

UNIVERZA V LJUBLJANI

FILOZOFSKA FAKULTETA

ODDELEK ZA BIBLIOTEKARSTVO, INFORMACIJSKO ZNANOST IN KNJIGARSTVO

IMPLEMENTACIJA METAPODATKOV NA PORTALU SISTORY

Diplomski seminar

Mentor: dr. Jan Pisanski

Katja Meden

Ljubljana, avgust 2017

Ključna dokumentacijska informacija

Ime in PRIIMEK: Katja MEDEN

Naslov diplomskega seminarja: Implementacija metapodatkov na portalu Sistory

Študijski program: Bibliotekarstvo in informatika

Smer: Informacijska znanost

Kraj: Ljubljana

Leto: 2017

Število listov: _____38_____ Št. slik: _____3_____ Št. tabel: _____4_____

Število prilog: _____5_____ Št. strani prilog: _____1_____

Število bibliografskih opomb: _____1_____

Mentor: dr. Jan Pisanski

Delovna praksa je potekala na: Inštitutu za novejšo zgodovino

Delovni mentor/mentorica: Igor Zemljič, uni. dipl. zgodovinar in bibliotekar

UDK: 004.774:001.103.2

Ključne besede: metapodatki, metapodatkovni standardi, aplikacijski profili, Dublin Core Metadata Element Set

Izveček: Z razvojem Interneta, ki prinaša čedalje več informacij, metapodatki in metapodatkovni standardi postajajo čedalje bolj pomembni koncepti v našem vsakdanu. Predstavljena in opisana sta aplikacijski profil ter prehod med aplikacijskim profilom SISTORY in DCTERMS, ki poda podlago za analizo implementacije metapodatkov na portalu. Gledano v celoti sta načina implementacije metapodatkov na portalu SISTORY koristna oziroma učinkovita, aplikacijski profil in prehod opravljata svoje naloge, čeprav se v procesu oblikovanja slednji zasuka iz poudarka uporabnika in poenostavljanja elementov na poudarek strojne berljivosti in bogatejšega opisovanja informacijskih objektov.

Title: Metadata Implementation on the SISTORY portal

Keywords: metadata, metadata standards, application profiles, Dublin Core Metadata Element Set

Abstract: In the age where Internet provides so many information and where most of them are only a simple click or a touch away, metadata and metadata standards are becoming more and more important concepts in an everyday use. This document analyses a structural design of SISTORY application profile and metadata crosswalk between SISTORY application profile and DCTERMS, which lays the basis for the analysis of metadata implementation on the portal SISTORY.si. In the end, we can conclude that, as a whole, metadata implementation has been useful and effective. Both, the application profile and crosswalk are fulfilling their intentions and are useful to the portal, even though the emphasis has shifted during the process from user orientation and metadata elements simplification to machine readability and richer metadata descriptions of information objects.

KAZALO

ZAHVALA.....	5
UVOD	6
1. PREGLED LITERATURE.....	7
1.1 APLIKACIJSKI PROFIL.....	7
1.2 PREHODI MED METAPODATKOVNIMI SHEMAMI.....	8
1.1.1 Absolutni in relativni prehodi	9
1.1.2 Proces prehoda	10
1.3 SISTORY.SI.....	12
1.3 APLIKACIJSKI PROFIL SISTORY	13
2. PRAKTIČNI DEL	17
2.1 OBLIKOVANJE PREHODA MED AP SISTORY IN DCTERMS.....	17
2.1.1 Harmonizacija	19
2.1.2 Semantična preslikava.....	21
3. PRIMER IMPLEMENTACIJE METAPODATKOV	23
3.1 IMPLEMENTACIJA NA NIVOJU XML-SHEME.....	24
3.2 IMPLEMENTACIJA NA NIVOJU SISTEMA	26
3.3 IMPLEMENTACIJA NA NIVOJU PORTALA.....	28
4. RAZPRAVA	32
5. ZAKLJUČEK	34
6. LITERATURA	35
7. PRILOGE.....	36

KAZALO TABEL:

Tabela 1: Primerjava absolutnega in relativnega prehoda	9
Tabela 2: Oblikovanje skupne sintakse prehoda	19
Tabela 3: Preslikave elementa Type	22
Tabela 4: Pregled implementacije AP in prehoda	31

KAZALO SLIK:

Slika 1: Struktura aplikacijskega profila HOPE	13
Slika 2: Prikaz elementov za slikovno gradivo v orodju za vstavljanje metapodatkov	26
Slika 3: Prikaz dveh elementov dcterms:type in njuna konteksta v istem zapisu	28

KAZALO DIAGRAMOV:

Diagram 1: Prikaz delovanja vseh treh nivojev v portalu Sistory	23
---	----

KAZALO PRILOG:

Priloga 1: Aplikacijski profil Sistory	36
Priloga 2: Semantična preslikava elementov	36
Priloga 3: Sistory Basic XML 2.0	36
Priloga 4: XML-zapis Anton Aškerc (MGML)	36
Priloga 5: Osnovni in vsi metapodatki Anton Aškerc (MGML).....	36
Priloga 6: RDFa-zapis Anton Aškerc (MGML).....	36
Priloga 7: Seznam kratic	37
Priloga 8: Izjava o avtorstvu.....	38
Priloga 9: Dovoljenje za objavo	39

ZAHVALA

Za strokovno pomoč, vodstvo in čas, ki mi ga je posvetil, se zahvaljujem mentorju dr. Janu Pisanskemu.

Za izvedbo analiznega dela diplomskega seminarja se zahvaljujem Inštitutu za novejšo zgodovino, še posebej dr. Andreju Pančurju in dr. Mojci Šorn za vso pomoč in nasvete pri izdelavi praktičnega dela.

Poleg tega se zahvaljujem še vsem, ki so kakorkoli pripomogli k izdelavi tega diplomskega seminarja.

UVOD

V današnjem svetu je z uporabo Interneta skoraj vsaka informacija na dosegu roke. Vsakdo, ki ima v lasti računalnik in dostop do svetovnega spleta, lahko ustvarja nove informacije in jih objavlja – kar ustvarja eksplozijo informacij. Tu pa se vmešajo metapodatki, ki so orodje organizacije pomembnih informacij. V večini so ti metapodatki vključeni v večji nabor elementov, imenovan metapodatkovni standard. Ti standardi nam skozi svojo sestavo omogočajo nadzor nad opisovanjem informacijskih objektov. Nadaljnji razvoj je prinesel različne možnosti prilagajanja in oblikovanja shem, med njimi tudi aplikacijske profile in prehode med metapodatkovnimi shemami, ki so ustvarjeni z namenom zadoščanja specifičnim potrebam opisovanja informacijskih objektov.

V informacijskem sistemu je potrebno zasnovati takšno metapodatkovno shemo, ki nam bo omogočala da kar najbolj podrobno opišemo informacijski vir in ga na portalu prikažemo na takšen način, da bo služil svojemu namenu – uporabi. Zato je namen diplomskega seminarja predstaviti in analizirati način implementacije metapodatkov »Zgodovina Slovenije – SIstory«, spletnega portala, ki omogoča dostop do informacijskih objektov s področja slovenskega zgodovinopisja. Pri tem sta v ospredje postavljena aplikacijski profil SIstory in prehod med metapodatkovnima shemama (med aplikacijskim profilom SIstory in DCTERMS). Poleg načina, na katerega sta vključena v prikaz metapodatkov na portalu, pa nas kot raziskovalno vprašanje zanima tudi, kako učinkovita oziroma smiselna je bila nadgradnja prikaza.

