

МЕТА-ДАННИ – ХАРАКТЕРИСТИКИ

МИЛЕНА МИЛАНОВА

Използването на термина „мета-данни“ в съвременната теория и практика, свързана с регистрацията на информационни ресурси (традиционни и дигитални), налага неговото изясняване. Концепцията за мета-данните се появява и се развива с появата на електронните ресурси и необходимостта от тяхната регистрация. Всъщност мета-данните са познати на библиотечната общност много преди появата на електронните документи. Това са библиотечните каталози и правилата и стандартите за библиографско описание на документалните източници.

Съществуват не малък брой дефиниции на термина мета-данни. Той се определя и разбира по различен начин от множеството професионални организации, които проектират, създават, съхраняват или използват различни информационни системи и ресурси. Някои го използват като го отнасят към машинно четима информация, докато други го използват само за записи, които описват електронни ресурси. До средата на 90-те години на XX век терминът е използван преди всичко от общности, обменящи геофизическа информация. Те са разбирали мета-данните като съвкупност от всички данни, необходими за идентификация, представяне, обмен, техническото управление, изпълнение и употреба на данни, съдържащи се в дадената информационна система(1).

Думите на Arlene Taylor(2) обобщават причините за съществуването на множество дефиниции и понятия, включени под общия термин мета-данни: „Хората имат предвид различни неща, когато обясняват понятието „мета-данни.“ В основата си дискусиите се отнасят до някои комбинации или до някои от следните концептуални компоненти: информационния пакет и атрибутите му, стандарти за съдържание (правила за описание на пакета), схеми с мета-данни, елементи на мета-данни, записи с мета-данни и кодиращите формати. Като пример за илюстрация на това твърдение може да послужи MARC форматът. Защото някои разглеждат MARC само като кодиращ формат, други го обясняват като схема с мета-данни.“

Най-разпространеното и най-популярно определение за мета-данните е „данни за данните“ или „информация за информацията“. В тази дефиниция се вземат предвид всички ресурси във всякакъв формат, електронни или традиционни.

Друго направление в дефинирането на термина е неговото тясно използване за записи, които описват само електронни ресурси.

Някои от съществуващите дефиниции са:

- Структурирана информация, използвана, за намиране, достъп и управление на информационни ресурси, най-

вече в дигиталното пространство (International Encyclopedia of Information and Library Science, 2003)(3).

- Структурирани, кодирани данни, които описват характеристиките на информационното поведение на обекти в помощ на идентификацията, откриването, оценяването и управлението на описваните обекти (Майкъл Горман)(4).

- Структурирани данни, които описват ресурс, същност, обект или други данни. Те биха могли да се използват независимо от мястото, употребата, както и управлението на даден ресурс, същност и др.(5).

- Информация, която се отнася към една или друга част от информация, която може да съществува под отделни физически форми. Накратко, данни за данните. Всеки тип описание може да бъде разглеждано като мета-данни. Примери могат да бъдат библиотечната каталожна информация, кодираните заглавия на текстови файлове и данните за шофьорските свидетелства. В света на информационните технологии терминът често е използван, за да определи данни, които се отнасят към дигитални ресурси, достъпни в мрежата(6).

- Мета-данните описват разнообразни атрибути на информационните обекти и им придават значение, съдържание и структура. Теорията и практиката на описателните мета-данни е позната област за мнозина, като корените ѝ са в каталогизирането на печатни произведения. В дигиталната сфера допълнителни категории от мета-данни се появяват, за да поддържат движението и управлението на файловете(7).

В библиотечното пространство, мета-данните са използвани обикновено за

някои формални схеми за описание на ресурси, приложими към някои типове обекти, дигитални или недигитални. Традиционните библиотечни каталози са форма на мета-данни. MARC 21 и наборът от правила, използвани в него, подобно на AACR2, са стандарти за мета-данни. Други схеми за мета-данни са развити, за да опишат различни типове текстови или нетекстови обекти, които включват издадени книги, електронни ресурси, помощни средства за търсене на архивна информация, обекти на изкуството, образователни и обучителни материали и научни материали(8).

