

Görüşler / Opinion Papers

A Brief Introduction to Metadata

Ann Hanlon*

Abstract

The term metadata is becoming ubiquitous in library circles. This short introduction to metadata attempts to give a brief description of what metadata is and what new ends it represents. This article also proposes some key questions about metadata: What attributes make metadata a tool for more uniform resource discovery for electronic resources on the World Wide Web? And how does metadata differ from our more traditional cataloguing tools, such as MARC? Finally, I offer a short annotated bibliography to more in-depth information about metadata.

Keywords: *Metadata, Internet, World Wide Web, Electronic resources-cataloguing, MARC.*

Introduction

The term metadata is becoming ubiquitous in library circles, but what does it mean? The concept is one that librarians, especially cataloguing librarians, are already familiar with: creating information about information in order facilitate discovery and to place information in an orderly fashion with other related and unrelated information. So what sets metadata apart from what librarians have already been doing, and what new ends does it represent?

In many ways, metadata is an extension of cataloguing. It has been proposed as a way to provide more uniform resource discovery for electronic resources, particularly on the World Wide Web. And in that respect, it provides much of the same information we have been filling catalogue records

* *Electronic Resources Manager, Bilkent University Library.*

Metadata Kavramına Kısa Bir Giriş

Mehmet Emin Küçük* (Çev.)

Özet

Metadata terimi kütüphanecilik alanında sürekli karşımıza çıkan terimlerden bir haline gelmiştir. Konuyla ilgili bu kısa giriş, metadatanın ne olduğu ve ne tür yeni gelişmeler getirdiğine ilişkin kısa bir tanıtımı hedeflemektedir. Bu makale ayrıca metadatta hakkında bazı temel soruları ortaya koymaktadır: Hangi nitelikleri metadatayı World Wide Web üzerinde elektronik kaynakların tek biçim kaynak keşfinde bir araç yapmaktadır? Ve MARC gibi daha geleneksel kataloglama araçlarımızdan nasıl bir farklılık göstermektedir? Son olarak, metadatta hakkında daha ayrıntılı bilgi içeren açıklamalı bir kısa bibliyografya sunulmaktadır.

Anahtar sözcükler: Metadata, Internet, World Wide Web, Elektronik kaynaklar-kataloglama, MARC.

Giriş

Metadata terimi kütüphanecilik alanında sürekli karşımıza çıkan terimlerden biri haline gelmektedir, fakat metadatta ne demektir? Metadata kavramı kütüphanecilere, özellikle katalogculara, kaynak keşfini gerçekleştirebilmek için bilgi hakkında bilginin yaratılması ve bilgiyi diğer ilişkili ve ilişkisiz bilgilerle bir düzen içinde bir araya getirmek özellikleriyle yabancı değildir. Bu durumda, kütüphanecilerin hâlihazırda yaptıklarından metadatayı ayıran nedir ve metadatta ne tür yenilikler ortaya koymaktadır?

Birçok yönden metadatta kataloglamanın bir uzantısıdır. Özellikle World Wide Web üzerinde, elektronik kaynakların daha fazla tek biçim kaynak keşfinin sağlanmasında bir yol olarak önerildi. Ve bu açıdan yaklaşıldığında, onlarca yıldır katalog kayıtlarına girmekte olduğumuz –eser adı, yazar, yayın tarihi, yayıncı, konu vb.- aynı bilginin çoğunu içermektedir. Ancak, yeni meta-

* Yrd. Doç. Dr.; Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü (mkucuk@hacettepe.edu.tr).

with for decades – title, author, date of publication, publisher, subject matter, etc. However, there are a few key reasons for the development of new metadata schemes: One is to simplify cataloguing so that non-librarians are able to create metadata for the fast growing resources available on the World Wide Web. A second reason is to make it possible to create standard metadata tags that can be assigned to materials from a wide range of communities and across a wide range of document types – not only for web pages and document-like objects, but for images, and interactive resources as well.

