JLIS.it

Bibliographic control in the fifth information age

Gordon Dunsire[a]

a) Independent Consultant, http://orcid.org/0000-0003-2352-0802

ABSTRACT

Bibliographic control is concerned with the description of persistent products of human discourse across all sensory modes. The history of recorded information is punctuated by technological inventions that have had an immediate and profound effect on human society. These inventions delimit five 'information ages'. It is now the Fifth Information Age, characterized by the ubiquitous use of powerful portable information processing devices for peer to peer communication across the entire planet. All such discourse is recorded during transmission and is copied to persistent storage media.

In the Fifth Information Age, the end-user is immersed in and interacts with a global ocean of recorded information. The interaction is continuous and ubiquitous, and never passive. Every interaction increases the volume of data; all aspects are recorded, including the time, place, and nature of the interaction, and details of the 'reader' and their 'book'. The roles of cave 'artist', scribe, printer, publisher, encoder, broadcaster, librarian, and other mediators are no longer differentiated from 'author'. The distinction between data and metadata is completely blurred: data becomes metadata as soon as an information resource is named by its creator.

The challenge for bibliographic control is the reconciliation of globalization and personalization via localization. The bibliographic ecosystem is very different and the activities and imploded roles of the end-user must be taken into account by professional agents.

KEYWORDS

Bibliographic control; Semantic Web; metadata; information retrieval.

JLIS.it

Paper

The context in which 'bibliographic control' takes place has been evolving at a fast pace for the past 30 years. Usage of the term was initially confined to written materials held in library collections, but has broadened to cover a wider range of information resources held in a wider range of collections. As a result, it is necessary to clarify the definition that is used in this paper.

The report of the Library of Congress Working Group on the Future of Bibliographic Control published in 2008 defines bibliographic control as "the organization of library materials to facilitate discovery, management, identification, and access" (Library of Congress Working Group on the Future of Bibliographic Control, 2008).

The IFLA Library Reference Model (LRM) published in 2017 is intended to cover "everything considered relevant to the bibliographic universe, which is the universe of discourse …" (Riva, Le Bœuf, and Žumer 2017, 20). The LRM is an entity-relationship model that consolidates three previous models for bibliographic records, authority data, and subject authority data published by the IFLA (International Federation of Library Associations and Institutions) as part of its development of "universal bibliographic control" (UBC). Although IFLA ceased its core support for UBC in 2003, development of bibliographic standards continues and the concept of UBC was "reaffirmed" as a set of principles in 2012 (IFLA, 2012). These principles are focused on the role of national bibliographic agencies and international coordination, and they include archives and museums in their scope.

The scope of the LRM is given by the definition of its broadest entity "Res": "Any entity in the universe of discourse" (ibid.). Dictionary definitions for the term 'discourse' emphasize written or spoken communication, and some specify a scholarly or formal context. For example, as of April 12, 2021, the online dictionary Dictionary.com gives two general definitions: "communication of thought by words; talk; conversation" and "a formal discussion of a subject in speech or writing, as a dissertation, treatise, sermon, etc." However, the LRM clearly intends a broader scope, beyond language-based materials, by giving examples of image, cartographic, and music resources. The LRM also restricts the definition to recorded communication: a resource is assumed to be embodied in a persistent carrier that can be accessed in the future, so speech must be recorded or transcribed if it is to be described.

The term 'bibliographic control' is defined by Dictionary.com in April 12, 2021 as "the identification, description, analysis, and classification of books and other materials of communication so that they may be effectively organized, stored, retrieved, and used when needed". No distinction is made between archive, library, and museum collections, and objects of control are "materials of communication".

This paper will therefore assume that bibliographic control includes all forms of recorded human communication. The 'bibliographic universe' is the set of all products of human discourse that forms the collective memory of Homo sapiens, and 'bibliographic control' is its management for future access and use.

Relevance

The bibliographic universe requires control because the organization of human memory is nec-

JLIS.it

essary for social cohesion and cultural evolution. Recorded discourse is communication through time and across distances greater than the unassisted range of human senses.