Важността на мета-данните се изразява в няколко направления. Наличието на богати, смислово логични мета-данни ще доведе до **повишаване на ефективността** при търсенето на информация и нарастване на достъпността. Те позволяват търсенето в колекции, съставени от много елементи, както и създаването на виртуални колекции от материали, които са разположени в множество отделни бази данни.

В дигиталния свят не е трудно за отделния обект от дадена колекция да бъде дигитализиран, след което да получи самостоятелност както от собствената каталогизираща информация, така и от връзките му към други обекти в същата колекция. Мета-данните имат критичната роля да запазват контекста на съществуване на дадена информация. Те документират връзките между обектите, идентифицират автентичността, структурната и процедурна интегрираност, а както и степента на завършеност на информационните обекти.

Увеличаване използването на диги-

талните информационни системи също се подкрепя от наличието на мета-данни. Добре структурираните мета-данни могат да улеснят почти безкрайното число от начини за търсене на информация и представяне на резултатите, като постигат преодоляване на физическите бариери, съществуващи пред различните потребители.

Мета-данните играят важна роля при **разграничаването на разнообразните версии**, създавани от различните дигитални форми на даден обект. Те трябва да осигуряват връзка към разнообразните версии и да представят еднаквото и различното във всяка една версия. Мета-данните трябва също така да дават възможност за различаване вариантите дигитализирани версии от постоянните копия на оригиналния или родствения обект. Пряко свързан с този въпрос е и въпросът за **легалността**. Мета-данните позволяват да се записват много нива на права и репродуктивна информация, която съществува за информационните обекти и техните разнообразни версии. Мета-данните също така документират легални или донорски изисквания, които са били наложени на обектите.

Друга първостепенна важност на мета-данните може да бъде отнесена към процеса на миграция на информационните обекти при разработване и използване на следващи генерации от компютърен софтуер и хардуер. Съществени са мета-данните, носещи информация за това как даден документ е бил създаден и поддържан, какво е поведението му и какви са връзките му с други информационни обекти. Тази информация трябва да бъде записана за информационните ресурси, за да останат

те **достъпни и понятни** с времето.

Не на последно място мета-данните имат своята важност при **подобряване на самата система и нейната икономика**. Техническите данни за benchmark (стандартни схеми за оценяване), много от които могат да бъдат събирани автоматично от компютър, са необходими, за да развиват и усъвършенстват системите, така че да ги направят по-ефективни и по-резултатни от техническа и икономическа гледна точка. Данните могат да бъдат използвани в планирането на нови системи(9).

Мета-данните се подчиняват на четири основни принципи, определени от Erik Duval(10):

√ Модулност (Modularity)

Този принцип се изразява при изграждането на схемите с мета-данни. Проектантите създават нови компоненти, базирани на съществуващите схеми с мета-данни, като се възползват от наблюденията върху най-добрите практики, вместо да измислят наново елементи. Елементите с данни на различни схеми, като речници и други изграждащи блокове, могат да се комбинират чрез синтактично или семантично коопериране на принципа на „лего“. Блоковете могат да се изградят и в по-големи структури. Важно понятие, свързано с този принцип, е понятието за **пространство на имената (namespace)**. То е фундаментална част от инфраструктурата на уеб. Най-просто казано, пространството на имената е формална колекция от термини, които управляват според определен алгоритъм. Основният протокол на уеб, http-пространството на имената, гарантира уникалността на адресирането в мрежата. Предметният рубризатор на

Библиотеката на Конгреса, LCSH (Library of Congress Subject Headings) – пространство на имената, управлява правилата за създаване и определяне на предметните рубрики(11).