All of this activity is, of course, the result of the Internet. Judith Ahronheim, Head of Original Cataloging at the University of Michigan explains that, “as the Internet has grown and developed, it has metamorphosed from a medium for communication to a medium for resource discovery, a function that was not in the mind of its originators” (1). New metadata standards are being developed to aid resource discovery in the shared environment of the World Wide Web, while traditional cataloguing was developed for the closed environment of the online public access catalogue (OPAC).

Furthermore, the resources that are available online represent a much wider range than had been catalogued in libraries. It is conceivable to go online and retrieve traditional text documents in electronic form, as well as images, animation, digital surrogates for primary resources, and interactive resources, all in a single sitting. The point of metadata, like cataloguing, is to make that retrieval more precise.

And while metadata is being considered as a solution for more precise discovery on the World Wide Web, it has documentary and preservation functions as well. There are three conceptual types of metadata:

1. Descriptive, 2. Structural, and 3. Administrative:

1. Descriptive metadata is the most similar to traditional bibliographic cataloguing. This is the information that is used to discover an electronic resource and includes information such as author, title and subject. For example, MARC fields 100, 245 and 600, and the AACR2 content standards are forms of descriptive metadata. The metadata standards Dublin Core, EAD, and TEI all include similar descriptive elements as well.

data şemalarının geliştirilmesinde birkaç temel neden bulunmaktadır: İki, World Wide Web (Web) üzerinde elde edilebilen hızla artan kaynaklara yönelik metadatanın kütüphaneci olmayanlar tarafından oluşturulabilmesi için kataloglamayı basitleştirmektedir. İkinci neden ise, çok farklı belge türlerine (sadece web sayfaları ve doküman benzeri nesnelere değil, görüntülere ve etkileşimli kaynaklara da) ve çok farklı gruplardan gelen materyale verilebilecek standart metadata etiketlerinin (tag) yaratılmasını olanaklı kılmaktır.

Tabii ki tüm bu faaliyetler İnternet'in bir sonucudur. Michigan Üniversitesi Özgün Kataloglama Birimi sorumlusu Judith Ahronheim, "İnternet, büyümesi ve gelişmesiyle birlikte bir iletişim medyasından, yaratıcılarının aklında olmayan bir fonksiyona, bir kaynak keşif aracına (medya) dönüşmüştür" (1) açıklamasını yapmaktadır. Geleneksel kataloglama, çevrim içi halka açık katalog (OPAC) dar çevresi için geliştirilmişken, yeni metadata standartları World Wide Web'in paylaşımlı çevresinde kaynak keşfine yardımcı olmak için geliştirilmektedir.

Bundan başka, çevrim içi olarak elde edilebilen kaynaklar, kütüphanelerde kataloglanan kaynaklardan çok daha geniş yelpazede yer almaktadır. Görüntülere, canlandırılmalara, temel kaynakların sayısal kopyalarına ve etkileşimli kaynaklara olduğu gibi elektronik formdaki geleneksel metin dokümanların hepsine, bir oturumda çevrim içi erişim olarak düşünülebilir. Metadata'nın hedefi, kataloglamada olduğu gibi, erişimi daha isabetli (precise) kılmaktır.

Metadata, World Wide Web'de daha isabetli keşifde bir çözüm olarak düşünülürken, dokümantasyon (documentary) ve koruma (preservation) fonksiyonlarına da sahiptir. Metadata'nın üç kavramsal tipi bulunmaktadır: 1. Tanımlayıcı (descriptive) 2. Yapısal (structural) 3. Yönetimsel (administrative):

1. Tanımlayıcı metadata, geleneksel bibliyografik kataloglamaya en benzer olanıdır. Yazar adı, eser adı ve konu gibi bilgileri içeren ve elektronik kaynağın keşfedilmesinde kullanılan bilgi, tanımlayıcı metadata'dır. Örneğin 100, 245 ve 600 MARC alanları ve AACR2 içerik standartları, tanımlayıcı metadata biçimlerindenidir. Dublin Core, EAD ve TEI metadata standartlarının hepsi de benzer tanımlayıcı unsurları içerirler.