V prvem poglavju diplomskega seminarja je predstavljena teoretična osnova konceptov aplikacijskega profila in prehoda, kjer so izpostavljene njihove temeljne lastnosti in struktura. V drugem poglavju je v praktičnem delu seminarja opisan in analiziran proces oblikovanja prehoda. V tem delu se nahaja tudi tretje poglavje seminarja, ki na konkretnem praktičnem primeru zapisa informacijskega objekta posmrtna maske Antona Aškerca prikaže načine implementacije aplikacijskega profila in prehoda.

1. PREGLED LITERATURE

1.1 APLIKACIJSKI PROFIL

Koncept aplikacijskega profila (AP) je po Lei Zeng in Qin (2008) definiran kot skupek izbranih elementov iz dveh ali več različnih metapodatkovnih standardov, sestavljenih v novo, prilagojeno shemo. Tak koncept je večinsko uporabljen na določenem področju, v katerem vladajo specifične potrebe. Vzdrževalci sistema, portala ali drugih storitev lahko uporabijo koncept aplikacijskega profila in s tem oblikujejo nove ali prilagajajo obstoječe metapodatkovne sheme, ki tako kar najboljše zadovoljijo potrebe sistema.

V osnovi ločimo tri vrste aplikacijskih profilov:

- aplikacijski profili z elementi iz več shem – najpogostejša definicija koncepta aplikacijskega profila predstavlja shemo, v kateri so elementi iz različnih metapodatkovnih shem združeni v enotni nabor elementov in uporabljeni za aplikacijo na specifičnih področjih (lokalna aplikacija);
- aplikacijski profili, ki so osnovani na eni sami metapodatkovni shemi, vendar so njene lastnosti prilagojene za specifično uporabo oziroma aplikacijo;
- aplikacijski profili, ki uporabljajo lastne imenske prostore – takšni aplikacijski profili v svojem naboru metapodatkov uporabljajo lastne elemente. Takšne aplikacije so redke, večinoma se v takšnih primerih imenske prostore le bolj natančno definira (Vilar, 2015/2016).

Pri tem imajo vse vrste aplikacijskih profilov določene skupne lastnosti oziroma določene omejitve.

Prva izmed takih lastnosti so že prej omenjeni imenski prostori oziroma *namespaces*, formalne zbirke terminov, ki enolično identificirajo nabor imen elementov (Vilar, 2015/2016). Tako lahko med seboj ločimo elementa z enakimi imeni – predpona, ki stoji pred imenom elementa, označuje imenski prostor, iz katerega ta element izhaja. Slednja je hkrati okrajšava za celotni URI, s katerim je imenski prostor označen. Prej omenjena lastnost, ki se navezuje na imenske prostore znotraj aplikacijskih profilov, je omejitev, ki preprečuje vpeljevanje novih imenskih prostorov oziroma novih podatkovnih elementov znotraj aplikacijskih profilov. Torej so lahko uporabljeni le že predhodno obstoječi elementi in njihovi imenski prostori. Heery in Patel (2000) ugotavljata, da se v aplikacijskem profilu še vedno lahko

ustvarja nove imenske prostore, pod pogojem, da se novi imenski prostor in elemente deklarira in tudi nadalje vzdržuje.

Naslednja izmed lastnosti aplikacijskih profilov je njihova zmožnost, da natančneje določijo že obstoječe definicije elementov. Definicije elementov iz različnih shem lahko prilagodi oziroma posebej opredeli, po večini pa gre le za zožitev ali drugačne omejitve definicije. Aplikacijski profili lahko tudi določajo dovoljene sheme in omejujejo vrednosti elementa. Večinoma se nabor omeji z vključevanjem kontroliranih slovarjev, tezavrov in drugih orodij za omejitve semantike, prav tako pa se vrednosti elementov lahko omeji z določanjem shem za zapis vrednosti, kot so sheme za zapis jezikov ali datumov (Heery, Patel, 2000).

1.2 PREHODI MED METAPODATKOVNIMI SHEMAMI

»Prehod med metapodatkovnimi shemami oziroma tako imenovan *crosswalk* predstavlja prevajanje elementov, semantike in sintakse iz enega metapodatkovnega standarda v drugega« (Chan, Lei Zeng, 2006) oziroma je specifikacija za preslikavo enega metapodatkovnega standarda z drugim (St. Pierre, LaPlant, 1998).

Pri tem termin preslikava (*mapping*) izraža intelektualno aktivnost identificiranja semantično ekvivalentnih elementov v različnih metapodatkovnih shemah, medtem ko termin prehod med metapodatkovnima shemama (*crosswalk*) označuje dokumentacijo, ki je rezultat preslikave elementov, kjer so prikazani semantično ekvivalentni elementi iz različnih metapodatkovnih shem (Lei Zeng, Qin, 2008).

Oblikovanje prehodov med metapodatkovnimi shemami je eden izmed najbolj učinkovitih in najbolj uporabljenih načinov za zagotavljanje interoperabilnosti. Pri tem so elementi iz enega nabora metapodatkov povezani z drugim naborom metapodatkov. Čeprav se pri tem element iz enega standarda preslika v drug element, s katerim se po opisu, namenu in definiciji najbolje ujema, se poleg tega preslikajo še njegovi podatki oziroma vrednosti in njegova semantika – preslika se pomen teh elementov. Te postopke se včasih imenuje tudi semantično preslikavanje (*semantic mapping*) (Vilar, 2015/2016).

1.1.1 Absolutni in relativni prehodi

Pri oblikovanju prehodov med metapodatkovnimi standardi lahko uporabimo dva pristopa:

- 1.) absolutni prehod (*Absolute crosswalking*) – Ta pristop zagotavlja semantično ekvivalentnost elementov iz izvorne in ciljne metapodatkovne sheme. To zagotavlja tako, da pri preslikavi v prehod uvršča le elemente, ki so si semantično enakovredni, vendar ta pristop ne deluje dobro pri konverziji podatkov;
- 2.) relativni prehod (*Relative crosswalking*) – V tem pristopu pa semantična enakovrednost elementov ne igra tako velike vloge kot v absolutnem prehodu. S tem pristopom vsem elementom izvorne sheme preslikamo element ciljne sheme, pri tem pa ni važno, ali sta si elementa med seboj semantično enakovredna (Chan, Lei Zeng, 2006).

Tabela 1: Primerjava absolutnega in relativnega prehoda (oblikovano po Chan, Lei Zeng, 2006)

Izvorna shema/Ciljna shema	Absolutni prehod	Relativni prehod
VRA Core 3.0	Dublin Core	Dublin Core
Technique	----	Format
Location.Current Repository	----	Contributor Coverage

Tabela 1 prikazuje primerjavo oblikovanja prehoda med elementi iz VRA Core 3.0 in Dublin Core. Ker absolutni prehod zahteva semantično ekvivalentnost med elementoma izvorne in ciljne sheme, le-ta v tem primeru ne obstaja. Relativni prehod pa preslikavo vseeno omogoča, saj sta si elementa podobna, čeprav ne semantično ekvivalentna.

Absolutni prehod z natančnostjo na področju semantike prinaša večjo enakovrednost elementov v preslikavi, vendar s tem močno zmanjšuje število elementov, ki jih lahko preslikamo. Po drugi strani pa to rešuje relativni prehod, vendar s seboj prinaša problem, saj elementi niso dejansko semantično ekvivalentni. Tako je relativni prehod bolj uporaben za oblikovanje prehoda iz bolj kompleksne v preprostejšo shemo (Chan, Zeng, 2006).

1.1.2 Proces prehoda

Te lastnosti so pomembne podrobnosti v naslednjih štirih fazah oblikovanja metapodatkovnega prehoda – harmonizaciji, semantičnem prevajanju, pravilih in pretvorbi.

