of thesauri). In addition, idiosyncratic and malicious metadata may be created. This is why a mixed metadata approach is introduced.

10. Mixed metadata approach

It is indicated that the involvement of users does not substitute the roles played by metadata experts. Thus, the emergence of a mixed metadata approach where user created metadata augments expert-created metadata. The metadata creation process no longer remains to be the sole prerogative of the metadata expert. The Web 2.0 collaborative environment would now allow users to participate in both adding and re-using metadata. If implemented, this would benefit both the institution and the user.

Both standards and socially-constructed metadata have limitations. In contrast, the strengths of one complements the weakness of the other. Standards do not scale to describe the growing size of digital collections, but socially-constructed approaches render the platform of user metadata creation hence scalability issues are addressed. Standards also fail to incorporate user terminologies whereas the latter cater for metadata diversity. Conversely, socially-constructed metadata approaches require the semantic and syntactic structures afforded by standards. The two approaches are therefore complementary - hence the case for a mixed metadata approach. A mixed metadata approach ensures the inclusion of the users' terminologies and better incorporates future changes in the language of users. It caters to the socio-cultural and linguistic diversity inherent in users. A mixed metadata approach provides a required balance between structured schemas with that of flat and unstructured but diverse and usable metadata, hence ontologies, taxonomies, and folksonomies can be used together (Morville 2005). This mixed metadata approach provides metadata enrichment.

11. Continuous enrichment of cultural heritage collections

The principle of metadata enriching implies a continuous process of adding, augmenting and improving expert-created (a priori) metadata with user-created (post-hoc) metadata. The latter provides the diversity and breadth of interpretations of information objects, whilst a priori metadata serves as a focus for presenting basic structured, standards-based, metadata to users. Once users get access to information objects using a priori metadata, they can start enhancing it using their own terminologies, interpretations or descriptions. Enriching is thus characterised by a constant flux. Enriching can be contrasted with the current metadata principle of sufficiency and necessity, which focuses on creating simple metadata in a 'complete state'.

Whilst it is technically possible that libraries can implement the principle of enriching within their own institution, it can only be effectively and efficiently implemented if another principle, the principle of metadata linking, is embraced and implemented. The principle of linking enables libraries to continually enrich their existing metadata with metadata that resides outside their boundaries. Metadata that is linked, both with internal and external data sources, results in interlinked metadata, thus offering users the ability to seamlessly navigate between disparate information objects. The principle of metadata linking ensures that metadata values are granularly structured, uniquely and persistently identified and interlinked, thus, bringing together disparate metadata sources. However, in order to realise metadata linking, the metadata that is to be linked-to must be made openly accessible. The principle of metadata openness makes explicit that institutions communally benefit from making their metadata available in an open, re-usable and re-combinable format. Whilst there can be various degrees of openness, the highest degree of openness where metadata is open, machine process-able and interlinked provides the greatest benefit, in terms of opportunities for re-use.

Users can continue to enrich collections with metadata and contributions could be made through crowdsourcing, tagging, reviews and ratings. These contributions are mapped to an existing structure such as thesauri or ontology. This encashment happens in a continuous manner rather than a deterministic, one step, process. It does not aim to anticipate future metadata needs based on a priori metadata. Instead by using a priori metadata as a structure to get access to post-hoc metadata, it learns from the collective metadata intelligence. A non-deterministic view of enriching and filtering prefers to record user preferences post-hoc rather than attempting to anticipate users' needs. In addition, the principle of enriching takes the view that terminologies change through time. It takes an interpretivist rather than an objectivist ontological point of view.

Librarians have grappled with the question of objectively ascertaining users' requirements through surveys, interviews and feedback forms. However, the answers thereby obtained can only make library metadata user-centred but not user-driven, the latter being a more desirable outcome. Only direct involvement, in the creation of metadata on the part of users, can assure that metadata is indeed truly user-driven. User-driven metadata enriching can empower users to be proactive creators, collaborators and partners. In the new metadata paradigm, users co-own the metadata. They can participate not only in the co-creation process, but also in its management and curation. Since the main objectives for affixing metadata to information objects is to enable its discoverability, metadata should be judged by its relevance in meeting this objective, rather than the way it depicts an information object in an objective and ontological manner. Terminologies change over time, hence, the metadata should be able to be cognisant of and reflect such changes.

Enriching is a continuous process of adding, enhancing and improving metadata content. Benefiting from the wisdom of the crowds effect (Surowiecki, 2004), a network effect (Shirky, 2008), the Long Tail (Anderson, 2006) where each contribution slowly aggregates, enriching aims to collect as much metadata as possible, and thus facilitating metadata diversity.