2. Structural metadata provides information necessary to activate and view an electronic resource. In the case of traditional bibliographic resources, structural metadata might include information about separate essays in an anthology or the contents of multiple records in an album set. For electronic resources, this information might also include the names of applications necessary to view the resource (QuickTime, RealPlayer, etc...), or the scheme used to encode it (HTML, PDF, etc.).
3. Administrative metadata includes rights and management information, such as copyright status, contact information, and publication date. In the case of electronic resources, it might also include version numbers and update information.

There are many metadata schemes in development. Among the most well developed and essential to the library community are the Dublin Core Metadata Initiative (DCMI), EAD (Encoded Archival Description) and TEI (Text Encoding Initiative). Perhaps the most popular scheme for electronic resources is the Dublin Core, a metadata scheme developed in cooperation with librarians at OCLC. Comprised of a set of 15 elements, the Dublin Core is intended to be a simple scheme for describing a wide range of electronic resources. EAD is being adopted at archives and special collections in order to make their finding aids electronic and searchable on the World Wide Web. Accordingly, the descriptive, structural and administrative elements in EAD are specifically geared towards the elements of a finding aid, while Dublin Core is more amenable to document-like objects. The TEI is an encoding scheme developed for electronic textual materials and is intended as header type information at the beginning of the encoding of an electronic document to record bibliographic history, and provenance information as well as information about the electronic text itself.

How are these elements encoded? Traditional cataloguing records have existed separately from the resources they describe. In the case of electronic documents it is possible to include metadata information in the electronic encoding of the document.

Most of us are familiar with HTML, the markup language used to make documents available on the World Wide Web. HTML itself is an encoding

2. Yapısal metadata bir elektronik kaynağı aktif hale getirme ve göstermede gerekli bilgiyi sağlar. Geleneksel bibliyografik kaynakları örnek olarak ele alacak olursak, yapısal metadata bir antolojide yer alan farklı denemeler hakkında ya da bir albümde yer alan birden fazla müzik eserinin içeriği hakkında bilgi içermesine benzetilebilir. Elektronik kaynaklar için yapısal metadata bir kaynağı izlemek için gerekli uygulamanın adını (QuickTime, RealPlayer vb.) ya da kaynağı kodlamada (encoding) kullanılan şemayı (HTML, PDF vb.) içerebilir.
3. Yönetimsel metadata telif durumu, iletişim bilgisi ve yayın tarihi gibi haklar ve yönetim bilgisini içerir. Elektronik kaynakların durumunda güncelleme bilgisi ve sürüm (version) numarasını da içerebilir.

Geliştirilmekte olan birçok metadata şeması bulunmaktadır. Dublin Core Metadata Initiative (DCMI), EAD (Encoded Archival Description) ve TEI (Text Encoding Initiative) kütüphanecilik dünyasına gerekli olan ve en iyi geliştirilen şemalar arasındadır. Belki de elektronik kaynaklara yönelik en popüler şema, OCLC'de kütüphanecilerin iş birliğiyle geliştirilen Dublin Core'dur. 15 öğeden oluşan Dublin Core, farklı elektronik kaynakların tanımlanmasında basit bir şema olayı amaçlamıştır. EAD, malzemelerini elektronik olarak erişilir ve World Wide Web üzerinde taranabilir kılmak için özel koleksiyonlar ve arşivlerce kabul gördü. Buna bağlı olarak, Dublin Core daha çok doküman benzeri nesnelere yönelirken, EAD'daki tanımlayıcı, yapısal ve yönetimsel öğeler, keşif destek öğeleri olarak tasarlandılar. TEI, elektronik metinlere yönelik olarak geliştirilen bir kodlama şemasıdır ve elektronik metin hakkındaki bilginin yanı sıra provönans bilgisini içerir ve bibliyografik geçmişini kaydetmede elektronik doküman kodlanırken, başındaki başlık tipi bilgi olarak amaçlanmıştır.

Bu öğeler nasıl kodlanır? Geleneksel katalog kayıtları tanımladıkları kaynaklardan ayrı olarak tutulmuşlardır. Elektronik dokümanların durumunda ise, doküman elektronik olarak kodlanırken metadata bilgisini dokümana eklemek olanaklıdır.