Обявяването на пространството на имената позволява на дизайнерите на схеми с мета-данни да дефинират контекста на отделния термин, което осигурява неговата уникалност. Всяка схема с мета-данни получава уникален префикс, който определя нейната принадлежност. Така например можем да различим схемата на Dublin Core по префикса „dc“; IEEE-LOM (IEEE Learning Object Metadata) схемата с мета-данни ще я различим по префикса „lom“. В началото на RDF (Resource Description Framework) записи, декларацията на пространството на имената дава както името на набора от елементи, използвани, за да се опишат атрибутите на информационния ресурс, така и URL или интернет адреса, на който е представен официалния синтаксис. Пространството на имената, което представлява набор от кодирани полета, е понятие, което подчертава модулността на мета-данните и е наричано още: набор от елементи с мета-данни, схема с мета-данни, система от мета-данни или пространство на имената.

√ Разтегливост (Extensibility)

За да могат да бъдат използвани мета-данните според нуждите на тяхното приложение, те трябва да позволяват разтягане. Това означава те да могат да приспособяват идеята на основната схема, като добавят допълнителни елементи, които да бъдат нагодени за определена употреба според отделни или специфични нужди на дадена област, без да

се правят големи компромиси със съвместимостта, която е обезпечена от основната схема. Някои от елементите с мета-данни могат да бъдат открити във всяка схема с мета-данни, подобно на създателя на даден информационен ресурс или неговото идентифициране (URI/URL). От друга страна съществуват множество полета, които са специфични за определени приложения или общности.

Така например, съществуват схеми с мета-данни, създадени и поддържани от професионални общности, които са базирани на Dublin Core елементите с мета-данни, при което специфични полета за съответния домейн са били добавени при изграждането на новата схема с мета-данни, за да разтеглят използваемостта на базовата схема (в случая Dublin Core Metadata Element Set) за нуждите на съответната общност. Естествена последица, резултат от принципа на разтегливостта е, че архитектурата на мета-данните изисква да бъде конструирана така, че лесно да може да съчетава идеята на основната схема с допълнителни елементи, така че да се приспособи за специфични локални или домейн-специфични нужди без сериозни компромиси със съвместимостта на базовата схема. С други думи, някой използвайки други приложения, който се натъква на тези разширения, да е способен да ги игнорира и дори да ги направи използвани за всеки общ елемент в тези две схеми (базовата и специфичната)(12).

√ Усъвършенстване (Refinement)

Усъвършенстването най-просто представя възможността да се създава много „по-фино“ изработено описание

на ресурси, ако потребителят желае това. Или това е възможността на схемите с мета-данни да позволяват на ползвателите си да избират нивото на детализираност, подходящо за даденото приложение. Концепцията за усъвършенстване включва две субконцепции. Първата е допълнение към качеството, което прави значението на елементите по-специфично (т.е. прецизиране на значението). *Илюстратор, автор, композитор* или *скулптор* са примери за различни типове термини от много по-общия термин *създател*. *Дата на създаване, дата на промяна и дата на приемане* имат много по-тясно значение от атрибута *дата*. Втората субконцепция на принципа на усъвършенстване включва определяне специфичността на отделна схема или стойността на даден елемент, които дефинират целия диапазон от стойности на елемента. Ако е възможно да установим, че стойността, зададена на отделен елемент от мета-данни, е била избрана от контролиран речник или е била конструирана акуратно според определен алгоритъм, това може да направи този елемент много по-използваем, особено при автоматизирани процеси. Ако използваме отново атрибута *дата*, кодирането на дати и време е пример за използването на кодиращ стандарт, за да се отстрани двусмислеността от представянето на стойностите в мета-данните. Поредицата 03/06/02 се интерпертира като Март 6, 2002 в Северна Америка и като Юни 3, 2002 в Европа и Австралия. При използването на кодиращ стандарт датата може да бъде кодирана по недвусмислен начин във формат гтгг-мм-дд (2002-03-06, например: MARC21, UNIMARC). Други

елементи, например предметните рубрики, могат да използват традиционните библиотечни контролни речници (например: LCSH) също така успешно, за да се разшири достъпът до електронните ресурси чрез предмета. Друг пример може да бъде десетичната класификационна схема на Дюи (Dewey Decimal Classification) като мултилингвистична класификационна схема, която дълго се използва в традиционната библиотечна среда, която също така добре може да се приложи и към електронни ресурси. Употребата на специфични кодиращите формати в схемите с мета-данни, позволява увеличаване на яснотата при машинните процеси, както и улеснява разбирането от страна на потребителите(13, 14).