1. Harmonizacija. V tej fazi poteka usklajevanje izvorne in ciljne metapodatkovne sheme – treba je oblikovati enotno sintakso. Tu se usklajujejo lastnosti, organizacija elementov in terminologija, kar omogoča oblikovanje skupne sintakse metapodatkovnega prehoda. Poteka najpreprosteje v tabeli, pri tem pa je treba upoštevati posebnosti elementov – oblikovanje sintakse ni preprosto, saj se preslikave v večini primerov ne ujemajo popolnoma, možne so namreč različne preslikave:
 - »ena-v-ena« (*One-to-One mapping*) – elementa iz izvorne in ciljne sheme sta si semantično ekvivalentna;
 - »ena-v-mnogo« (*One-to-Many mapping*) – element izvorne sheme se lahko preslika v več elementov ciljne sheme;
 - »mnogo-v-ena« (*Many-to-One mapping*) – več elementov izvorne sheme se ujema in preslika v en sam element ciljne sheme;
 - »ena-v-nič« (*One-to-none mapping*) – element izvorne sheme ni ekvivalenten nobenemu izmed elementov ciljne sheme, se ne preslika.
2. Semantično preslikavanje. Po končani stopnji harmonizacije se začne vzpostavljati zasnovo za semantično preslikavanje, kjer se vizualno s pomočjo tabele vzpostavijo odnosi med posameznimi elementi, ker se element izvorne sheme preslika v element ciljne sheme, ki mu je semantično ekvivalenten.
3. Pravila. Naslednja stopnja je določevanje pravil oziroma lastnosti, ki jih bo prehodna shema določala – pri tem je treba določiti, kakšne temeljne lastnosti metapodatkov bo prehod imel, da se bo s tem čim bolj približal namenu, ki ga želi doseči, in pri tem ohranjal ohranjati konsistentnost z drugimi shemami. Avtorja LaPlant in St. Pierre (1988) sta orisala lastnosti, ki so sorodne večini metapodatkovnih standardov. Te lastnosti naj bi se med procesom oblikovanja prehoda izpostavile, generalizirale in prilagodile za uporabo v procesu prehoda, pri čemer bi se ta proces poenostavil.

Primeri takih lastnosti:

- semantične definicije (pomen elementov – definicije);
- univerzalna identifikacija elementa (ID, label);
- število ponavljanj in obveznost (določi se, ali je element obvezen ali neobvezen na podlagi dovoljenega števila ponavljanj);
- odnosi – treba je opredeliti in izpostaviti odnose med elementi;
- zapis vrednosti elementa – prosto besedilo, numerični zapis ali zapis, omejen s kontroliranimi slovarji.

4. Pretvorba. Zadnja faza pa je pretvorba oziroma realizacija zasnovanega prehoda – izvaja se na podlagi pravil in odločitev iz prejšnjih faz, rezultat prehoda pa je metapodatkovni opis (Vilar, 2015/16).

Oblikovanje metapodatkovnih prehodov je kompleksen proces, pri katerem se pojavljajo problemi. V večini primerov metapodatkovne sheme po končanem procesu niso popolnoma ekvivalentne, prav tako pa se lahko težave pojavijo znotraj procesa preslikavanja elementov. Med procesom preslikavanja se zato pogosto pojavljajo različni problemi. Taki so na primer problemi preslikavanj na načine »ena-v-mnogo«, »mnogo-v-ena«, »ena-v-nič«, kjer se uporablja relativni prehod, kar pomeni, da se semantična ekvivalentnost elementov in posledično sheme znotraj prehoda izgubi. Ravno tako je pogost problem ponavljanja elementov – ena izmed shem ponavljanje elementa dovoljuje, druga ne. Taki problemi in še precej drugih se redno pojavlja v procesu preslikave, nekateri izmed njih pa so nerešljivi.

Druga variacija postopka oblikovanja po LaPlantu in St. Pierreu (1998) pa je sestavljena iz enakih postopkov, vendar sta tu izpostavljeni le dve fazi: harmonizacija in semantično preslikavanje. Faza »pravila« je uvrščena znotraj faze harmonizacije, medtem ko faza »pretvorba« poteka znotraj semantičnega prevajanja in je še vedno potek oblikovanja končnega izdelka oziroma dokumenta – prehoda med metapodatkovnima shemama. V praktičnem delu tega diplomskega seminarja je predstavljen postopek oblikovanja po slednji razvrstitvi shem, predvsem ker je bilo treba znotraj sintakse upoštevati tudi različne lastnosti obeh shem, po katerih je potem oblikovana skupna sintaksa prehoda iz aplikacijskega profila v shemo DCTERMS.

1.3 SISTORY.SI

Definicija na spletni strani portal definira oziroma opiše kot »spletni sistem slovenskega zgodovinopisja, ki poleg dostopnih historičnih vsebin, relevantnih za proučevanje slovenske zgodovine, omogoča ustvarjanje raziskovalnih podatkov in njihovo uporabo v raziskovalnem okolju« (Zgodovina Slovenije SIStory, 2016).

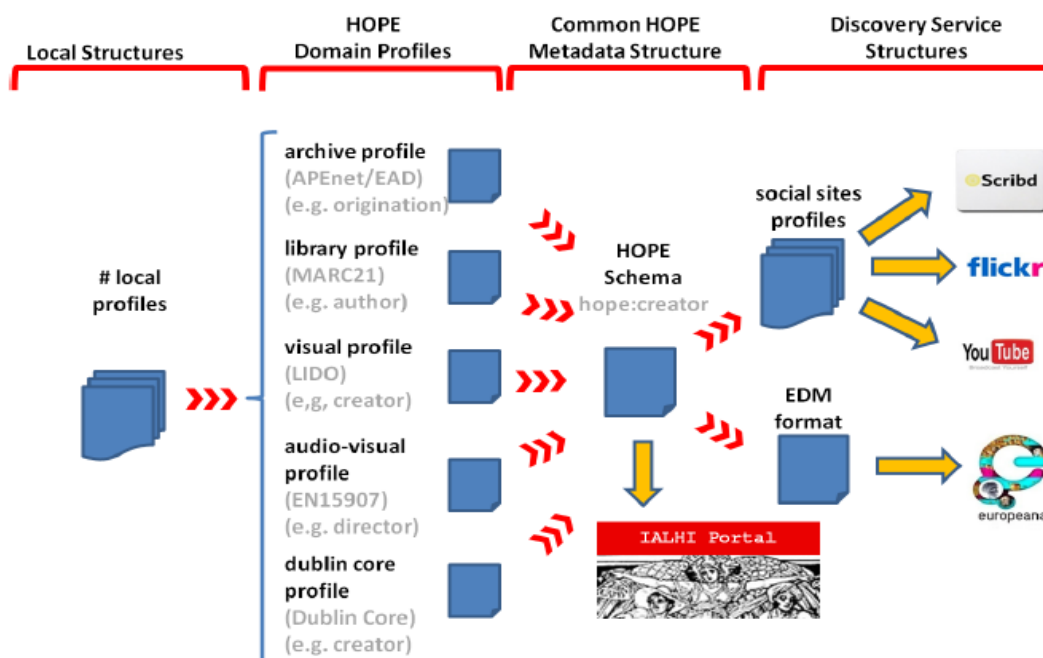
Portal je soroden drugim znanim slovenskim portalom, kot so Kamra, Museums.si in Dlib, vendar pa se od njih razlikuje pri obravnavanju informacijskih virov. Medtem ko se prej omenjeni portali koncentrirajo večinoma le na posamezno vrsto gradiva, se portal SIStory osredotoča na različne vrste gradiva v sklopu raziskovalnega dela ene same domene. Osnova raziskovalnemu delu so tako imenovani »primarni viri«, rezultat raziskovalnega dela pa so lahko različne publikacije (literatura). Slednje so kot nov vir lahko predstavljene na konferenci, lahko pa so podatki iz te publikacije uvrščeni v bazo podatkov. Vsak informacijski objekt je tako lahko obravnavan oziroma opisan kot primarni vir ali kot rezultat raziskovalnega dela – isti vir se lahko pojavlja kot različni zapis v različnem kontekstu, kar pa zahteva ločene metapodatke za vsak zapis.