Çoğumuz, dokümanları World Wide Web'de elde edilebilir kılmada kullanılan işaretleme dili HTML'i tanıyoruz. HTML'in kendisi bir kodlama standardıdır. Bir başka deyişle, tarayıcınıza (browser) bir dokümanı nasıl gösterece-

standard. In other words, it includes tags that tell your browser how to display a document. For example, the tag <p> tells the browser that a new paragraph is beginning, and the tag </p> tells the browser that the paragraph has ended. HTML is actually a subset of a standard called SGML or Standard Generalized Markup Language. SGML describes the logical structure of a document in order to communicate how it should be presented. Using something called a DTD, or Document Type Definition, it can be customised to respond to a wide range of metadata tags. EAD, TEI and even MARC have all devised or are developing DTD's to take advantage of the SGML's flexibility. New versions of the major browsers are being configured to read SGML and most variants on the DTD in order to take full advantage of the information that can be stored there.

Metadata or MARC?

Metadata is making a rather solid entrance into the librarian's toolbox, and with the importance of electronic resources indisputable, no librarian is in a position to ignore it. However, it is still important to consider metadata alongside the librarian's current machine-readable cataloguing standard, MARC. As pointed out above, metadata and MARC have similar goals, and MARC is even being fine-tuned to respond to the unique requirements of electronic resources and the Internet.

Michael Gorman, Dean of Library Services at California State University, Fresno, has taken up this issue and pointed out that metadata, like MARC, is merely a "framework" standard rather than a content standard (2). The fields in MARC are intended for particular content (245 is for main entry, the 100 field for personal name main entry, etc...), while AACR2 provides the content standards. There are a number of content standards under discussion for metadata, though as of yet there are no overarching standards. This is deliberate to a certain extent as one goal of the many metadata initiatives is to leave content decisions to the various communities.

Another elemental difference between MARC and metadata is in their complexity. MARC is a long established and well developed framework for bibliographic materials. Metadata is not an established standard in and of

ğini belirten etiketleri (tags) içerir. Örneğin <p> etiketi tarayıcınıza yeni bir paragrafın başladığını ve </p> paragrafın sona erdiğini belirtir. HTML aslında SGML ya da Standard Generalized Markup Language olarak isimlendirilen standarttan türemiştir. SGML, dokümanın nasıl gösterileceğine ilişkin iletişimi gerçekleştirmek için bir dokümanın mantıksal yapısını tanımlar. Bir DTD ya da Document Type Definition adı verilen bir şeyi kullanarak, farklı metadata etiketlerine karşılık vermede özelleştirilebilirler. EAD, TEI ve hatta MARC, SGML'in esneklik avantajından yararlanmak için DTD'ler geliştirmektedirler ya da planlamaktadırlar. Temel tarayıcıların (browser) yeni sürümleri SGML'i ve kaydedilebilen bilginin tüm avantajlarından yararlanabilmek için DTD üzerindeki değişkenleri okuyacak biçimde yapılandırılmaktadır.

Metadata mı MARC mı?

Metadata kütüphanecinin dünyasına oldukça yoğun bir biçimde girmekte ve elektronik kaynakların öneminin tartışılmazlığıyla hiç bir kütüphaneci metadata'yı gözardı edecek durumda değildir. Ancak, metadata'nın yanında kütüphanecinin mevcut makinece okunabilir standardı MARC'ı göz önünde tutmak hala önem göstermektedir. Yukarıda belirtildiği gibi, metadata ve MARC benzer amaçlara sahiptir. MARC hatta İnternet ve elektronik kaynakların diğerlerine benzemeyen gereksinimlerini karşılamada yeniden düzenlenmiştir.

Fresno'daki California State Üniversitesi kütüphane hizmetleri sorumlusu Michael Gorman konuyu ele almış, MARC gibi metadata'nın içerikten ziyade yapı standardı olduğunu belirtmiştir (2). AACR2 içerik standartlarını sağlarken MARC'ta yer alan alanlar belirli bir içeriği hedeflemiştir (245 eser adı temel girişi, 100 alanı yazar adı temel girişi gibi ...). Henüz öne çıkan bir standart olmamakla birlikte, metadata tartışmalarında yer alan bir dizi içerik standardı bulunmaktadır. Metadata girişimcilerinin amaçlarından biri de içeriğe yönelik kararları farklı topluluklara bırakmaktır.