√ **Многоезичност** (Multilingualism)

Последният от четирите принципа на мета-данните, е основен за запазване лингвистичното и културно разнообразие при използването на схемите с мета-данни. Есенцията на принципа се състои в това да пригоди архитектурата на мета-данните към съобразяване с езиковите и културните различия. Всъщност може да се каже, че този принцип представлява парадокс. Първите три принципа, спазени в схемите с мета-данни, подпомагат интернационализма и глобализацията при описанието на отделни ресурси. Стандартите гарантират постигане на основната цел на световната мрежа като информационна глобална среда, за което е необходимо да се използват инструменти, като контролирани речници и внимателно дефинирани пространства на имената. Уеб, като глобална информационна система, е важна с това, че позволява безпрецеден-

тен достъп до ресурси в глобален аспект. Ако подобни ресурси не са направени достъпни за потребителите на техните родни езици, с подходящи набори от символи, и мета-данни, подходящи за управлението на тези ресурси, уеб ще провали потенциала си за глобална информационна система. Едновременно с това, нуждата от локализация изисква адаптирането на същите тези стандарти към локален контекст. Основна отправна точка в представянето на глобална архитектура на мета-данни е преводът на спецификации и стандартни документи на различни езици. Десетичната класификационна схема на Дюи, например, е преведена на различни езици. DCMI (Dublin Core Metadata Initiative) поддържа списък от преводи на техните основни документи. Също така The European Workshop on Learning Technologies, поддържа преводи на LOM (Learning Object Metadata) спецификацията. По този начин даден контролиран речник може да бъде използван според локални нужди и да бъде прилаган от създателите на мета-данни, независимо от майчиния език.

Друго съществено направление е включването на описание на езикови и други културални аспекти на ресурсите в схемите с мета-данни, откъдето следва тяхната способност да идентифицират алтернативни версии на ресурса на различни езици така добре, както и самия оригинал.

На едно по-техническо ниво е важно за глобалното адаптиране на стандартите двете специфични направления (глобализация и локализация) и начинът на кодирането на тези направления да са доколкото е възможно по-неутрални

от средата. Например, не би било уместно да дефинираме стойност на елемент с данни като образователен контекст в случаите, когато тя е специфична само за една национална система. Също така, кодирането е най-добре да се базира на цифрово представяне на елементите или техните стойности, въпреки че съществува широка практика да се употребяват някои форми на „псевдо-английски“ (HTML- етикетите са типичен пример: етикет се отнася към понятието „List Item“ (списък на единиците).

Многоезичността е един от аспектите от сродния въпрос за мултикултурализма, който се наблюдава при представянето на данни в различни страни. Другоезичен, мултикултурален въпрос е този за „дясно-ляво“, посоката, в която текстът се представя на екрана и чете. Английският и европейските езици спазват посоката от ляво на дясно, а арабският и иврит от дясно на ляво. Тук се нареждат и въпросите за начина на представяне на датите в различните календари; културните значения по подразбиране на определени икони, пиктограми; стандарти от практиката (например: подредбата на имената). Въпреки, че тези въпроси са извън специфичния контекст на мета-данните, те са също част от нуждата им да описват релевантни характеристики на ресурсите по начин, който е във връзка и зачита езиковите и културни различия(15, 16).

Най-основно може да се каже, че съществуват три типа мета-данни:

✓ **Дескриптивни** мета-данни - описват ресурси с цел откриване и идентификация. Може да включват елементи, подобни на заглавие, анотация, автор и ключови думи.

√ **Структурни** мета-данни - посочват как съставните обекти да бъдат съчетани заедно. Например, как страниците са подредени при оформяне на част от даден документ.