Recorded discourse carries the information that allows humans in different family groups to co-operate with each other in larger social units. The persistence and accumulation of recorded memory drives culture and its evolution. The inheritance of recorded memory is essential for cultural identity; the bibliographic universe is synonymous with cultural heritage. The management of recorded memory improves its utility and functionality in this context.

Recorded memory is an intermediary stage in the communication of a message from one person to another. The message is transmitted and then frozen in time; the message waits to be received at some unknown time in the future by some unknown person. The focus of bibliographic management is therefore the connection between the message and the receiver: what happens, after the memory is recorded, to the product that is recorded discourse?

The five laws of library science proposed by S.R. Ranganathan support this point of view (Ranganathan, 1931). The need for bibliographic control is driven by all five of the laws, although the terminology reflects a narrow focus on the written, and in particular printed, products that characterized libraries at the time. As of April 12, 2021, the Wikipedia article on "Five laws of library science" describes several subsequent attempts to modernize the scope of the laws and augment them to take account of the impact of more recent innovations in communication and information technologies.

The second and third laws are "Every reader his or her book" and "Every book its reader" respectively. The model is readily extended to all of recorded memory: the 'book' is the message, the recorded memory, the product of human discourse, and the 'reader' is the receiver of the message. The terms will be used in this paper with these general meanings.

The first and fourth laws are "Books are for use" and "Save the time of the reader" respectively. The primary factors affecting the delivery of the book to its reader – the recorded message to its recipient – are its portability, reproducibility, and findability. Portability determines if the book is taken to the reader, or the reader to the book. Reproducibility determines if the book can be accessed by more than one reader at a time. Findability determines if the book exists and how it is to be accessed by the reader. This last factor is the realm of bibliographic metadata: data about data, a book that describes other books so that readers can access their contents, the organization of the products of recorded memory.

The fifth law is "A library is a growing organism". The number of books increases over time. Recorded memory grows as time goes by.

Information ages

The ongoing evolution of human society and culture is punctuated from time to time by an innovation in communication technology that has a revolutionary impact. Such an innovation is followed by a significant increase in the complexity of interactions and activity across all social groups world-wide. Profound changes take place in commercial, legal, religious, and other cultural systems that affect all aspects of personal life.

Four specific innovations have had the greatest impact on the recording of human discourse. These are writing, printing, telecommunication, and the Internet.

JLIS.it

Each innovation provides a fundamental change in one or more of the basic aspects of preserving human memory and providing subsequent access to it. This results in a significant change in basic cultural and social concepts and processes; a paradigm shift. The innovation evolves through further invention and continues to influence many aspects of social interaction and development until the next innovation. It is useful to categorize the timespan between innovations as an 'age', and specifically as an 'information age'. The beginning and ending of each timespan are not precise dates, and they vary from place to place. Individuals and groups may recognize the potential for change that the innovation represents, but the actual impact of the innovation is not predictable during and immediately after the transition. Four innovations delimit five information ages; the present is the Fifth Information Age.

First Information Age

The First Information Age is the timespan before the invention of writing. It is pre-literate by definition, and is labelled "prehistoric" despite the existence of products of human discourse in the form of images and manufactured objects.

The production of a painting or sculpture takes time and requires specialist skills and tools, so such products are expensive. The fragility and perishability of available carrier materials means that only objects made of hard substances such as stone and images preserved under rare special conditions have survived. How widespread was the recording of human discourse is not knowable, but human groups were nomadic and small: Paleolithic and Mesolithic hunter-gatherers.

In this age, most social and cultural memory is conveyed into the future, beyond the individual memory of a person, through an oral tradition that cannot be recorded (until the invention of writing).

The content of the discourse that is recorded is mostly representational, depicting the things of interest in the local environment. Some content is symbolic and abstract, but the context is unknown. The meaning or intention of recording the content cannot be determined; only the 'art' can be appreciated in the context of modern aesthetics.

Reproduction of the recorded memory is as expensive as manufacturing the original. Each carrier of the content is a one-off, a singleton manifestation in the terminology of the LRM.