Začetki portala segajo v september 2008, z verzijo SIStory 1.0, ki je bila v tistem času zelo enostavna različica portala, kot ga poznamo danes. Vsebovala je osnovno shemo Dublin Core s 15 elementi, ki je sčasoma prerastla v DCTERMS. V letu 2013 je potekal projekt nadgradnje prvotne verzije SIStory 1.0 v SIStory 2.0, ki je kot enega izmed rezultatov nadgradnje predstavil novi aplikacijski profil SIStory, ki je ostal v uporabi do trenutno najnovejše verzije portala – SIStory 3.0, ki je bila ustvarjena leta 2016. V tej nadgradnji je bila predstavljena osrednja tematika tega diplomskega seminarja, prehod iz aplikacijskega profila SIStory v metapodatkovno shemo DCTERMS za namen enostavnejšega prikazovanja elementov na portalu SIStory.

V okviru portala SIStory sta vzdrževani tudi dve večji lastni podatkovni bazi – Baza smrtnih žrtev druge svetovne vojne s popisom oseb, ki so umrle v obdobju od 1941 do 1945 zaradi vojnega nasilja ali posledicami vojne in Zgodovinarski indeks citiranosti – baza citatov iz slovenskih zgodovinskih monografij, osrednjih znanstvenih časopisov in revij. Poleg tega je portal vključen v DARIAH-SI (*Digital Research Infrastructure for the Arts and Humanities*), kjer sodeluje v projektu SI-DIH, iskalniku, ki omogoča iskanje podatkov po različnih repozitorijih s področja humanistike in umetnosti.

1.3 APLIKACIJSKI PROFIL SISTORY

Namen aplikacijskega profila je v splošnem prilagajanje metapodatkovne sheme tako, da je v uporabi na določenem področju. V primeru aplikacijskega profila Sistory (v nadaljevanju AP Sistory) gre za aplikacijski profil, ki je prilagojen za uporabo na področju zgodovinopisja – na portalu Sistory. Kot je bilo omenjeno, je v letu 2013 potekal projekt nadgradnje portala, katerega del je bilo tudi oblikovanje aplikacijskega portala Sistory – razlog za to odločitev je dejstvo da le elementi DCTERMS niso bili dovolj, da bi lahko zadostno opisali informacijske vire različnih vrst. Predvsem se je treba zavedati, da ima večina teh vrst virov »lastne« metapodatkovne sheme, ki se ukvarjajo izključno s tem, kako tako vrsto virov čim bolje opisati (primer: za opisovanje arhivskih virov obstaja metapodatkovni standard EAD). Zato je bilo treba iskati elemente, ki bi tak opis omogočali – te pa so našli v aplikacijskem profilu HOPE. Da pa lahko lažje predstavimo strukturo aplikacijskega profila Sistory, je treba pred tem opisati metapodatkovni shemi, ki ga sestavljata in iz katerih si »izposoja« elemente. HOPE (*Heritage of People's Europe*) povezuje digitalne zbirke o socialni zgodovini in delavskem gibanju ljudi v Evropi, ki jih deli na portalu Europeana (HOPE, 2013). Za opisovanje svojih dokumentov so razvili aplikacijski profil HOPE, ki ga sestavljajo elementi različnih metapodatkovnih shem.



Slika 1: Struktura aplikacijskega profila HOPE (Lemmens et al, 2011)

Slika 1 razlaga strukturo aplikacijskega profila HOPE, kjer je del, pomemben za razumevanje sestave AP SIStory, prikazan v drugem stolpcu z oznako domenski profili HOPE (*HOPE Domain Profiles*). Slednji prikazuje elemente drugih metapodatkovnih shem, oblikovane v profile, ki omogočajo opisovanje različnih vrst gradiva. HOPE tako vsebuje izbrane elemente iz različnih metapodatkovnih shem:

- arhivski profil – elementi iz EAD 2002 in ISAD(G),
- knjižnični profil (skupina z elementi za opisovanje knjižničnega gradiva) – MARC21 in ISBD Consolidated,
- vizualni profil (za slikovno gradivo) – LIDO in Spectrum 2.0,
- avdiovizualni profil – EFG in EN15907
- profil Dublin Core– uporabljeni elementi iz Dublin Core.

Naslednja shema, iz katere si AP SIStory sposoja elemente, pa je Dublin Core. Dublin Core oziroma *Dublin Core Metadata Element Set* (DCMES) je metapodatkovna shema, ki služi kot osnova precej bolj kompleksnim shemam in je še do danes ena najbolj poznanih in uporabljenih shem. Osnovna shema je vključevala 15 elementov, oblikovanih za namen opisovanja različnih tipov informacijskih objektov. Ta set elementov temelji na principu enostavnosti, zato je bila poimenovana *Simple Dublin Core* iz tega pa se je postopoma začela razvijati kompleksnejša in natančnejša verzija Dublin Cora, imenovana *DCMI Metadata Terms* oziroma DCTERMS. Ena izmed vidnejših razlik oziroma nadgradenj sheme je bila vključitev kvalifikatorjev (*qualifiers*).

AP SIStory si tako sposoja elemente in tudi določene lastnosti iz obeh metapodatkovnih shem. Če pogledamo aplikacijski profil HOPE in ga primerjamo s Prilogo 1 (Aplikacijski profil SIStory), v kateri so predstavljeni elementi znotraj AP SIStory, lahko opazimo določene podobnosti s HOPE.

Prva izmed takih podobnosti je samo dejstvo, da gre v obeh primerih za aplikacijski profil. Kot je že bilo omenjeno zgoraj, HOPE uporablja elemente iz petih različnih metapodatkovnih shem in jih združuje v posamezne skupine glede na vrsto gradiva– EAD, MARC21, LIDO, EN15907 in DC. AP SIStory si iz HOPE sposoja izbrane elemente iz vsake izmed skupin elementov (oziroma profilov). Tako torej AP SIStory sicer združuje elemente HOPE in DCTERMS, v resnici pa ti elementi izhajajo iz različnih metapodatkovnih shem – v Prilogi 1

to prikazuje stolpec »originalni imenski prostori«. Iz tega lahko razberemo, iz katerega imenskega prostora je element originalno izviral pred priključitvijo v aplikacijski profil HOPE (AP HOPE) in pozneje AP Sistory. Aplikacijski profil Sistory je v marsikaterem pogledu zelo podoben AP HOPE – do te mere, da bi se lahko vprašali, ali je Sistory le preimenovan aplikacijski profil HOPE. Temu pa ni tako, saj aplikacijski profil HOPE predvideva uporabo le osnovne metapodatkovne sheme Dublin Core (osnovnih 15 elementov). Ker je bila za AP Sistory predvidena uporaba kvalifikatorjev, je bilo treba profilu priključiti elemente DCTERMS.

Tudi organizacija elementov v profile je podobnost, ki je bila povzeta po HOPE – to je sicer delno vidno iz Priloge 1, kjer so elementi profila organizirani v 33 sklopov. Zadnji štirje sklopi elementov so poimenovani Arhivski elementi, Knjižničarski elementi in Za slikovno gradivo ter Za avdio-vizualno gradivo in vsebujejo določene izbrane elemente, ki so bili prevzeti iz aplikacijskega profila HOPE. Hkrati pa si od HOPE ne izposoja vseh elementov, temveč le določene, ki so potrebni za opisovanje informacijskih virov za različne vrste gradiva (arhivske, knjižnične dokumente, slikovne in avdio-vizualne). V Prilogi 1 so sklopi in elementi, ki izhajajo iz aplikacijskega profila HOPE, za jasnost ločeni od elementov DCTERMS.