√ **Административни** мета-данни - предоставят информация, която подпомага управлението на ресурси, подобно на това кога и как е бил създаден даден ресурс, типа на файла и друга техническа информация и кой може да има достъп до него. Съществуват няколко подмножества на административните мета-данни; два, които понякога са посочвани като отделни типове мета-данни са:

- мета-данни за управление на права, които са част от притежаващите интелектуални права;

- мета-данни за консервация, които съдържат информация необходима за архивите и опазване на ресурсите(17).

На таблица 1 се виждат видовете мета-данни с техните цели и примерни елементи, както и употреба. Таблицата е според он-лайн ръководство от интернет портала на Cornell University Library.

Същото определение на видовете мета-данни е възприето от Priscilla Caplan(18).

Както бе споменато, съществуват и други видове мета-данни, които някои изследователи посочват като основни видове. Според Anne J. Gilliland(20) мета-данните са административни, дескриптивни, за консервация, технически и за употреба. Мета-данните за консервация са свързани с управлението и опазването на информационните ресурси. Те документират физическото състояние на ресурсите или действията, предприети за опазване на физическата и дигитал-

ната версия на ресурса, например: опресняване на данни и миграция. Техническите мета-данни са свързани с това как функционира системата или с поведението на мета-данните. Такива са: документацията за хардуера и софтуера; информация за дигитализацията, например: формати, коефициент на компресиране, скалираща програма; следене на времето за отговор на системата; данни за автентичност и сигурност, например: кодиране, пароли. Мета-данните за употреба са свързани с нивото и вида на използване на информационните ресурси: показвани записи, следене на използването и на потребителите, съдържателна повторна употреба и информация за различните версии.

Една ранна класификация (от 1998 г.) предлага четири типа мета-данни: мета-данни, свързани с откриването на ресурсите; мета-данни за наличност на ресурса; мета-данни за използването на ресурса; мета-данни за администрация и контрол на ресурса(21).

Този преглед на определенията, принципите и характеристиките на мета-данните показва, че за съвременния изследовател това понятие може би не е „панацея“ за решаването на проблемите, свързани с каталогизирането на електронни ресурси, но е достатъчно важно, за да е необходимо неговото проучване и добро познаване. В съвременната теория и практика концепцията за мета-данните е основен инструмент, който позволява постигането на съвместимост между информационните системи на различни общности, като използва добре познати на библиотечните професионални среди механизми.

Таблица 1. Видове мета-данни(19)

Тип	Цели	Примерни елементи	Примерна употреба
Дескриптивни мета-данни	<p>Описват и идентифицират информационните ресурси;</p> <ul style="list-style-type: none"> • на локално ниво (система) – позволяват търсене (например: търсене на колекции с изображения, с цел да се намерят художници или животни); • на ниво уеб – позволяват на потребителите да откриват ресурси (например: търсене в уеб за намиране на дигитализирани колекции от поезия). 	<ul style="list-style-type: none"> • уникални идентификатори (PURL, Handle); • физически атрибути (медия, условни размери); • библиографски атрибути (заглавие, автор/създател, език, ключови думи). 	<p>Handle PURL (Persistent Uniform Resource Locator) Dublin Core MARC HTML Meta Tags</p> <p>Контролирани речници подобни на: Art and Architecture Thesaurus, Categories for the Description of Works of Art</p>
Структурни мета-данни	<p>Подпомагат навигацията и представянето на електронните ресурси:</p> <ul style="list-style-type: none"> • осигуряват информация за вътрешната структура на ресурсите, включвайки страници, раздели, номер на глави, индекси и съдържания; • описват връзки между материалите (например: фотография В се съдържа в ръкопис А); • свързва сродни файлове и знаци (например: файл А с JPEG формат на архивното изображение файл В). 	<p>Структурира етикети, подобни на заглавна страница, съдържание, глави, части, печатни грешки, индекси, връзки между суб-обекти (например: фотография от дневник).</p>	<p>SGML XML Encoded Archival Description (EAD) MOA2, Structural Metadata Elements Electronic Binding (Ebind)</p>
Административни мета-данни	<p>Подпомагат както срочното, така и дългосрочното управление и развитие на дигиталните колекции:</p> <ul style="list-style-type: none"> • включват технически данни за създаване и качествен контрол; • включват управление на правата и контрол на достъпа, както и условията на ползване; • запазват информация за действия. 	<p>Технически данни, подобно на типа на на скенера и модсда, резолюция, големина на битоветс, цетовоа схема, формат на файла, компресия, източник на светлина, собственик, дата на печатни права, лимит за копиране и разпространение, лицензна информация, действия за консервация (цикли на опресняванс, миграция и др.)</p>	<p>MOA2, Administrative Metadata Elements National Library of Australia, Preservation Metadata for Digital Collections CEDARS</p>

