

Радостина Тодорова

Увод

Динамичното развитие на съвременните информационни технологии оказва влияние върху сектори, свързани със съхранение и осигуряване на достъп до данни и информация. Университетските библиотеки не могат да останат извън тези процеси, поради спецификата на аудиторията, която обслужват, и необходимостта от развитие на различни информационни продукти и услуги. В практиката се очертава тенденция за използване на различни софтуерни приложения в „облака“ за изпълнение на ключови библиотечни функции, като на първо място е осигуряването на открит достъп до научна информация. С навлизането на „облачните“ технологии, които могат да са отделни софтуерни разработки, платформи или информационни инфраструктури, възникват редица въпроси, свързани със сигурността на библиографските данни и неприкосновеността на личните данни, с които библиотеките неизбежно боравят.

Не на последно място трябва да се отбележат и проблемите, свързани с качеството на метаданните, които в редица случаи се въвеждат от крайните потребители на тези системи – автори, преподаватели и докторанти, учени и изследователи в академичните институции. Данните и метаданните имат важна функция в общество, основано на знанието. Различните сфери на стопанския живот, управлението и науката зависят от тях, поради което и проблемът за тяхното качество може да се окаже фундаментален. Една от важните предпоставки за създаване на организационна среда, подкрепяща и насърчаваща споделянето и повторното използване на данни и метаданни, е развитието на култура за споделяне на знание у представителите на академичната общност.

Хранилища на данни в „облака“

„Облачните технологии“ са част от естествената еволюция на съвременната виртуализация, ориентирана към услугите. До голяма степен навлизането на облачните приложения в ежедневието на организациите и хората се дължи на възможността те да се абстрахират от дейности като експертно управление и контрол на технологичната инфраструктура. Това е нова и перспективна парадигма за доставяне на IT и компютърни услуги. Облачната технология е модел, който прави възможен мрежовия достъп до споделени ресурси като интернет мрежи, сървъри, хранилища за масиви от данни и софтуерни приложения с минимално участие или управление от доставчика на услугата. С навлизането на този технологичен бизнес модел потребителите могат да използват различни приложения чрез леки преносими устройства, като мобилни телефони, лаптопи и PDA устройства.

Моделите за обслужване чрез облак са няколко. От най-голямо значение за библиотеките е моделът „софтуер като услуга“ (Software-as-a-Service – SaaS) – доставка на софтуер като услуга през интернет, което премахва необходимостта от инсталиране и стартиране на дадено приложение на собствени компютри на организацията. Именно на този принцип се изграждат системите за институционални научни архиви. При тези системи контролът и управлението на основната физическа и виртуална инфраструктура на облака, включително мрежи, сървъри, операционни системи и др. се осъществяват от доставчика на облачния софтуер, който може да е самата университетска библиотека, или IT звено в университета. Данните и метаданните

се предоставят от самите потребители – преподаватели и докторанти, учени и изследователи, които депозират в системата своите произведения.

Предимствата на системи от модела SaaS, с които се характеризират и изградените чрез тях хранилища на данни, са: лесно администриране, автоматично обновяване и управление, универсална съвместимост, т. е. всички потребители ще имат една и съща версия на софтуера, лесно сътрудничество между различните нива потребители (автор, библиографски редактор и администратор), което предполага и глобална достъпност.

Например EPrints, първият и най-популярен софтуер за изграждане на институционални хранилища, освен към специалистите администратори е значително ориентиран и към потребителите. Разработчиците се стремят постоянно да премахват бариерите пред депозиращите научно съдържание, но и да създават механизми за подобряване качеството на метаданните и като цяло гарантиране информационната стойност на дигиталния архив, посредством намаляване времето за осъществяване на депозита, възможност за импорт на данни от други хранилища или чрез други услуги, автоматично попълване на полета за по-бързо въвеждане на данни и контрол върху имената (name authorities). Системата се придържа изцяло към принципа за еднократно въвеждане и многократно използване на данните. Тя може да работи съвместно с други системи от типа „библиографски мениджъри“, различни десктоп и Web 2.0 приложения, RSS емисии и имейл известяване. Потребителите лесно могат да съставят библиографии по различни критерии и в различни формати, да експортират данни, да генерират собствени RSS емисии и да ги интегрират в други уеб приложения.