Prav tako je treba obrazložiti določene posebnosti elementov v aplikacijskem profilu – čeprav je AP Sistory skupek izbranih elementov med HOPE in DCTERMS, pa za potrebe portala vsebuje tudi določene lastne elemente:

- MENU (*sistory:menu*) – Meni sklopa (zbirke), v okviru katerega je objekt objavljen na portalu Sistory;
- Sistory:Title (*sistory:title*) – Angleški prevod originalnega naslova objekta. Trenutno se element ne uporablja več, njegovo funkcijo pa je prevzel element *dcterms:alternative*;
- Avtor(-ji) (*sistory:author*) – Ime in priimek avtorja dela. Soroden elementu *dcterms:creator*, vendar pa se po definiciji nanaša le na osebo kot avtorja dela, ne pa tudi organizacije;
- Predmet, Vsebina (*sistory:subject*) – Slovenske in angleške ključne besede nekontroliranega vokabularja. S tem se nadomešča originalni element *dc:subject*. Za razliko od njega se v sklopu Predmet, Vsebina vpisuje v dve posebni polji, kjer se lahko ključne besede vpisuje po lastni presoji;

- Opis, Povzetek (*sistory:description*) – Opisna predstavitev publikacije na slovenski in angleški spletni strani portala Sistory. Ta element razširja pomen elementa *dcterms:description*;
- Sistory Type (*sistory:type*) – Za razliko od sorodnega elementa Type, ki opisuje samo tip gradiva, se s *Sistory Type* označuje tudi njegovo funkcionalnost na portalu Sistory;
- Publikacija (*sistory:publication*) – Digitalne datoteke objavljenih objektov, skupaj z njihovimi metapodatki;
- Zbirka (Collection) (*sistory:collection*) – Zbirka je skupek publikacij, del katerih je tudi objavljeno delo (ožji pomen od *isPartOf*) (Pančur, 2013).

Večinoma so ti elementi sorodni elementom DCTERMS, s katerimi si delijo ime, ne pa tudi pomena. Ustvarjeni so bili, ker ti spremenjeni pomeni elementov dopolnijo, kar elementi DCTERMS ne morejo. Vsi elementi Sistory pa se ne prikazujejo na zapisu metapodatkov, temveč so določeni namenjeni le za rabo v sistemu. Zato so ti elementi v shemi AP Sistory pridobili imenski prostor *sistory:* (Priloga 1), s katerim se jih loči od elementov z istim imenom. Podrobnejši opis strukture in lastnosti pa je podan v praktičnem delu v podpoglavju Oblikovanje prehoda med aplikacijskim profilom Sistory in DCTERMS.

2. PRAKTIČNI DEL

Leta 2016 je potekala že omenjena nadgradnja portala SIStory v različico 3.0. Ena izmed sprememb na področju implementacije metapodatkov na portalu je bila tudi vključitev prehoda med AP SIStory (metapodatkovno shemo SIStory 2.0 od leta 2013) in DCTERMS. Ta naloga je potekala med opravljanjem prakse na Inštitutu za novejšo zgodovino. Postopek oblikovanja prehoda bo predstavljen v tem delu diplomskega seminarja, saj je razumevanje prehoda in njegove sestave potrebno za razumevanje rezultatov.

2.1 OBLIKOVANJE PREHODA MED AP SISTORY IN DCTERMS

V procesu oblikovanja prehoda med AP SIStory in DCTERMS sta v ospredju dve pomembnejši fazi – harmonizacija in semantična preslikava. Obe sta bili podrobneje prikazani v pregledu literature, kjer je harmonizacija predstavljena kot oblikovanje skupne sintakse izvirne in ciljne sheme. Da to lahko naredimo, je treba poznati obe shemi in njuni sintaksi. Za opis in pozneje primerjavo obeh sintaks bo pri lastnostih izpostavljena terminologija (definicije elementov in njihova uporaba), tip vnosa elementov (obveznost elementov) in način zapisa vrednosti elementov (uporaba shem za kodiranje podatkov). Terminologija aplikacijskega profila SIStory za opis elementa sledi naslednji obliki:

- Definicija: opredelitev pomena elementa;
- element HOPE: elementi, dodani aplikacijskemu profilu, v večini izhajajo iz metapodatkovne sheme HOPE;
- Originalna definicija: definicija, dodeljena v shemi HOPE;
- Izvorni element: URI-povezava na element iz izvirne sheme;
- Izvorna definicija: definicija elementa v izvorni shemi;
- XML-element: navedba imena, oznake elementa, kot se prikazuje znotraj XML-sheme AP SIStory;
- Tip vnosa: navedba, ali je element obvezen ali opsijski;
- Uporaba: opisana sta predvidena uporaba elementa in vpis v vmesnik (Pančur, 2013).

Kot je mogoče opaziti, ima terminologija SIStory več različnih definicij – vsi elementi pridobijo definicijo, originalno in izvorno definicijo pa le elementi, ki izvirajo iz HOPE. Slednje naj bi pripomoglo k lažjemu razumevanju aplikacijskega profila.

Naslednja izmed »kategorij« za oblikovanje skupne sintakse je tip vnosa elementov. Obveznost elementa je tu podana z omejitvijo števila pojavljanj elementa v zapisu (minimalno, maksimalno število pojavljanj). Pri tem imajo obvezni elementi različno število pojavljanj, neobvezni elementi pa v večini minimalno število pojavljanj 0 in maksimalno neskončno. AP SISTORY pri tem uporablja obvezne in neobvezne elemente. Obvezni elementi in omejitve števila pojavljanj, ki jih AP SISTORY zahteva, so elementi Menu_Slo, Jezik, Naslov in SISTORY Type.

V terminologiji AP SISTORY so vsakemu elementu podani definicija in navodila, kako uporabljati element znotraj orodja za vnos metapodatkov. Na podoben način je predstavljena tudi uporaba kontroliranih slovarjev in drugih kodirnih shem – določeni elementi imajo priporočeno uporabo določenih kontroliranih slovarjev in shem za kodiranje podatkov. Najpogosteje uporabljeni so W3CDTF za zapis datumov, za oznako jezika se uporablja že v orodje vstavljen kontrolirani slovar (po ISO639-3), DCMI Type za oznako vrste informacijskega objekta in licenca Creative Commons 2,5 za vstavljanje HTML-kode za uravnavanje avtorskih pravic (Pančur, 2013).

V tem prehodu DCTERMS igra vlogo izhodiščne sheme, ki jo je ravno tako treba opisati in iz nje izluščiti pomembne lastnosti za poznejšo fazo harmonizacije oziroma oblikovanja skupne sintakse. Vsak izmed elementov je po terminologiji DCMI Terms (2012) opremljen z atributi. Osnovni atributi za opis elementa so Ime (*Name*), Oznaka (*Label*), URI, Definicija in Tip Elementa (*Type of term*), pri tem pa so lahko tem dodani tudi dodatni atributi, kot so Komentar, Glej in podobni. DCTERMS nobenega izmed elementov ne predvideva kot obveznega, neobvezni elementi pa imajo število pojavljanj omejeno na minimalno število 0 in maksimalno število neskončno. Za omejitve zapisa vrednosti elementa je v Dublin Core priporočena uporaba standardov za podatkovno vsebino, kot so DCMI Type, LCSH in UDC. Za kodiranje v metapodatkovnem zapisu pa so med najbolj uporabljenimi standardi standard W3CDTF za zapis datumov, ISO639-3 za zapis jezikovnih kod, URI, Period in Point pa za označevanje časa in geografske lokacije.