Metadata ve MARC arasındaki bir diğer temel fark karmaşıklıkta yatar. MARC, bibliyografik araçlara yönelik olarak çok uzun zaman önce ve iyi bir biçimde geliştirilmiş yapıdır. Metadata bir standart olmadığı gibi bir standart olarak ta geliştirilmemiştir. Dublin Core gibi bazı metadata türleri diğerlerine göre daha yaygınlaşmıştır.

itself. Some varieties of metadata are more established than others, such as the Dublin Core.

In the case of the Dublin Core the goal is to make the framework as simple and straightforward as possible, or as Judith Ahronheim describes it, as “a kind of lowest common denominator”. In Diane Hillman’s introductory document for the Dublin Core she explains, “The Dublin Core element set has been kept as small and simple as possible to allow a non-specialist to create simple descriptive records for information resources easily and inexpensively while providing for effective retrieval of those resources in the networked environment (3).” In other words, Dublin Core is being developed as a less labour intensive alternative to traditional cataloguing, with the hope that the precision and uniformity of traditional cataloguing can somehow be retained or at least approximated.

Michael Gorman argues for more critical reflection on rationales such as these before we throw MARC overboard in favour of metadata for electronic resource description. As libraries in Turkey begin to grapple more and more with the implications of the World Wide Web and the demand for electronic resources, I propose that it is worthwhile to take his challenge to heart. Metadata standards have been developed with the attributes of electronic resources in mind, but MARC has recently added new fields, such as the 856 field, to accommodate the new features in electronic resources that are unique and unlike traditional bibliographic entities. And new initiatives such as CORC, (Cooperative Online Resource Catalog) from OCLC, use both Dublin Core and MARC to collectively catalogue online resources that have been selected by participating libraries. The cooperative nature of this venture helps make electronic resource description less labour intensive through the time honoured practice of copy cataloguing.

I hope that this very brief introduction will inspire other librarians in Turkey to consider the advantages and disadvantages of metadata, as well as to consider more closely the challenges and opportunities presented by the vast range of resources available on the World Wide Web. For a more in-depth discussion of the issues presented here, I have included a short annotated bibliography as well as some additional resources that provide more in-depth guides to metadata.

Dublin Core'un durumunda, amaç yapıyı mümkün olduğu kadar açık ve basit yapmak ya da Judith Ahronheim'in tanımladığı gibi "bir tür ortak yol"dur. Diane Hillman, Dublin Core'u tanıtan yazısında, "Dublin Core element seti, ağda yer alan kaynaklara erişimi sağlarken, uzman olmayanların bilgi kaynaklarının basit tanımlayıcı künyelerini kolay ve ucuzca oluşturması için mümkün olduğu kadar basit ve sınırlı tutulmuştur" (3) açıklamasını yapmaktadır. Bir başka deyişle, Dublin Core geleneksel kataloglamaya göre daha az emek gerektiren bir seçenek olarak, geleneksel kataloglamanın tek biçimliği ve erişim isabetinin bir şekilde korunabileceği ya da ona yakın olabileceği umuduyla geliştirilmektedir.

Michael Gorman, elektronik kaynakların tanımlanmasında metadata için MARC'ı bir kenara atmanın olumsuzluğunu tartışmaktadır. Türkiye'deki kütüphanelerin World Wide Web'in karmaşıklığıyla uğraşmaya ve elektronik kaynaklara talebin başlamasıyla, bu sorunla baş etmeyi ciddiye almanın önemli olduğunu belirtmek istiyorum. Metadata standartlarının elektronik kaynakların niteleyicileriyle geliştirildiğini göz önünde tutmakla birlikte, MARC elektronik kaynaklarda yer alan özel ve geleneksel bibliyografik araçlarda yer almayan özellikleri verebilmek için son dönemde 856 alanı gibi yeni alanlar eklemiştir. OCLC'nin CORC'u (Cooperative Online Resource Catalog) gibi yeni girişimler, üye kütüphanelerce seçilmiş çevrim içi kaynakların ortak kataloglanmasında hem Dublin Core'u hem de MARC'ı kullanmaktadır. Bu ortak girişim, bildiğimiz kopya kataloglama aracılığıyla elektronik kaynakların tanımlanmasını daha az emek yoğun kılmaktadır.