Качество на метаданните

Проблемите, свързани с качеството на метаданни в мрежова среда, са особено актуални след широкото навлизането на протокола OAI-PMH, който претърсва и извлича големи количества метаданни от оперативно съвместими хранилища на ресурси в областта на науката, при което е налице непосредствена връзка между качеството и съответствието с практики на библиографското описание. В специализираната литература авторите се обединяват около мнението, че проблемите с качеството възникват именно в този контекст, тъй като все по-сложни ресурси се описват от необучени хора, вместо от добре обучени библиотекари или чрез автоматизирани средства. Проблемите стават още по-сериозни, когато метаданните от едно хранилище или архив станат част от по-широка система, която има за цел оперативна съвместимост.

Повечето автори констатират, че в отделното хранилище първоначално качеството на метаданните може да е високо, но в една обобщена среда, същите метаданни могат да имат ниско качество. Това се случва, защото параметрите за качество, които са валидни в оригиналния контекст на даден архив, могат да бъдат различни от параметрите, приети в агрегираната среда. В условията на непрекъснато увеличаване, от една страна на броя на цифровите хранилища, на техните обеми и не на последно място – на разнообразието на съхранявани информационни източници и формати, а от друга страна, на развитието на все по-нови системи за обединено търсене и извличане на информация, проблемът за осигуряване на общо и достатъчно равнище на качество става основен за развитието на научните архиви.

C. Yasser анализира и сравнява проблемите с качеството на метаданните, описани в литературата, като ги класифицира в пет категории:

- неправилни стойности – записите с метаданни съдържат стойности, които не представляват правилно даден ресурс, макар отделните елементи да са приложени правилно;
- неправилни елементи – стойностите са подходящи за описване на ресурса, но са предназначени за грешен елемент;
- липсваща информация – записът с метаданни не е пълнен;

- загуба на информация – някои параметри, характеризиращи информацията са изгубени (по време на процеса на мигриране на метаданни от един формат в друг);
- противоречиви стойности – различни стойности, които по отделно правилно характеризират, но събрани заедно могат да компрометират функционалността на системата.

Качеството на данните и метаданните засяга ефективното използване на информационните инфраструктури и на отделните ресурси, каквито са институционалните научни архиви, които играят ключова роля в обществото на знанието. Оценката и предприемането на целенасочени действия за осигуряване качеството на метаданните във всеки отделен научен електронен архив е една от основните задачи, от които зависи и неговото развитие.