Бележки

1. Gill, T., Gilliland, A. J., Woodley, M. S. Introduction to Metadata. Pathways to Digital Information. Online Edition, Version 2.1, p. 1, http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html, достъпен на 06.02.2007.
2. Цит. по: **Intner, S., Lazinger, S., Weihs, J.** Metadata and its impact on libraries. London, 2006, p. 5.
3. Пак там.
4. Цит. по: **Tedd, L., Large, A.** Digital Libraries. Muenchen, 2005, p. 89.
5. Metadaten, <http://www.ddb.de/standardisierung/metadaten/metadaten.htm>, достъпен на 13.03.2007.
6. Structural Metadata Dictionary for LC Digitized Material, Version 1.03, <http://memory.loc.gov/ammem/techdocs/repository/gengloss.html#metadata>, достъпен на 07.02.07.
7. Moving Theory into Practice: Digital Imaging for Libraries and Archives. 5. Metadata, <http://www.library.cornell.edu/preservation/tutorial/preface.html>, достъпен на 07.02.2007.
8. Understanding Metadata. Bethesda, 2004, p. 1, <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>, достъпен на 11.01.2007.
9. Gill, T., Gilliland, A. J., Woodley, M. S. Introduction to Metadata. Pathways to Digital Information. Online Edition, Version 2.1, p. 8-9, http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html, достъпен на 06.02.2007.
10. **Duval, E., Hodgins, W., Sutton, S., Weibel, S.** Metadata Principles and Practicalities. // D-Lib Magazine, Vol. 8, 2002, N 4, <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/april02/weibel/04weibel.html>, достъпен на 28.01.2007.
11. Пак там.
12. **Intner, S., Lazinger, S., Weihs, J.** Metadata and its impact on libraries. London, 2006, p. 7-8.
13. Пак там, с. 8.
14. **Duval, E., Hodgins, W., Sutton, S., Weibel, S.** Metadata Principles and Practicalities. // D-Lib Magazine, Vol. 8, 2002, N 4, <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/april02/weibel/04weibel.html>, достъпен на 28.01.2007.
15. **Intner, S., Lazinger, S., Weihs, J.** Metadata and its impact on libraries. London, 2006, p. 8.
16. **Duval, E., Hodgins, W., Sutton, S., Weibel, S.** Metadata Principles and Practicalities. // D-Lib Magazine, Vol. 8, 2002, N 4, <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/april02/weibel/04weibel.html>, достъпен на 28.01.2007.
17. Understanding Metadata. Bethesda, 2004, p. 1, <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>, достъпен на 11.01.2007.
18. Цит. по: **Nero, L.** Cataloguing digital resources. // Library Review, Vol. 54, 2005, N 2, p. 104.
19. Moving Theory into Practice: Digital Imaging for Libraries and Archives. Metadata Types, <http://www.library.cornell.edu/preservation/tutorial/preface.html>, достъпен на 07.02.2007.
20. Gill, T., Gilliland, A. J., Woodley, M. S. Introduction to Metadata. Pathways to Digital Information. Online Edition, Version 2.1, p. 4, http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html, достъпен на 06.02.2007.
21. Цит. по: **Intner, S., Lazinger, S., Weihs, J.** Metadata and its impact on libraries. London, 2006, p. 11-12.