2.1.1 Harmonizacija

Iz opisov aplikacijskega profila Sistory in metapodatkovne sheme DCTERMS smo izpostavili skupne lastnosti, ki jih v fazi harmonizacije preoblikujemo v skupno sintakso. Tabela 2 prikazuje izpostavljene lastnosti obeh shem in skupno shemo po poteku faze harmonizacije.

Tabela 2: Oblikovanje skupne sintakse prehoda

	AP Sistory	DCTERMS	HARMONIZACIJA
Terminologija	<ul style="list-style-type: none"> Definicija Element HOPE Originalna definicija HOPE Izvorni element Izvorna definicija XML-element Tip vnosa Uporaba 	<ul style="list-style-type: none"> Ime (<i>name</i>) Oznaka (<i>label</i>) URI Definicija Tip elementa (<i>Type of term</i>) 	<ul style="list-style-type: none"> Ime (<i>name</i>) Oznaka (<i>label</i>) Definicija (vse definicije – Sistory HOPE (originalna) in izvorna definicija) Tip vnosa (<i>Type of term</i>) Uporaba URI
Tip vnosa	Obvezni elementi: <ul style="list-style-type: none"> MENU_SLO/ANG Jezik Naslov Sistory Type 	Nima obveznih elementov.	Obvezni elementi: <ul style="list-style-type: none"> MENU_SLO/ANG Jezik Naslov Sistory Type Ostali elementi so neobvezni.
Sheme za kodiranje podatkov	DCMI Type, IMT, LCSH, UDC, W3CDTF, ISO639-3, (ARRS), Creative Commons 3.0, Period, Point	DCMI Type, IMT, LCSH, UDC, W3CDTF, ISO639-2, ISO639-3, URI, Period, Point	V uporabi so sheme za kodiranje podatkov iz obeh shem.

Tabela 2 prikazuje skupne lastnosti AP SIstory in DCTERMS, združene v skupni sintaksi, uporabljeni za oblikovanje prehoda med obema shemama. Pri tem so bile upoštevane tri ključne lastnosti – terminologija, tip vnosa elementov in uporabljene sheme za kodiranje podatkov. Terminologija AP SIstory je bila ustvarjena po vzoru terminologije iz DCTERMS, kar se odraža na določenih skupnih poimenovanjih (Ime Elementa, Definicija, URI, Oznaka). To je sicer pomenilo veliko prednost pri oblikovanju skupne terminologije, vendar je bilo pri tem treba upoštevati tudi terminologijo AP SIstory.

Primer takega problema je vključevanje vseh definicij elementov z namenom lažjega sledenja izvora elementov. Gre za nekakšno dokumentacijo elementov za prihodnje nadgradnje portala. V skupni sintaksi so tako združeni termini iz obeh shem.

Naslednja izmed lastnosti, ki jih je treba vključiti v skupno sintakso, je tip vnosa oziroma način ponavljanja elementov. DCTERMS v svoji sintaksi ne predvideva nobenih obveznih elementov, vendar pa je na portalu treba prikazati vsaj osnovne podatke o informacijskem objektu, ki pa so zato označeni kot obvezni elementi. V obeh sintaksah so elementi omejeni s številom ponavljanj.

Nazadnje pa se je skupna sintaksa oblikovala še po shemah za kodiranje podatkov (standardi za podatkovne vrednosti) – po primerjavi sintakse obeh shem je bilo ugotovljeno, da tako AP SIstory kot DCTERMS uporabljata enake sheme za kodiranje podatkov. Pri harmonizaciji je to preprosto pomenilo, da se v skupni sintaksi uporablja vse skupne sheme.

2.1.2 Semantična preslikava

Kot je bilo predstavljeno v teoretičnem delu, poznamo dva koncepta oblikovanja prehoda oziroma preslikavanja elementov – absolutni in relativni prehod. Elementi obeh shem, udeleženi pri preslikavi si med seboj niso bili semantično ekvivalentni. Ker pa v tem primeru semantična ekvivalentnost ne igra večje vloge pri uporabi, je bil pri oblikovanju uporabljen relativni prehod. V nasprotnem primeru se elementi, ki si semantično niso ekvivalentni, ne bi mogli preslikati, s tem pa bi propadel tudi smisel oblikovanja prehoda. Preslikava je potekala s pomočjo tabele Excel – Semantična preslikava elementov (Priloga 2). Vsak izmed elementov aplikacijskega profila Sistory je dobil preslikavo v element DCTERMS. Prehod je bil dodatno opremljen z URI-povezavami vsakega izmed preslikanih elementov, opredeljenim načinom zapisa in dodatnimi zahtevami pri vpisu metapodatka.

Preslikava je potekala na način »ena-v-ena« (*one-to-one*) in »mnogo-v-ena« (*many-to-one*). Preslikava »ena-v-ena« je potekala le na elementih DCTERMS, ki so del aplikacijskega profila Sistory, kar praktično pomeni, da gre za preslikavo elementa samega vase – element *dc:date* se tako preslika iz izvirne sheme v ciljno shemo kot element *dc:date*, prav tako to velja za ostale elemente in njihove kvalifikatorje. Preslikava »mnogo-v-ena« pa je potekala na elementih HOPE v aplikacijskem profilu Sistory. Več elementov HOPE se je tako preslikalo v en sam element DCTERMS. Načina preslikav elementov sta vidna na primeru prehodov elementa Type v Tabeli 3, kjer je prva preslikava elementa, ki izvirno spada v shemo DCTERMS (*dcterms:type*), preslikava elementa v samega sebe oziroma je to preslikava »ena-v-ena«. V isti element pa se preslika še trinajst elementov iz metapodatkovne sheme HOPE (znotraj aplikacijskega profila), kar pa predstavlja preslikavo »mnogo-v-ena«.

Tabela 3: Preslikava elementov DCTERMS in HOPE v element *dcterms:type*

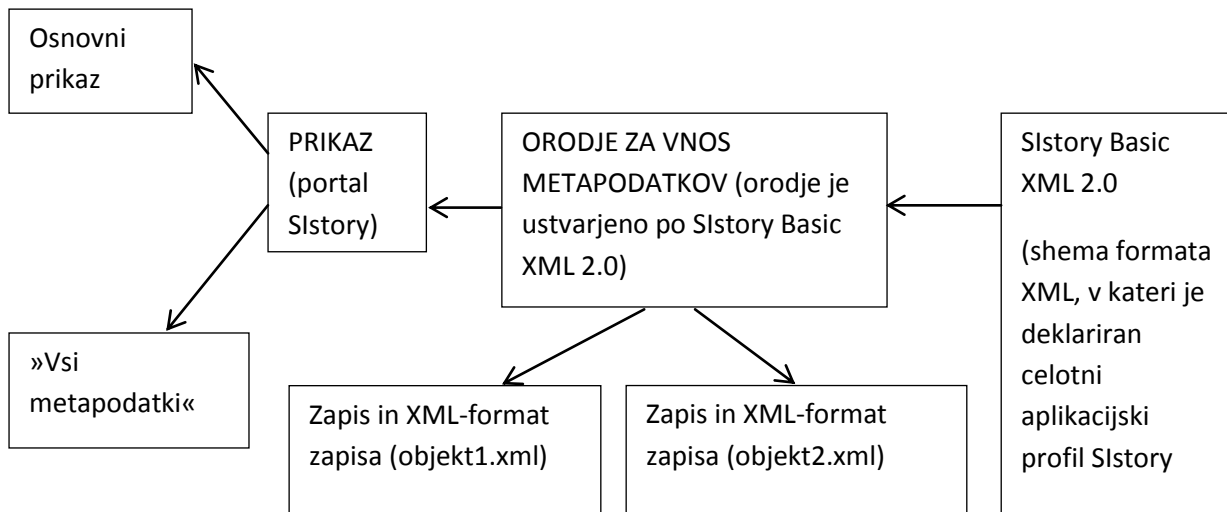
Aplikacijski profil	Dublin Core
Type	dcterms:type
archlevel	dcterms:type
archGenreForm	dcterms:type
libLevel	dcterms:type
libGenMaterialD esignation	dcterms:type
visLevel	dcterms:type
visObjectName	dcterms:type
visTechique	dcterms:type
visTechAttribute	dcterms:type
avLevel	dcterms:type
avIstalationType	dcterms:type
avCarrierType	dcterms:type
avColourSystem	dcterms:type
avSoundSystem	dcterms:type