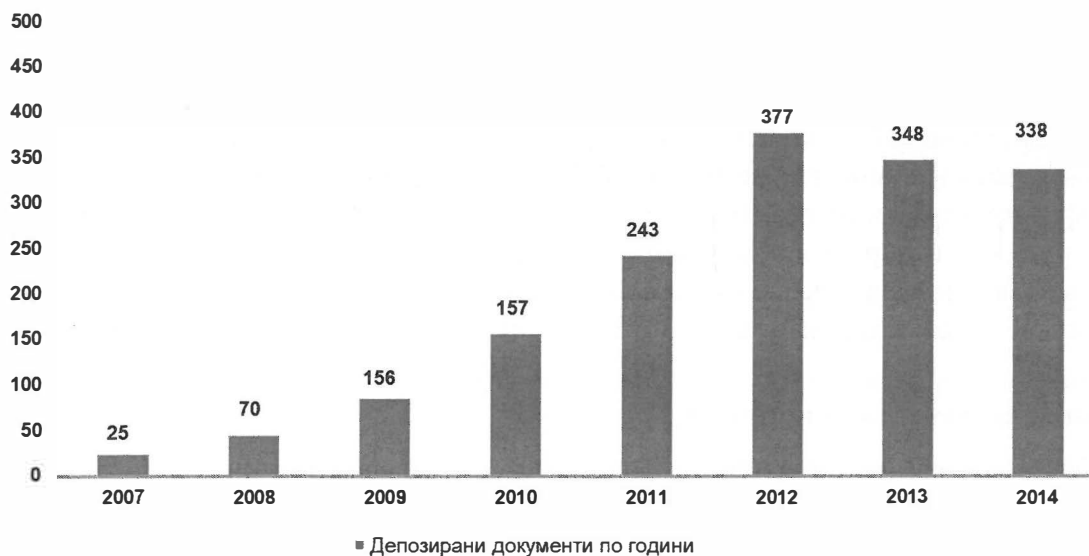
Опит за качествена оценка

Научният електронен архив на Нов български университет е първият архив с открит достъп до научна информация, създаден в България. Мисията му е да бъде единно, отворено електронно пространство, което осигурява дълготрайно съхранение и достъп до научните постижения на академичната общност в университета. Освен това чрез него се провежда институционална политика, насърчаваща свободния обмен на знания. Депозитът на документите и попълването на метаданните се осъществява от самите автори – преподаватели (щатни и хонорувани) и докторанти, като в случай на съавторство е необходимо поне един от авторите да бъде свързан с университета.

Принадлежащ към т. нар. „Зелен път“ и институционален архив по своята същност, той е уникален с това, че депозирването на документи в него не е задължително. Учените от университета депозират произведения освен на български и на различни други езици. Друго предизвикателство е многообразието от документи, плод на научната или творческа дейност, които могат да бъдат депозирани, като: статия, публикувана в списание, сборник или вестник; част, глава от книга; монография; материал от конференция, конгрес (презентация, доклад или постер); цяла книга; докторска дисертация или хабилитационен труд; патент; артефакт; шоу, изложба; музикална композиция; изпълнение (музикално); изображение (фотография); видео; дигитален звукозапис; числови данни; експериментални данни и анализи; учебни ресурси: лекции, упражнения, тестове и др.

Тези условия поставят пред екипа, който администрира архива, предизвикателството ежедневно да обучава, убеждава, съдейства и насърчава университетската общност да се възползва от предимствата на открития достъп до информация. Но една от главните задачи е осигуряването на контрол върху метаданните, което е от изключително значение за потребителите на информация в интернет. От създаването си през 2005 г., вече 10 години екипът неформално анализира практиката за самостоятелно депозиране на документи, които вече наброяват над 1700 (виж фиг. 1).

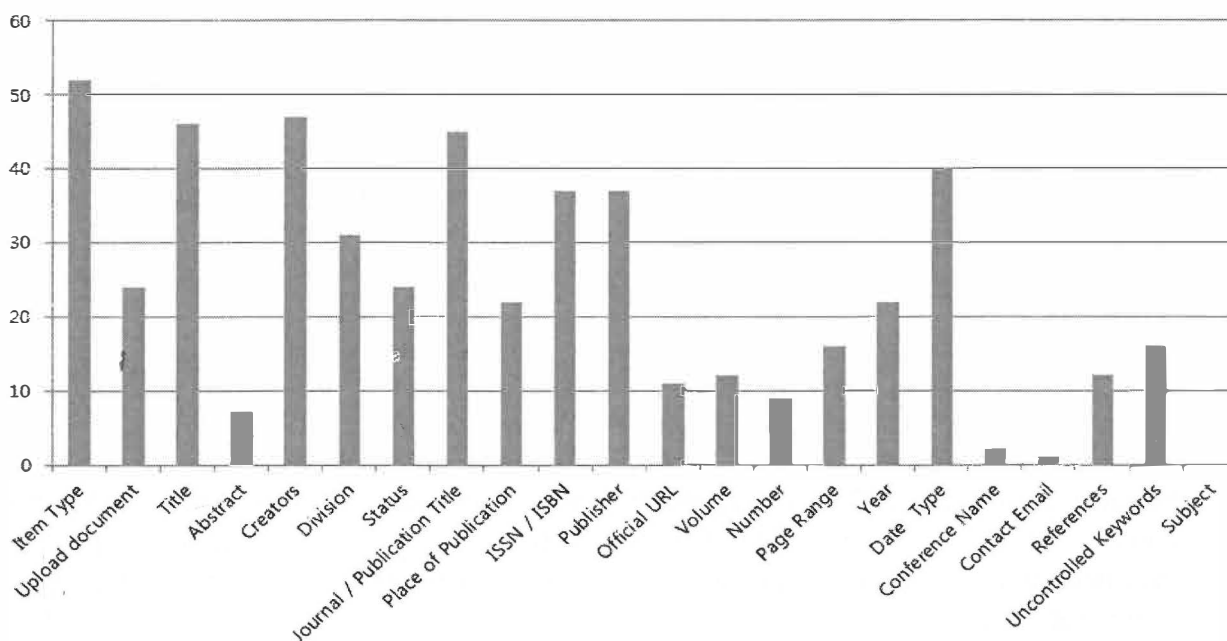
С цел повишаване качеството на предоставяните метаданни и подобряване на обучението се проведе шестмесечен мониторинг на предлаганите от авторите записи през периода от м. ноември 2013 г. до април 2014 г. През периода са депозирани общо 155 документа от 34 автори, преподаватели и докторанти в университета, които са обект на изследването. Проследени са попълнените метаданни в 22 библиографски полета. Използван е изключително опростен модел за определяне на качеството на метаданните за разлика от споменатия по-горе модел на Yasser. Като некачествени метаданни са отчетени тези, в които библиографският редактор на архива е направил промяна.