Access to recorded discourse is very limited. Images carried by cave paintings are often located in the furthest reaches of the cave. The reader must be taken to such a book to access it, and this seems to have been a religious or ritualistic activity. Portable sculptures must be small and light enough to be transported along with the other possessions of hunter-gatherer social units. Fragile carriers such as wood and soft stone are easily destroyed, small objects are easily lost, and such books are very rare. What has survived is now curated in museum collections.

Second Information Age

The Second Information Age begins with the invention of writing, the symbolic representation of language. Writing allows the recording of linguistic discourse. The act of speaking is readily transferred to the acts of writing and reading. The recording of discourse in specific aspects of human culture becomes common-place.

JLIS.it

The content of recorded linguistic discourse is descriptive and much more expressive than images and objects. There is immediate benefit in recording the 'word' in commercial, legal, and religious systems; social agreement is no longer reliant on the oral tradition or individual human memory. Peer-to-peer communication over long distances between persons who are known to each other, the writing of letters, becomes possible.

In this age, carriers remain singletons, such as manuscripts and paintings, but reproduction requires only the skills of the scribe or copyist. Reproduction has the same costs as the manufacture of the original manuscript, but this is less expensive than copying a painting or object. The process of reproduction is industrialized with the development of the scriptorium. Centralization of reproduction leads to centralization of storage, and the first libraries appear.

Access to recorded memory becomes easier. Readers who can travel independently can go to the scriptorium or library. Writing is applied to flat surfaces, and the third dimension of the cave or figurine is not required. This allows and encourages portability by embodying the message in materials such as clay, bark, bone, and textiles. Some writing is monumental, such as the Code of Hammurabi stele, and the reader must go to the book, but many products of discourse can be carried by hand to the reader. Not many survive because of the perishability of portable carriers.

Third Information Age

The Third Information Age begins with the mechanization of printing. Printing is a development of the industrialization of writing that involves the mechanical reproduction of writing and images. Development of the technology begins in the Second Information Age with the use of seals for stamping text onto clay or paper. The content is usually a name that confers ownership or authority on an accompanying manuscript. The technique evolves to cover the content of a page of text or a drawing in a larger stamp made of wood, stone, or some other hard material that can be sculpted. This speeds up the production of copies of texts and images, but preparing a seal or stamp is expensive and the range of discourse that is recorded in this way remains very limited.

The Second Information Age ends with the development of movable type and printing presses which industrialize the mechanics of reproduction. Manufacture and reproduction of the products of discourse becomes much less expensive, and there is a corresponding increase in the quantity of such products. Reproduction becomes part of the process, and the existence of multiple identical copies becomes the norm. The products of recorded discourse become more common-place, but are mediated by the printer who has the skills to set the type and operate the press.

There is an immediate and significant increase in the range of persons whose memory is recorded. A greater proportion of depictive content is manufactured and distributed using the new technologies, to cater for readers who are illiterate or who do not understand the language of a text; a picture bridges linguistic barriers. Scholarly communication becomes industrialized with the development of printed journals.

Access becomes easier. The reader has a choice of copies of the book, located in multiple places, and the book is easy to transport. Printers and booksellers become 'high street' services, and modern libraries begin to develop.

JLIS.it

Fourth Information Age

The Fourth Information Age begins with the invention of digital telecommunication. The development of the transmission of information over large distances required new techniques for correcting signal errors while increasing the size of the message; this stimulated the evolution of digital technologies.

Most forms of telecommunication require the message to be encoded so that it can be transmitted. The message is decoded back into its original form when it is received. The application of telecommunication technologies to discourse usually requires the discourse to be recorded as part of the encoding and decoding processes.

Encoding allows all forms of content to be transmitted, including music, speech, and static and moving images. In this age, the range and quantity of recorded discourse increases again. Electromagnetic media become available for the persistent storage of memory. Digital encoding allows the content and carrier of the book to be created, manufactured, distributed, and accessed in an integrated, seamless, and intangible infrastructure. Reproduction is unavoidable and invisible; a temporary copy of the product of discourse is automatically created in every encode/decode transaction and it is trivial to make that copy persistent.

There are no physical barriers to access, and access becomes localized; the book always goes to the reader, wherever the book and the reader may be. Transportation is instantaneous; the reader gets the book when and where the reader wants it.