12. Metadata diversity better conforming to users' needs

Metadata diversity implies the inclusion of a multitude of potentially conflicting metadata ascribed to information objects by users. Metadata diversity increases the likelihood of conforming to the multitude of perspectives and interpretations of various groups of potential users. In relation to idiosyncratic (personal) metadata entries, it is important to recognise that a given metadata entry that might be considered trivial for a general user might be important for the one who created it, since it is likely that the latter will search with those keywords. Such idiosyncrasy can be managed by providing personalised presentation, which can be managed through appropriate metadata filtering. The new metadata paradigm may thus permit the inclusion of metadata descriptions (interpretations) of information objects that may seem in opposition. In such instances, it is important that the diversity of the various interpretations is maintained. Nevertheless, such a metadata paradigm should not include random entries; instead, it should cater towards semantic and meaningful metadata whilst at the same time maintaining the diversity of interpretations.

13. Evaluation of digital cultural heritage metadata

Web 2.0 proponents, such as Shirky (Shirky, 2005, 2008, 2010), and Weinberger (2007), argue that Web 2.0 content is neither anarchic nor “anything goes”. However, unlike the contemporary (print or Web 1.0) model where the editing, vetting and the quality control process is performed before content is published; the Web 2.0 paradigm implements the process of filtering “after-the-fact (post-hoc)”. According to Shirky (2005), “There's an analogy here with every journalist who has ever looked at the Web and exclaimed ‘Well, it needs an editor.’ The Web has an editor, it's everybody. In a world where publishing is expensive, the act of publishing is also a statement of quality; the filter comes before the publication. In a world where publishing is cheap, putting something out there says nothing about its quality. It's what happens after it gets published that matters. If people don't point to it, other people won't read it. But the idea that the filtering is after the publishing is incredibly foreign to journalists.” Shirky (2005) furthermore contrasts the new model with library cataloguing, where classification schemes and categories are pre-determined.

User-generated metadata without structure and contextual linking would remain unusable. Hence, it needs to be structured, uniquely identified, indexed and linked to information objects. Existing Linked Data (RDF) technologies could be used to enable such structuring and linking (Alemu et al., 2014).

Quality is an important but contentious issue. Unless it is malicious metadata, metadata for cultural heritage collections can be as diverse and rich incorporating multiple user interpretations. So, the quality metrics used for user-created metadata should be as inclusive as possible. Quality should be measured not for the sake of ontological or objective truth but for its usefulness. In a mixed metadata approach the issue of metadata quality should be weighed considering the relevance of metadata to support findability and discoverability. Further, it is metadata's usefulness for finding and discovering information objects, rather than on whether it is objectively accurate (truthful) or not, or on whether it is ‘good enough’, rather than on whether it is ‘perfect metadata’, that should be considered. Should typos be corrected? What about idiosyncratic/personal metadata? Should users themselves be involved in spotting, correcting malicious and false entries?

As Gartner (2015, p.36) admits “the metadata universe, like the physical, can be a messy place”. The answer seems to be working out some way of filtering it, honing it down so that it becomes usable and lets us find what we want. This is the model proposed in a recent book by Getaneh Alemu and Brett Stevens; they use the epithet ‘enrich then filter’ for a model that should let us have the best of both worlds, diverse, vibrant metadata and focussed, relevant search results. Their idea is that we encourage metadata from all sources: the ‘expert’-created records that have been the norm for so long and community-generated ones that come from anyone willing to provide it. We then filter this rich but confusing and always changing body of metadata as and when we need it to meet our requirements”. (p.104).

14. Practical implications and conclusion

Carletti (2016) indicates that European nations expend a huge sum of money towards digitisation of cultural heritage objects, albeit 90% non-digitised material yet. Carletti contends if the digitised 10% (about 300 million) is accessible and well used by users. Out of the 300 million, as Carletti reports only 34% is accessible online. In this paper, I argued that enriched metadata plays a crucial role to drive content to where users are – improvising the discoverability of valuable digital heritage objects, hence justifying the return on investment for digitisation, aggregation, system development and salaries for the people involved in the development and management of these systems. To this end, participatory and collaborative Web 2.0 approaches would help to enrich digital heritage collections. Put another way, rather than considering the community of users as passive consumers of the services, they can be considered as proactive collaborators, actively engaged. As a result, digital heritage objects could be enriched by their users who consult them: viewers, readers or listeners. As Paul Otlet in the 1930s envisioned, annotations, about-ness information and the social space of documents could be used to generate new information – which he calls the book ‘*about the book*’ (Wright, 2007). This goes in tandem with Vannevar Bush’s notion of associative trails (Bush, 1945); and with a Linked Data turn, Tim Berners-Lee’s Semantic Web (Berners-Lee, 1997; Berners-Lee, 1998; Berners-Lee, et al., 2001). Overall, as Tim O’Reilly notes Web 2.0 offers a collective intelligence (O’Reilly, 2005) from the long-tail of variable contributions (Anderson, 2006) or wisdom of the crowds (Surowiecki, 2004).