Фиг. 1. Депозирани документи по години

Резултати от изследването

Анализът на събраните данни очертава типичните грешки, които депозиращите автори допускат при метаданните и причините, които ги обуславят. Установени са общо 513 грешно попълнени метаданни или средно 3,3 грешки във всеки запис. Топ 5 от грешките в цялата изследвана съвкупност са в полетата: Item Type, Creators, Title, Journal/Publication Title и Date Type (виж фиг. 2). На първо място прави впечатление проблемът с определяне типа на документа, създателите му, като най-често се обърква поредността на авторите. Проблемът със заглавията произлиза от интерпретирането им, от разликата във формулирането им при подаване за печат и впоследствие редактиране на заглавието от издателя. Авторите интерпретират и заглавието на източника, в който е бил публикуван техният материал, както и годината на публикуване.

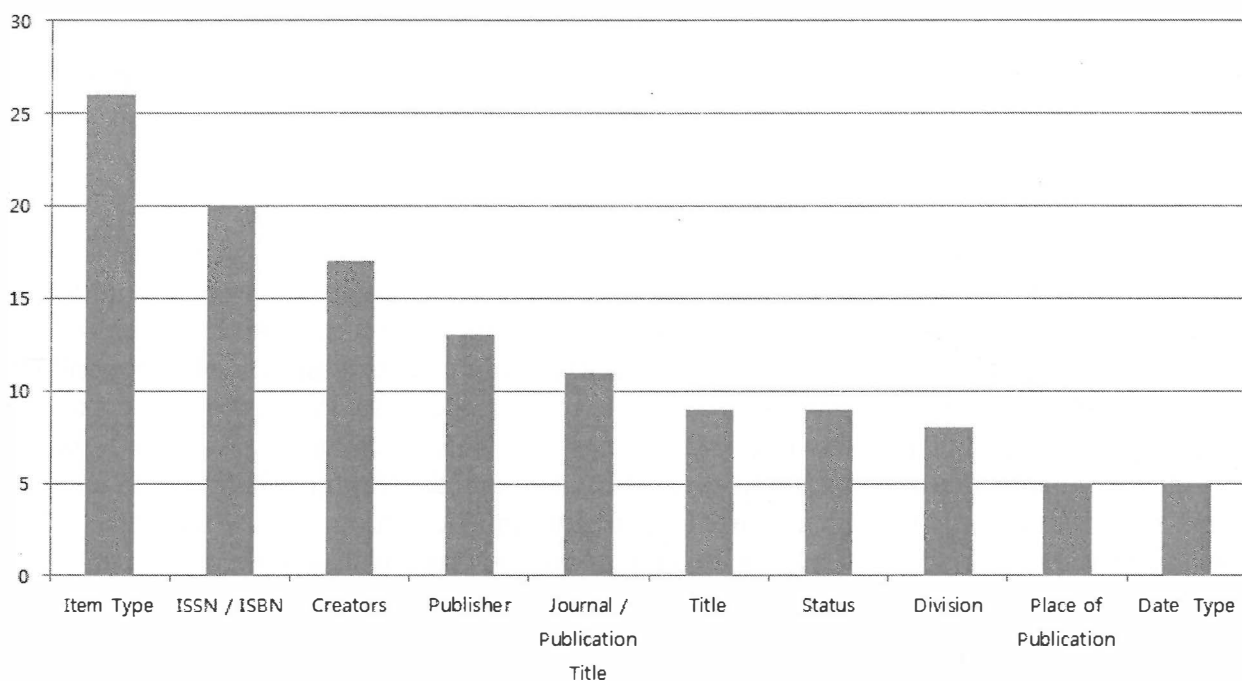


Фиг. 2. Брой некачествени метаданни по полета

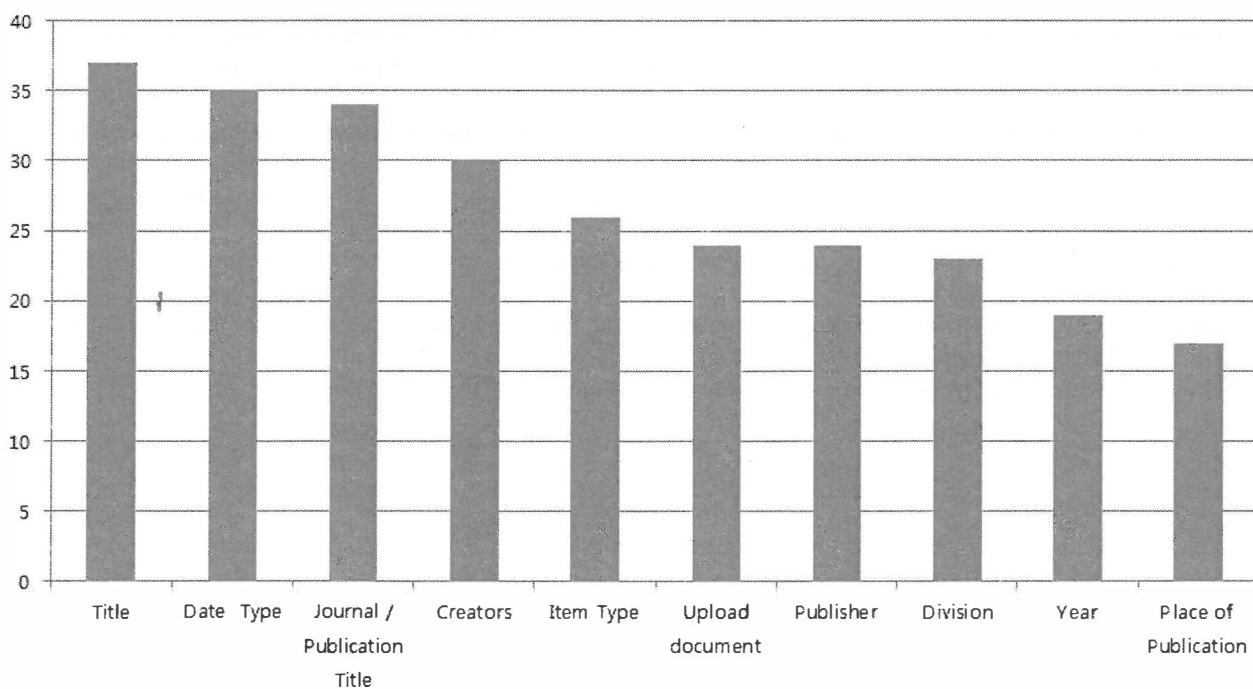
Прави впечатление липсата на проблем с метаданните в полето Subjects. Тук депозиращите имат възможност да избират от предварително дефинирана онтология от предметни категории, чрез които да разкрият съдържанието на своите документи. За авторите не е проблем да

предоставят имейл за контакт, а незначителен процент попълват неправилно полето с наименованието на конференция, в която са участвали. Изследването отчита честотата на полагане на допълнителни усилия от авторите за създаване на допълнително резюме (несъществуващо в основния документ) на език различен от езика на документи, като прави впечатление, че много малък процент от депозиращите не го правят.