Fifth Information Age

The Fifth Information Age begins with the invention of the Internet. The Internet globalizes digital telecommunication networks linked to powerful data processing machines and allows the participation of nearly every living human in discourse over a distance.

Digital encoding and decoding are a necessary process for discourse using the Internet. All discourse is recorded on persistent digital media. The deletion of recorded memory, "the right to be forgotten" (ICO, n.d.), has become a cultural and social issue, in a complete reversal of the First Information Age and 'the right to remember'. An example of the impact on bibliographic control is the initiative by NISO on "author name changes" (NISO, 2021)

The World-Wide Web is an application of the Internet that allows any person to take on and combine the roles of author, publisher, printer, distributor, and reader. The book includes every email, social media post, chat or webinar conversation, blog, website, or search ever made by every reader.

Reproduction is a built-in automatic feature. Overt reproductions of recorded memory are made to ensure persistence of cultural heritage, improve access, and retain evidence of discourse.

The "Internet of things" is a result of the miniaturization of computer chips as digital encoding, storage, and decoding devices. The reader and the book exist in the same local space and time. The perceived benefits of allowing 'all cookies' ensures that recording is ubiquitous and constant; the 'user' is immersed in an ocean of recorded/recording memory. The reader is every individual human; the book is a collection of all digital human memory.

JLIS.it

Metadata

The development of metadata for bibliographic control arises in the Third Information Age. The quantity and availability of printed products stimulated an increase in collections of recorded memory by social groups and individuals. Such collecting began in the Second Information Age with the development of libraries of manuscripts, but these were rare because of the expense of obtaining or reproducing hand-made products. Printing allowed wealthy individuals to accumulate private collections for pleasure, research, and status, and for a greater range of commercial, legal, religious, and scholarly organizations to develop repositories of information to support their activities. As collections grew in number and size, it became useful to record the collector's memory of what the collection contained, and to organize access to the collection to find and select a specific product of discourse. Is the item in the collection, and if so, where is it located? "As the number of books available to collectors like [Hernando Colón] grew, and new ways of organizing them became necessary, a list of authors in alphabetical order probably seemed a fairly unproblematic place to start … the alphabetical list forces the librarian, and the users of the library, to attribute each of the books to a single, named author, in a sense 'inventing' the notion of the author (or at least their centrality) as a matter of necessity" (Wilson-Lee, 2018, 209-210).

The content of metadata is essentially descriptive, and therefore linguistic in form. Textual metadata can be sorted and ordered using the syntax of the language of description, and it is much easier to formulate search and retrieval queries in the same syntax. Textual metadata can be transformed into spoken word, using a screen-reader, or visual symbols such as colour-coded categorizations. On the other hand, depictive metadata content is of limited utility. A thumbnail image is a representation or depiction of the whole image, not a description of it. Essentially, the reader reads a (metadata) book in order to find a (data) book.

The Third Information Age therefore stimulated and supported the printing of metadata as a result of the printing of books. The Fourth Information Age stimulated the internationalization of metadata creation, reproduction, and distribution. The MARC formats were initially developed to be "a vehicle for the exchange of bibliographic information between systems with independent computer facilities" (Morton, 1986). The Fifth Information Age allows the reader to be the author and publisher of metadata – the cataloguer – as well as being the author and publisher of a book that is being described.

Current approaches to metadata are rooted in the paradigms of the Third and Fourth Information Ages. The impact of the Fifth Information Age on bibliographic control is at its beginning and the detail belongs to the unknown future, but it will be profound. Some of the main characteristics of the bibliographic future are already emerging, including identity management, data provenance, open world application, and the authenticity of consensus.

Identity management

The management of identity is essential to the functionality of metadata. An identifier is a label that distinguishes the referent from other things. Effective information retrieval processes require that the subject of a metadata description is identified: is the individual book or associated entity that is being described the one that the reader wants?

JLIS.it

Identity management is the basis of classical authority control, a development of the concept of 'author' from the Third Information Age. The nature of discourse, and human culture itself, differentiates names and titles in specific social contexts only; there is no global system that makes the distinction based on universal physical contexts such as space and time. A person is not a cultural artefact, but is a natural phenomenon that cannot exist in two places at the same time. The same person has different names; the same name can refer to multiple persons. This is surely a result of larger, settled groups in the Second Information Age. More generally, the same individual is labelled with different identifiers, and the same identifier is used for different referents, across different human cultures. Much of this diversity is driven by local context and by the difficulties of assigning identifiers that are agreed at global level.