For now, the digital heritage is characterised by a top-down, expert-created/curated metadata using tightly rigid standards. Granted that current Web 2.0 applications have their share of limitations (unstructured and unfiltered) and perhaps not always fit for purpose to persuade librarians/archivists to involve users as metadata co-creators through user tags, comments, reviews, links, ratings (likes and dislikes), recommendations) and crowdsourcing, however, when the right set of incentives are designed in the system, the Web 2.0 paradigm has significant benefits (Carletti, 2016). As Campbell and Fast (2006) warned “We ignore ‘mob indexing’ at our peril; by refusing to change our methods or our approaches in the face of collaborative systems, we run a serious risk of allowing information architects to become anachronisms in emerging information environments”. As Alex Wright in his book, *Cataloguing the World* (2014) writes, Otlet aimed at not merely collecting the world’s knowledge but interconnect it through a universal bibliography. His vision was to build a world wide web of knowledge plus conversations around these documents. It is these conversations, these annotations, these social metadata that are missing in contemporary digital heritage.

It is important thus, cultural heritage institutions embrace the theory of metadata enriching and filtering (Alemu & Stevens, 2015). For this to happen, as Weinberger (2012) indicates libraries and archives need to conceive “the library as a platform. In that sense, by becoming a platform the library can better fulfill the abiding mission it set itself: to be a civic institution essential to democracy”. As Shirky (2008) aptly notes “revolution does not happen when society adopts new technology, it happens when society adopts new behaviours”. Granted that Web 2.0 approaches to content creation and collaboration have limitations. But granted also traditional GLAM classification and metadata approaches are equally riddled with limitations. The volunteers of Wikipedia recognise its limitations and that is why they are passionate in their efforts to continually improve it. Wikipedia is never a finished product. As Weinberger (2007), Shirky (2008), Howe (2009) argue, the strength of Web 2.0 and crowdsourcing approaches lies in the principle of collective intelligence where the contributors continually enhance, edit, filter and improve it. “Increasingly skilled and capable of organised, sophisticated collaboration, amateurs are competing successfully with professionals in fields ranging from computer programming to journalism to the sciences. The energy and devotion of the amateur comprises the fuel for the crowdsourcing engine” (Howe, 2009, p.1). Another important question to ask may be: who is the amateur and who is the expert? Increasingly, domain experts contribute to Web 2.0 applications. This

is proved to be the case for Galaxy Zoo. As Howe (2009, p. 28) argues that “crowdsourcing efforts generally attract people both with and without professional credentials”.

Collaborative and participatory approaches are made possible through the adoption of the Web 2.0 paradigm. These approaches are not in opposition to contemporary, top-down, quality controlled and structured systems. As argued in this paper, a mixed metadata approach where the two can complement each other is suggested. The Web 2.0 paradigm is often associated with its instantiations such as tagging, rating and crowdsourcing applications but that is a minimalist view for it represents a broader social-constructivist and interpretivist philosophical world view, having thus overarching implications on collections, metadata functions and services. There is a shift from collections to platforms, from hierarchies to networks, content consumers to content co-creators (prosumers), read-only to read-write and top-down/vertical to bottom-up/horizontal (Anderson, 2006; Calhoun, 2014; Carletti, 2016; Gartner, 2016; Shirky, 2008; Tammaro, 2016; Weinberger, 2007; Weinberger, 2014).

Two-way communication, users as contributors, the Wisdom of Crowds (collective intelligence), variable participation and openness are some of the central components of the Web 2.0 paradigm (O'Reilly, 2005). It is often referred to as the read-write web or the social web. It is not just user-centred but also user driven. Web 2.0 brought about a socio-cultural shift in the way organisations operate and communicate with their customers (Lagoze, 2010; Miller, 2005; O'Reilly, 2005). Their potential impact for libraries is noted (Alexander, 2006; Casey & Savastinuk, 2006; Evans, 2009; Kroski, 2008; Miller, 2005). As Kroski (2008) indicates, Web 2.0 allows users to read, write and contribute content.