Очаквано има разлика при качеството на метаданните, предоставяни от авторите, след разделянето им в две групи – на участвали в обученията за развитие на информационната компетентност, по отношение на открития достъп до научна информация и такива, които не са участвали в подобни обучения. Идентифицирани са типичните грешки и при двете групи (виж фиг. 3 и фиг. 4).



Фиг. 3. Брой некачествени метаданни по полета, допуснати от автори, участвали в обученията



Фиг. 4. Брой некачествени метаданни по полета, допуснати от автори, които не са участвали в обученията

Първата група автори, участвали в обучителни семинари са 18, а депозираните от тях материали са 93. Установени са 140 некачествени метаданни или средно 1,5 грешки в запис. Топ 5 грешки допускани в групата на преминалите през обучение автори са в полетата: Item Type, ISSN/ISBN, Creators, Publisher и Journal/Publication Title. Втората група се състои от 16 автори, които не са посещавали обучителни семинари, като депозираните от тях материали са 62. Установени са 373 некачествени метаданни или средно 6,0 грешки в запис. Топ 5 грешки допускани в групата на непреминалите през обучение автори са в полетата: Title, Date Type, Journal/Publication Title, Creators и Item Type.

Сравнителният анализ на резултатите на двете групи разграничава поведението и навичките на авторите, посещавали обучения от тези, които не са посещавали такива. Ясно се очертава зависимостта групата на необучените да допуска три пъти повече грешки спрямо групата на участвалите в обучения. Най-важната констатация обаче е, че обучените автори депозират по-често. В разглеждания период всеки от тях е архивирал средно 5,2 документа, за разлика от необучените – само 3,9.

Заклучение

Проведеното изследване и анализът на резултатите допринасят за утвърждаване на доброволния депозит като част от университетската култура за свободно споделяне на научно знание. Установената зависимост, че преминалите обучение автори депозират по-често показва, че именно това е пътят, по който преподавателите и докторантите могат да бъдат приобщени към културата за свободно споделяне на знания в университета. Резултатите от пилотното изследване ще подпомогнат обновяването на съществуващата обучителната програма чрез поставяне на нови акценти, свързани с качеството на метаданните, придобиване на знания и умения за успешно попълване на данните в проблематичните полета. От друга страна, опитът за оценка на първия електронен научен архив у нас предоставя изпитани и полезни практики, които могат да бъдат използвани от другите развиващи се архиви в България, а също и от университетски библиотеки в други страни, развиващи своите научни архиви в подобна среда.

Използвана литература:

1. Тодорова, Р. Университетската библиотека в научната комуникация: нови предизвикателства. // Библиотеки в трансформация: проблеми, предизвикателства, нови възможности: Сборник доклади от XXIV Национална конференция, на ББИА, София, 5 – 6 юни 2014 г. София, 2014, с. 28 – 31.
2. Тодорова, Р. Иновативни подходи за развитие на информационната компетентност на преподаватели и докторанти в Нов български университет. // Информационна грамотност – модели за обучение и практики: Сборник с науч. доклади и съобщения от Научен семинар с международно участие, 18 – 19 окт. 2012 г. София, 2012, с. 344 – 350.
3. Shin, D. Beyond user experience of cloud service: Implication for value sensitive approach. // Telematics And Informatics, 2015, 32, pp. 33 – 44.
4. Tani, A., Candela, L., Castelli, D. Dealing with metadata quality: The legacy of digital library efforts. // Information Processing And Management, 2013, 49, pp. 1194 – 1205.
5. Yasser, C. An analysis of problems in metadata records. // Journal of Library Metadata, 2011, 11 (2), pp. 51 – 62.