The Fourth Information Age stimulated the development of global approaches to identifier management, generally limited to the book and its trade. Examples include the International Standard Bibliographic Number (ISBN) and International Standard Serial Number (ISSN) systems. The beginning of the Fifth Information Age saw the development of similar approaches to the identities of persons, including the author and therefore ultimately the reader, such as the International Standard Name Identifier (ISNI) and ORCID. However, it is not always a single person or group of persons that is being identified, and the cultural confusion of names and named persists, as ISNI's name suggests. As of April 12, 2021, the ISNI website states that it covers "public personas … such as pseudonyms, stage names, record labels or publishing imprints"; the LC/NACO Name Authority File remains under active development. The LRM includes an entity Nomen, the class of names of things, that is distinct from the things, such as agents, places, and timespans, themselves. This allows description of the name, such as usage, language, etc. to be separated from description of the thing that is named.

However, the Fifth Information Age eliminates half of the general problem, of the same identifier being used for different referents. The Internationalized Resource Identifier (IRI) system, based on the Uniform Resource Identifier (URI), is applicable to anything that can be described; that is, any thing that is the subject of bibliographic metadata. This is one of the necessary and fundamental aspects of the Internet, the World-Wide Web, and the linked open data of the Semantic Web. It is managed independently of any cultural application or context.

The assignment of more than one identifier to an individual thing cannot yet be eliminated. That would require all of the assigners of identifiers to agree on a preferred identifier and to supply a means of de-referencing it to a description of the thing it identifies. This was the approach of IFLA's UBC programme, and is the antithesis of the bottom-up construction of the Semantic Web. In the Fifth Information Age, authority control evolves into the management of linked data identifiers. The application of automated reasoning to connect the reader to the book is completely dependent on consistent and complete assignment of IRIs to readers, books, and associated entities. It is important that there is no ambiguity in what is being identified within the chosen data model, such as the LRM or BIBFRAME. The rules used in semantic reasoning are simple and they are applied by dumb machines; it is the metadata that is 'smart'.

JLIS.it

Data provenance

The Semantic Web is a globalized metadata retrieval system built on the World-Wide Web. It is based on description logic and has no intrinsic accommodation of "truth". The Semantic Web adheres to the AAA Principle: "anybody can say anything about any thing"; this is alternatively known as the AAA Slogan: "anybody can say anything about any topic" (Allemang and Hendler, 2011, 27). What is said in metadata may be true or false, in the same way that the content of any product of discourse may be true or false relative to the context in which it was created. Statements may be true when recorded, but are false when they are replayed; things change. Statements may be known to be false when recorded. "This statement is true" may be fake, and its author a liar. This is not just a cultural phenomenon. Discourse itself has in-built paradox, ranging from the "impossible" images of M.C. Escher to the linguistic paradox of Epimenides: "This statement is false" is false if it is true, and true if it is false.

These uncertainties mean that effective bibliographic control requires provenance for metadata. This is metadata that describes metadata, and has similar functionality to data provenance or "detailed information about the origin of data" (Glavic and Dittrich, 2007). For bibliographic metadata, provenance includes information about the author (cataloguer, curator, etc.), the application of content and encoding standards, and the date of creation. Data provenance has been accommodated in bibliographic control from the Fourth Information Age to support the coordination of shared catalogue records. For example, this is provided by leader and control fields in MARC formats, such as "Date and Time of Latest Transaction" (Library of Congress Network Development and MARC Standards Office, 1999). Another latent example is the use of brackets in International Standard Bibliographic Description (ISBD): "Square brackets enclose information found outside the prescribed sources of information and interpolations in the description" (ISBD Review Group, 2011, 22). The recording of bibliographic data provenance for more general purposes is given specific accommodation in the development of more recent standards such as RDA: Resource Description and Access (RSC Technical Working Group, 2016).