For digital heritage to thrive serving its communities, it should develop, as Lankes (2016) suggests, a platform whereby its users not only allow access to a huge trove of information resources but also facilitates conversations and knowledge co-creation. In Lankes' words, “a participatory platform allows a community to share its passions, expertise, and resources” (Lankes, 2016, p.115). The Web 2.0 paradigm opens such possibilities and affords endless possibilities for digital heritage. Whilst the infrastructural and technical challenges are equally important, Web 2.0 is first and foremost a socio-cultural and policy challenge, suggesting thus the need for the cultural heritage to redefine its core standards and policies. On the technical front, among other things, it requires a re-configuration of digital cultural heritage systems and interfaces to allow user co-creation. As Lankes (2016) indicates an emerging shift from collections to scholarship, from dissemination to action, from passive user engagement to proactive co-creation and co-ownership. As Lankes put it “the mission of librarians is to improve society through facilitating knowledge creation in their communities” – for which metadata co-creation is key.

As Paul Otlet argued “the catalogue guides the reader as far as the location of the book but not to the contents within and also relationships between documents” – where the social space of the book (user metadata, annotations and reviews) is missing. As the breadth and size of information resources grows by the day – “with an estimated 4 billion pages on the Web, which would take 57,000 years to read” (Boulton, 2014), it is the metadata, as David Weinberger said, that “liberates knowledge” from its confines and silos. Metadata describes, annotates, provides context and answers the "what, by whom, why, for whom and when" questions, thus helping users to find, identify, make sense, contextualise, filter through and use information resources such as print and e-books, journal articles and data. Whilst there are plethora of standards in use in libraries and cultural heritage institutions, they are not however without challenges. Some of these challenges have arisen due to the size and diversity of collections, ever changing technologies and changing users' expectations.

If the cultural and institutional readiness is in place, Web 2.0 collaborative technological platform affords users the ability to add metadata – and such metadata could be structured and linked in usable way. It is argued that metadata diversity through user tagging, user ratings, user reviews and user recommendations would enhance findability and discoverability, hence improve the usage of these information resources. This approach would help cultural heritage institutions address the needs of their

users through enhanced discovery services. Metadata enrichment can be supported through a mixed metadata paradigm where by user become proactive co-creators whilst librarians/archivists could provide the required structuring, linking, managing and making sense of the metadata created (generated) by users. Such metadata enrichment for digital cultural heritage collections would add value for users and enhance their experience. It also has practical implications for participatory digital heritage and citizen scholarship (Carletti, 2016). As Howard (2015, p.iii) argues “cultural heritage institutions have always had a role in allowing us to experience, explore and interpret our world by enabling people to engage with information in multiple forms through their mutual core functions of acquiring, organising, storing, providing access to and preserving information”. This mandate is partly realised through citizen participation and engagement by allowing users to co-create and manage metadata. The following are some of the concepts that are important to consider for the digital cultural heritage metadata.

- *Users as proactive metadata co-creators*
- *Metadata diversity*
- *Metadata scalability*
- *Variable metadata participation*
- *Metadata aggregation*
- *Network effect and wisdom of crowds*
- *Self-healing system*
- *Affixing provenance metadata*
- *Collective metadata intelligence*

As explored in this paper, hierarchical, structured and standards-based metadata on the one hand and bottom-up, collaborative and Web 2.0 approaches of user metadata co-creation on the other are not in opposition. As Wright (2007, p.235) argues “while the tension between hierarchies and networks may in the long run”. This is to say the networks organise themselves giving shape and meaning through structured metadata. Given that metadata itself is a social construct and since various users have diverse interpretations of cultural information objects, the digital heritage domain would benefit from metadata enrichment. This paper espouses a mixed metadata approach where both standards-based, top-down, librarian/archivist created metadata and socially-constructed, bottom-up, user-created metadata are combined and re-mixed to maximise the findability (discoverability) and usage of information resources. In this approach, users are pro-active contributors.

Developments in standards-based metadata approaches, Linked Data and discovery can help to uniquely identify, structure, aggregate, analyse and present socially-constructed metadata. This facilitates serendipitous discovery of information resources, metadata openness and sharing, identification of resource usage patterns, Zeitgeist and emergent metadata, facet-based navigation and metadata enriched with links (Alemu et al., 2014). Hence, authors (content creators), archivists, museum curators, librarians and most importantly users can co-create, contribute, enhance, enrich and manage metadata. It is argued, metadata enriching improves the search-ability, browse-ability, find-ability, preserve-ability, usability and management of digital heritage.