Umarım bu kısa giriş, World Wide Web üzerinde elde edilen kaynakların sunduğu olanakları, daha yoğun mücadeleyi, metadatanın avantaj ve dezavantajlarını göz önüne almada diğer Türk kütüphanecilere ilham verir. Bu yazıda aktarılan konuların ayrıntısına yönelik olarak, metadadata ayrıntılı açıklama getiren ilave kaynakların yanı sıra açıklamalı kısa bir bibliyografya aşağıda sunulmuştur.

References

- (1) Ahronheim, J. R. (1998). Descriptive metadata: Emerging standards. *Journal of Academic Librarianship*, 24(5): 395–403.

This is an excellent introduction to metadata. It also some sample records that might be very helpful in understanding how metadata is being applied. She also includes a helpful bibliography, though some links are now outdated.

- (2) Gorman, M. (1999). Metadata or cataloguing? A false choice. *Journal of Internet Cataloguing*, 2(1): 5–22.

Michael Gorman's presentation at the XVI Colloquium on Library Science Research is a thoughtful and thought provoking analysis of the drive toward metadata over current cataloging standards.

- (3) Hillman, D. (2001). Using Dublin Core. *Dublin Core Metadata Initiative. 2001 OCLC*. [Online]. URL: <http://dublincore.org/documents/20001/04/12/usageguide/> [12 April 2001].

Diane Hillman has written the latest introduction to the Dublin Core Metadata Initiative.

Additional Resources

- Dublin Core Metadata Initiative Official Website - <http://dublincore.org>
- Website of the TEI (Text Encoding Initiative) Consortium - <http://www.tei-c.org/>
- EAD (Encoded Archival Description) Official Website - <http://www.loc.gov/ead/>
- The XML Cover Pages - <http://www.oasis-open.org/cover/sgml-xml.html>

Robin Cover provides an overview as well as more in-depth information about SGML and XML.

Kaynakça

- 1 Ahronheim, J. R. (1998). Descriptive metadata: Emerging standards. *Journal of Academic Librarianship*, 24(5): 395–403.
Metadatataya mükemmel bir giriş. Metadatanın nasıl uygulandığının anlaşılmasında oldukça yardımcı olan örnek künyeleri de içermektedir. Bazı bağlantılar geçerliğini yitirmiş olsa da yararlı bir bibliyografya da bulunmaktadır.
- 2 Gorman, M. (1999). Metadata or cataloguing? A false choice. *Journal of Internet Cataloguing*, 2(1): 5–22.
Michael Gorman'ın XVI Colloquium on Library Science Research'teki sunumu mevcut kataloglama standartları yerine metadatataya yönelik eğilime nazik bir eleştirel analizdir.
- 3 Hillman, D. (2001). Using Dublin Core. *Dublin Core Metadata Initiative. 2001 OCLC*. [Çevrim içi]. Elektronik Adres: <http://dublincore.org/documents/20001/04/12/usageguide/> [12 April 2001].
Diane Hillman, Dublin Core Metadata Girişimi'ne ilişkin en son tanıtımı yazmıştır.

URL' / Ek kaynaklar:

- Dublin Core Metadata Initiative Official Website - <http://dublincore.org>
- Website of the TEI (Text Encoding Initiative) Consortium - <http://www.tei-c.org/>
- EAD (Encoded Archival Description) Official Website - <http://www.loc.gov/ead/>
- The XML Cover Pages - <http://www.oasis-open.org/cover/sgml-xml.html>

Robin Cover, SGML ve XML hakkında genel ve ayrıntılı bilgi vermektedir.