Provenance is a means of quality control. Knowing who created metadata helps to distinguish high-quality data created by trained professionals with ethics from low-quality data created by amateurs with bias. It is also important to know when metadata was created and what standards were used. Metadata theory and practice evolve just as much as any other form of discourse. How things were described in the past may be useless or misleading in a contemporary context. Provenance allows metadata from disparate sources to be aggregated without 'one bad apple' lowering the quality overall.

Open world

The Semantic Web also makes the Open World Assumption (OWA). The assumption is that the absence of metadata is not a description of absence, but simply a description that has not yet been made. Metadata may be added in the future, and there is no expectation that future metadata will be objectively or subjectively true. This is a consequence of the AAA principle and the paradox of discourse: there cannot be a complete description of a thing because an infinite number of false or unprovable statements can be added.

JLIS.it

Applications of bibliographic metadata based on closed-world assumptions become less efficient in the Fifth Information Age. A bibliographic record can no longer be a fixed and complete description of a book or the entities associated with it. Metadata will always accumulate, so the size of the 'record' increases through time. It is unlikely that any single application will need or want to use the whole set of metadata that describes an entity, but the set exists and cannot be ignored. The closed-world practice of updating erroneous or incomplete metadata is no longer tenable. Instead, it must be assumed that the original statement of metadata is 'out in the field' in multiple information retrieval systems where it is not feasible to update every copy. Revisions are made with new statements; erroneous statements are assigned appropriate data provenance.

Wikis that share data from multiple authors without central mediation have been involved in conflicts where statements are updated by one author and 'updated' back to the original statement by another author. Each author wants their version to be published and the other's version to be discarded. For example, Wikipedia has a published policy on "dispute resolution" that seeks consensus before arbitration is invoked. As a result, data provenance and version control systems built-in to wiki software have become an important tool in quality control and assurance. Nothing can be truly deleted in a wiki, and amendments can be 'rolled-back' to a previous version. Similar systems are required for metadata.

Imposing fees for the use of metadata in wide-area applications or for the copying of metadata to use in local applications is a barrier to the utility of metadata in the Fifth Information Age. It prevents open linking and discourages the reader's contribution of metadata to the global pool, for example through passive cookies or active crowd-sourcing.

Consensus

If any reader can make any metadata statement they want, with no distinction between 'fact' and 'fiction', how can any consistency or authenticity be determined?

In the Fifth Information Age, recorded discourse is cultural memory, and metadata is the organization of culture itself. What makes local culture consistent is local consensus. A social group agrees to a particular set of truths, reflected in its recorded memory, to maintain a consistent and persistent world view.

Consensus in metadata can be determined through analysis by machine and by the human mind. Statistical analysis of large sets of metadata accumulated from multiple sources can calculate consensus by matching similar statements and by using data provenance to detect bias from particular sources. This is basically how search engines work; relevance is determined by the automatic analysis of the links on a webpage, where the focus of the page is assumed to be the subject of the link, and the links to a webpage, where the page is the target of the link. The link itself is metadata; the subject and target are associated in some way.

Linked open data in the Semantic Web can be processed using semantic reasoning, a standard set of algorithms that can derive metadata statements from metadata statements. These algorithms are simple, reflecting the simple 'atomic' structure of the linked data subject-predicate-object triple. They are not a substitute for human intelligence and culture. These automated techniques are a tool for cataloguers, not a substitute for cataloguers or other humans.

JLIS.it

Human analysis of metadata may be conscious or subconscious. The reader carries out such analysis throughout their information seeking and retrieval activity. The conscious analysis of the relevance of data is a form of 'ask the audience' in a quiz show. This is a core feature of social media in the Fifth Information Age, where the audience is invited to like or dislike (choose a binary review of) a piece of data, a mini-book. Consensus is reflected in the numbers of persons who like or dislike the information and the balance between them. This is a very broad measure of the 'authenticity' of data or metadata. A more refined approach is to crowd-source contributions for specific sets of books by specific sets of readers.

Subconscious analysis is now possible using eye-tracking technologies. The reader has no control of how their eyes read a book or description of a book. Experiments show that it is not the linear scan that it appears to be in the conscious mind. The development of virtual reality, mimicking the immersive cultural memory of the Fifth Information Age, will stimulate the use of subconscious feedback technologies.