References

- Alemneh, D. G. (2009). *An Examination of the Adoption of Preservation Metadata in Cultural Heritage Institutions: An Exploratory Study Using Diffusion of Innovations Theory* (PhD), University of North Texas. Retrieved from: <http://digital.library.unt.edu/ark:/67531/metadc9937/?q=gelaw>
- Alemu, G., & Stevens, B. (2015). *An emergent theory of digital library metadata: Enrich then filter*. Waltham, Massachusetts: Chandos Publishing.
- Alemu, G., Stevens, B., & Ross, P. (2012). Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: A social constructivist approach. *New Library World*, 113(1), 38-54. doi:10.1108/03074801211199031
- Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). The social space of metadata: Perspectives of LIS academics and postgraduates on standards-based and socially constructed metadata approaches. *Journal of Library Metadata*, 12(4), 311-344. doi:10.1080/19386389.2012.735523
- Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2014). Toward an emerging principle of linking socially-constructed metadata. *Journal of Library Metadata*, 14(2), 103-129. doi:10.1080/19386389.2014.914775
- Alemu, G., & Stevens, B. (2015). *An emergent theory of digital library metadata: Enrich then filter*. Waltham, Massachusetts: Chandos Publishing.
- Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2014). Toward an emerging principle of linking socially-constructed metadata. *Journal of Library Metadata*, 14(2), 103-129. doi:10.1080/19386389.2014.914775
- Anderson, C. (2006). *The long tail: How endless choice is creating unlimited demand*. London: Random House Business Books.
- Boulton, J. (2014). In J. Boulton, *100 ideas that changed the web*. London, UK: Laurence King. Retrieved from <http://search.credoreference.com/content/entry/lkingideas/metadata/0>
- Bush, V. (1945). As we may think. *The Atlantic Monthly* (July 1945 issue). Retrieved from: <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>
- Calhoun, K. (2014). *Exploring digital libraries: Foundations, practice, prospects*. London: Facet Publishing.
- Cameron, F., & Kenderdine, S. (2007). *Theorizing digital cultural heritage: A critical discourse*. Cambridge, Mass. ; London: Mit.
- Carletti, L. (2016). Participatory heritage: Scaffolding citizen scholarship. *International Information & Library Review*, 48(3), 196-203. doi:10.1080/10572317.2016.1205367
- Casey, M. E., & Savastinuk, L. C. (2006). Web 2.0: Service for the next-generation library. *Library Journal*,
- Chan, L. M., & Zeng, M. L. (2006). Metadata interoperability and standardization – A study of methodology part I :Achieving interoperability at the schema level. *D-Lib Magazine*, 12(6).
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: SAGE Publications.
- de Boer, V., Melgar, L., Inel, O., Ortiz, C. M., Aroyo, L., & Oomen, J. (2017). Enriching media collections for event-based exploration. In E. Garoufallou, S. Virkus, R. Siatra & D. Koutsomiha (Eds.), *Metadata and semantic research: 11th international conference, MTSR 2017, tallinn, estonia, november 28 – december 1, 2017, proceedings* (pp. 189-201). Cham: Springer International Publishing. doi:10.1007/978-3-319-70863-8_18
- EU (2017). Decision (EU) 2017/864 of The European Parliament and of The Council of 17 May 2017 on a European Year of Cultural Heritage (2018). *Official Journal of the European Union*, L 131/1. Available from <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017D0864&from=EN>
- Floridi, L. (2000). On defining library and information science as applied philosophy of information. *Social Epistemology*, 16(1), 37-49.
- Gartner, R. (2008). *Metadata for digital libraries: State of the art and future directions*. (). Bristol: JISC Technology & Standards Watch.