Preslikava elementov je bila izvedena kot relativni prehod, v katerem je semantična ekvivalenca stranskega pomena. Vseeno je tudi za relativen prehod potrebna vsaj določena stopnja semantičnega ujemanja med elementi izvorne in ciljne sheme. Zato sem se pri tem opirala na definicije posameznih elementov v dokumentaciji AP Sistory (*Navodila za uporabo orodja za vnos metapodatkov*) in jih primerjala z definicijami v DCTERMS. Večjih problemov pri iskanju ustreznih semantično ekvivalentnih elementov za preslikavo ni bilo, saj je večina elementov ponovljivih v obeh shemah, prav tako pa ciljna shema ne predvideva svojih obveznih elementov, ki bi lahko potencialno otežili preslikavo elementov.

Po končani stopnji semantične preslikave je sledila realizacija prehoda med metapodatkovnima shemama. Način in stopnja implementacije aplikacijskega profila Sistory in prehoda med AP Sistory ter DCTERMS pa je prikazana v naslednjem poglavju na primeru zapisa o posmrtni maski Antona Aškerca iz zbirke *Odlivanje smrti*.

3. PRIMER IMPLEMENTACIJE METAPODATKOV

Diagram 1: Prikaz delovanja vseh treh nivojev v portalu Sistory



Celotna struktura portala je zasnovana na XML-shemi aplikacijskega profila, poimenovani Sistory Basic XML 2.0. V njej so deklarirani vsi elementi, njihovi atributi, števila ponavljanja elementov, sheme za kodiranje podatkov, ki so prisotni v sintaksi aplikacijskega profila itd. V tej shemi je torej temelj celotnega cikla nastajanja metapodatkov na portalu Sistory. Naslednji del je orodje za vnos metapodatkov, ki temelji na shemi Sistory Basic XML – po vzorcu in postavitvi elementov, atributov in ostalih delov je postavljen tudi vmesnik orodja za vnos metapodatkov. V njem je mogoče vnašati metapodatke za opisovanje informacijskih objektov na portalu. Vsakemu izmed končanih zapisov v orodju se generira tudi zapis v XML, ki ga je v orodju mogoče izvažati in uporabljati.

V zadnjem delu pa se zapis iz orodja za vnos metapodatkov prikaže na vmesniku portala Sistory, kjer sta ustvarjena dva različna prikaza – osnovni prikaz, ki prikazuje »osnovne« metapodatke, kot so Naslov, Avtor, Leto, Jezik in ID (*handle* povezava) zapisa, ter razdelek »Vsi metapodatki«, v katerem pa so predstavljeni vsi opisni metapodatki, ki jih z orodjem vnašamo na portal. Oba prikaza se za lažjo predstavo nahajata v Prilogi 5. Pri tem so v razdelku izpuščeni določeni elementi, kot so na primer sistemski elementi. Na primeru zapisa informacijskega objekta z naslovom »Anton Aškerc (MGML – Muzej in Galerije Mesta Ljubljane)« bomo v tem delu podrobneje pregledali celotno zgradbo in potek ustvarjanja metapodatkov na portalu ter na vsakem »nivoju« ugotavljali, ali sta AP Sistory in prehod implementirana na določenem nivoju, zakaj da ali ne ter na kakšen način.

3.1 IMPLEMENTACIJA NA NIVOJU XML-SHEME

Kot lahko vidimo po Diagramu 1, celotna struktura sistema in proces opisovanja informacijskih virov izhajata iz SIstory Basic XML 2.0. XML (*eXtensible Markup Language*) je razširljiv označevalni jezik, ki namesto točno določenih oznak/imen elementov omogoča uporabo lastnih oznak (W3C, 2016).

SIstory Basic XML 2.0 je shema, ki v formatu XML s svojimi označevalci deklarira celotno sestavo AP SIstory – njegove elemente, sheme za kodiranje, ponovljivost elementov itd. Kot je bilo predhodno omenjeno, se vsakemu izmed končanih zapisov v orodju generira tudi zapis v XML-formatu. Celotna shema SIstory Basic XML 2.0 je vključena v Prilogo 3, XML-zapis primera Antona Aškerca pa v Prilogo 4.

Znotraj tega zapisa so XML-elementi prikazani v obliki `<ime_elementa ime_atributa=»vrednost atributa«>`. Poleg elementov opisnih metapodatkov so v XML-zapisu prikazani tudi sistemski elementi, ki so zadolženi za zapis dodatnih informacij. Primeri takih elementov so na primer DATE/TIME_ADDED (zapišejo točen datum in časovno kodo, ko je bil zapis objavljen na portalu), STATUS (vrednosti 0 ali 1 obveščata o tem, ali je zapis aktiven ali ne, pri tem je številka 1 enaka vrednosti »aktiven«) ali LOCK (vrednosti sta ravno tako binarni in obveščata o tem, ali je zapis zaklenjen).

Če pogledamo, na kakšen način sta v XML-format implementirana AP SIstory in prehod, lahko to na primeru zapisa Antona Aškerca vidimo na tem delu:

```
<SLIKOVNI slikovniEType="visTechnique">Mavec, patiniran</SLIKOVNI>  
  <SLIKOVNI slikovniEType="visTechAttribute">Maska je v slabem  
stanju.</SLIKOVNI>  
  <SLIKOVNI slikovniEType="visTechAttribute">Na zgornjem delu vstavljen kovinski  
kavelj za obešanje.</SLIKOVNI>
```

Zgornji del XML-zapisa predstavlja elemente, ki se v orodju nahajajo v razdelku »Za slikovne elemente«, pri čemer je poimenovanje elementov vprašljivo – poimenovanje, za katero so se razvijalci Sistory odločili, ne sledi nikakršnim standardom. Hkrati pa so v tem delu pri oblikovanju imen elementov naredili napako poimenovanja: *visTechnique* in *visTechAttribute* izvirata iz metapodatkovnega standarda LIDO, ki opisuje muzejske informacijske objekte. XML-elementa za ta elementa bi morala biti poimenovana bližje muzealijam, kot na primer `<MUZEJ muzealijeElType>` ali podobno, saj bi morala odražati, kaj element opisuje. Napaka se je najverjetneje zgodila zaradi slik, ki so izpostavljene v prikazu zapisa. V izogib takšnim napakam bi bilo najbolje upoštevati standardna poimenovanja elementov (za XML-elemente bi uporabili originalna imena elementov) ali uporabljati za to določen standard oziroma smernice.