Effectively 'author', 'authority', and 'authenticity' blur into the control of culture by consensus.

Conclusion

The future of bibliographic control is as unpredictable as the future of writing, printing, telecommunication, or the Internet when they first appeared. In every case, there has been an immediate impact on human discourse and recorded memory, followed by a slower but profound impact on every aspect of human culture. Although the dates may be imprecise and localized, the timespan of each information age decreases by at least an order of (decimal) magnitude, from tens of thousands of years through a few thousand and a few hundred years to a few decades.

Syntactically rooted in the Second Information Age, conceptually rooted in the Third Information Age, and mechanically rooted in the Fourth Information Age, bibliographic control is struggling in the Fifth Information Age. The range and quantity of products of recorded discourse requires a shift in the focus of bibliographic control, from top-down to bottom-up with the 'professional' cataloguer distinguished from other readers by context, not process.

Bibliographic control is likely to be based on the Open World Assumption. It will involve the coordination of metadata created by professionals and amateurs with metadata created by machine analysis. Data provenance is essential to achieve this by providing context and supporting the management of quality control. Metadata is common and necessary in the Fifth Information Age. It is a social and cultural 'good' that is not best controlled by commercial interests.

The purpose and function of bibliographic control is to manage cultural identity in a global framework. The distinction between data and metadata is no longer useful, and bibliographic control will become indistinguishable from culture. The Fifth Information Age is the technological extension and immersion of personal and social mind.

JLIS.it

References

Allemang, Dean, and Jim Hendler. 2011. *Semantic Web for the Working Ontologist,* Second Edition. Amsterdam, Morgan Kaufmann.

Glavic, Boris and Klaus R. Dittrich. 2007. "Data Provenance: A Categorization of Existing Approaches" In 12. Fachtagung des GI-Fachbereichs "Datenbanken und Informationssysteme", Aachen, Germany, 7 March 2007 - 9 March 2007:227-241. Accessed April 12, 2021. http://dx.doi.org/10.5167/uzh-24450

ICO. nd. "Right to erasure". Accessed April 12, 2021. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/

IFLA. 2012. *IFLA Professional Statement on Universal Bibliographic Control.* Accessed April 12, 2021. https://www.ifla.org/files/assets/bibliography/Documents/ifla-professional-statement-on-ubc-en.pdf

ISBD Review Group. 2011. *ISBD : International standard bibliographic description, consolidated edition.* Berlin, De Gruyter Saur.

Library of Congress Network Development and MARC Standards Office. 1999. *MARC 21 Format for Bibliographic Data: 005 - Date and Time of Latest Transaction (NR).* Accessed April 12, 2021. https://www.loc.gov/marc/bibliographic/bd005.html

Library of Congress Working Group on the Future of Bibliographic Control. 2008. *On the Record: Report of The Library of Congress Working Group on the Future of Bibliographic Control.* Washington, D.C.: Library of Congress. Accessed April 12, 2021. https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf

Morton, Katherine D. 1986. "The MARC Formats: An Overview" In *American Archivist* Vol. 49, No. I/Winter 1986:21-30.

NISO. 2021. "NISO Members Approve Proposal for a New Recommended Practice to Update Author Name Changes". Accessed April 12, 2021. http://www.niso.org/press-releases/2021/04/niso-members-approve-proposal-new-recommended-practice-update-author-name

Ranganathan, S. R. 1931. *The Five Laws of Library Science, etc.* Madras, Madras Library Association.

Riva, Pat, Patrick Le Bœuf, and Maja Žumer. 2017. *IFLA Library Reference Model: a Conceptual Model for Bibliographic Information.* Den Haag, IFLA. Accessed April 12, 2021. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

RSC Technical Working Group. 2016. *RDA models for provenance data.* Accessed April 12, 2021. http://www.rda-rsc.org/sites/all/files/RSC-TechnicalWG-1.pdf

Wilson-Lee, Edward. 2018. *The Catalogue of Shipwrecked Books: Young Columbus and the Quest for a Universal Library.* London, William Collins.