- Gartner, R. (2016). *Metadata: Shaping knowledge from antiquity to the semantic web*. Cham, Switzerland: Springer.
- Gruber, T. (2007). *Ontology of folsonomy: A mash-up of apples and oranges*. *International Journal on Semantic Web & Information Systems*, 3(2)
- Hedstrom, M., Ross, S., Ashley, K., Christensen-Dalsgaard, B., Duff, W., Gladney, H., . . . Neuhold, E. (2003). *Invest to save: Report and recommendations of the NSF-DELOS working group on digital archiving and preservation*.
- Howard, K. (2015) *Educating cultural heritage information professionals for Australia's galleries, libraries, archives and museums: A grounded Delphi study*. PhD thesis, Queensland University of Technology. Retrieved from <http://apo.org.au/system/files/57651/apo-nid57651-60986.pdf>
- Howe, J. (2009). *Crowdsourcing: Why the power of the crowd is driving the future of business*. New York: Three Rivers Press.
- Kefalidou, Genovefa, Mercourios Georgiadis, Bryn Alexander Coles and Suchith Anand. 'Crowdsourcing Our Cultural Heritage'. In: Clare Mills, Michael Pidd and Esther Ward. *Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities*. Sheffield: HRI Online Publications, 2014. Available online at: <http://www.dhi.ac.uk/openbook/chapter/dhc2012-kefalidou>
- Kalay, Y. E., Kvan, T., & Affleck, J. (2008). *New heritage: New media and cultural heritage*. London: Routledge. Retrieved from <http://lib.mylibrary.com?id=106295&entityid=https://idp1.solent.ac.uk/idp/shibboleth>; <http://portal.solent.ac.uk/library/help/eresources/ebooks-help.aspx>
- Kärberg, T. and Saarevet. K. (2016). *Transforming User Knowledge into Archival Knowledge* D-Lib Magazine, Vol. 22, No. 3/4. Retrieved from: <http://www.dlib.org/dlib/march16/karberg/03karberg.html>
- Lagoze, C. *Accommodating simplicity and complexity in metadata: Lessons from the dublin core experience*.
- Lagoze, C. (2010). *Lost identity: The assimilation of digital libraries into the web* Available from *Lost Identity: the Assimilation of Digital Libraries into the Web*.
- Lankes, R. D. (2016). *The new librarianship field guide*. Cambridge, Massachusetts: The MIT Press.
- Lim, S., & Liew, C. L. (2010). (2010). *GLAM metadata interoperability*. Paper presented at the *The Role of Digital Libraries in a Time of Global Change*, 140-143.
- Lim, S., & Liew, C. L. (2011). *Metadata quality and interoperability of GLAM digital images*. *Ap*, 63(5), 484-498. doi:10.1108/00012531111164978
- Lourdi, I., Papatheodorou, C., Doerr, M.: *Semantic integration of collection description*. *D-Lib Magazine*. 15 (2009) retrieved from <http://www.dlib.org/dlib/july09/papatheodorou/07papatheodorou.html>
- Maness, J. M. (2006). *Library 2.0 theory: Web 2.0 and its implications for libraries*. *Webology*, 3(2)
- Miller, P. (2005). *Web 2.0: Building the new library*. *Ariadne*, 45
- NISO. (2004). *Understanding metadata*. Retrieved from: https://www.lter.uaf.edu/metadata_files/UnderstandingMetadata.pdf
- O'Reilly, T. (2005). *What is web 2.0: Design patterns and business models for the next generation of software*.
- Shirky, C. (2005). *Ontology is overrated: Categories, links, and tags*. *Clay Shirky's Writings about the Internet*,
- Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. London: Allen Lane.
- Smith-Yoshimura, Karen and Cyndi Shein. 2011. *Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews*. Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2011/2011-02.pdf>.
- Surowiecki, J. (2004). *The wisdom of crowds : Why the many are smarter than the few*. London: Abacus.
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, Mass. ; London: MIT Press.
- Tammaro, A. M. (2016). *Heritage curation in the digital age: Professional challenges and opportunities*. *International Information & Library Review*, 48(2), 122-128. doi:10.1080/10572317.2016.1176454
- UNESCO. (2003). *Charter on the preservation of digital heritage*. Retrieved from http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html

- Vander Wal, T. (2007, February 2). Folksonomy coinage and definition [Web log post]. Retrieved from <http://vanderwal.net/folksonomy.html>
- Weinberger, D. (2005). Tagging and Why It Matters. Retrieved from <http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/07-WhyTaggingMatters.pdf>
- Weinberger, D. (2007). Everything is miscellaneous: The power of the new digital disorder. New York, N.Y.: Henry Holt.
- Weinberger, D. (2014). Too big to know: Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room. New York: Basic Books.
- Wright, A. (2014). Cataloging the world: Paul Otlet and the birth of the information age. New York: Oxford University Press.
- Wright, A. (2007). Glut: Mastering information through the ages. Washington, District of Columbia: Joseph Henry Press.
- Zeng, M. L., & Qin, J. (2016). Metadata (2nd ed.). London: Facet Publishing.