Ker pa oba elementa ne pripadata DCTERMS, temveč aplikacijskem profilu HOPE, tu lahko vidimo implementacijo AP Sistory v XML-zapisu objekta – prisotni so elementi in atributi, ki so v osnovi del metapodatkovne sheme HOPE. SGML-jeziki (med katere spada tudi XML) dovoljujejo poljubno poimenovanje označevalcev, kar je odprlo nove možnosti poimenovanja. Čeprav bi ta možnost omogočala tudi implementacijo prehoda, so se razvijalci odločili, da vsak izmed elementov in atributov pridobi unikatno poimenovanje. Ta odločitev je posledično pomenila, da prehod med AP Sistory in DCTERMS v XML ne bo vključen. Verjeten razlog za to določitev je bilo ogromno število elementov, ki se preslikajo v element *dcterms:type*. Slednje bi lahko predstavljalo problem pri generiranju XML-formata zapisa iz informacijskega objekta (problem generiranja XML-formata zaradi uporabe relacijskih baz MySQL), kar bi lahko imelo za posledico napačen prikaz elementov v zapisu.

3.2 IMPLEMENTACIJA NA NIVOJU SISTEMA

Sistem, poimenovan SISTORY administracija ali na kratko admin, ob prijavi omogoča različne storitve, ki so povezane z vzdrževanjem portala SISTORY. Največji del sistema predstavlja orodje za vnos metapodatkov. Ob vnosu nove publikacije za objavo na portalu, morebitnem popravljanju ali posodabljanju že obstoječih zapisov je v sistem vgrajeno orodje za vpis metapodatkov. Orodje je osnovano na aplikacijskem profilu SISTORY – vsebuje vse elemente in kvalifikatorje v 33 sklopih elementov, pri tem so ti sklopi v vmesnik vstavljeni v enakem vrstnem redu, kot je prikazano v AP SISTORY.

Celotna zasnova elementov je pri tem upoštevala vodilno pravilo – orodje za vnašanje metapodatkov mora biti kar se da enostavno za uporabo, pri tem pa se ta enostavna uporaba nanaša na vnašalce metapodatkov, saj uporabniki nimajo dostopa do administracije. Čeprav je pri zasnovi prevladovalo to pravilo, pa so bile potrebne določene stopnje prilagoditve za sistem. Aplikacijski profil oziroma njegova implementacija se torej odražata že v sami strukturi, sintaksi, semantiki orodja za vnašanje metapodatkov. Glavna lastnost aplikacijskega profila pa je, da vsebuje elemente iz drugih shem. Na primeru zapisa posmrtna maska je ta koncept prikazan v sklopu elementov, ki so namenjeni opisovanju slikovnega gradiva.

Za slikovno gradivo	
visTechnique	Mavec, patiniran
visTechAttribute	Maska je v slabem stanju.
visTechAttribute	Na zgornjem delu vstavljen kovinski kavelj za obešanje.

Slika 2: Prikaz elementov za slikovno gradivo v orodju za vstavljanje metapodatkov

Na dani sliki sta vidni imeni obeh elementov, ki opisujeta sliko informacijskega objekta. Elementa ne izhajata iz sheme DCTERMS, temveč iz HOPE oziroma izvirno iz metapodatkovnega standarda LIDO.

Prehod med AP Sistory in DCTERMS je v tem sistemu zanemarjen oziroma je izpuščen zaradi jasnosti pomenov elementov in sledenja pravilu enostavnosti za vnašalca metapodatkov. Vključitev prehoda v sam sistem orodja za vnos metapodatkov bi namreč povzročila velike probleme pri vnašanju metapodatkov.

Ker so se pri preslikavi elementov iz AP Sistory v DCTERMS elementi HOPE preslikali na način »mnogo-v-ena«, je v preslikavi večje število elementov z enakim imenom (glej Tabela 3 – trinajst elementov se preslika v element *dcterms:type*). V primeru, da bi element v sistemu preslikali na način »mного-v-ena«, bi to povzročilo problem pri prepoznavanju konteksta elementa. Ta problem pa predstavlja najbolj verjeten razlog, ki je spremljal odločitev, da se znotraj sistema zanemari implementacijo prehoda.

3.3 IMPLEMENTACIJA NA NIVOJU PORTALA

Standardni prikaz, ki se nam odpre ob izboru zapisa, je sestavljen iz treh večjih delov. Prvi razdelek vsebuje sliko in osnovne metapodatke o informacijskem objektu – Naslov, Ime Avtorja, Vrsto gradiva, Leto, Ključne besede, Zbirko, Trajno povezavo oziroma Identifikator objekta in v tem primeru dodatni identifikator. Drugi del vsebuje datoteke za prenos objekta, pri čemer imajo tudi datoteke za prenos svoje lastne metapodatke, kot so naslov, ime, velikost datoteke, format in stalna povezava datoteke. Zadnji, za prikaz implementacije metapodatkov in prehoda najpomembnejši del, pa je razdelek »Vsi metapodatki«.

Razdelek »Vsi metapodatki« natančneje prikazuje način implementacije novega nabora elementov, v katerega je vključen prehod med aplikacijskim profilom in shemo DCTERMS. Vsakemu izmed elementov v razdelku je dodana oznaka s povezavo na spletno stran DCMI z opisom elementa Dublin Core. Iz Priloge 5 (Vsi metapodatki) sta bili za lažjo predstavo vzeti obe pojavitvi *dcterms:type* za primerjavo njunih kontekstov in implementacije AP SISTORY ter prehoda.

- *dcterms:type* ⓘ
 - Physical Object
- *dcterms:type* ⓘ
 - Mavec, patiniran
 - Maska je v slabem stanju.
 - Na zgornjem delu vstavljen kovinski kavelj za obešanje.

Slika 3: Prikaz dveh elementov *dcterms:type* in njuna konteksta v istem zapisu (SISTORY.si, 2016)

Na tem primeru dveh elementov z istim imenom lahko razložimo način implementacije AP SISTORY in prehoda v DCTERMS.

Prvi element *dcterms:type* se nanaša na vrsto informacijskega objekta po kontroliranem slovarju DCMI Type. V primeru tega zapisa gre za fizični objekt (*Physical Object*), saj je objekt v tem primeru odlita maska Antona Aškerca. V drugem primeru pa se element *dcterms:type* nanaša na material in način, ki sta bila uporabljena za izdelavo maske, ter opis trenutnega fizičnega stanja maske. To ponavljanje elementov z istim imenom je posledica preslikave »mnogo-v-ena«, kot je bilo prikazano v Tabeli 3. Prvi primer je torej originalni element *dcterms:type*, ki izhaja iz metapodatkovnega standarda DCTERMS, medtem ko je drugi element preslikava elementov iz sheme HOPE.

Pri ponavljanju elementov z enakim imenom pa lahko prihaja do problematike razumevanja elementov – imena elementov bi bila lahko precej preprosteje poimenovana prostobesedno. Ta problem se je pojavil tudi pri oblikovanju prehoda za portal SIstory. V samem začetku razvijanja novega vmesnika prikaza je bil prehod za razdelek »Vsi metapodatki« načrtovan kot način poenostavljanja elementov in je s tem zagotavljal lažjo uporabo. Primarni uporabniki portala so raziskovalci na Inštitutu za novejšo zgodovino, ki pa v večini uporabljajo le osnovni prikaz z osnovnimi elementi (Naslov, Avtor, Leto, Jezik, URI). Tako se razdelka z vsemi opisnimi metapodatki informacijskega vira poslužujejo le redki.

To je bil najverjetneje razlog, da so se razvijalci na Inštitutu za novejšo zgodovino odločili za zasuk iz razvijanja prehoda za uporabnike na strojno berljivost elementov. Prvotna ideja je bila namreč vključitev RDFa¹ le v glavo (*header*) dokumenta. Pri tem je bila s strani razvijalcev izbrana shema DCTERMS kot ciljna shema. DCTERMS je sicer primerna izbira za predstavitev metapodatkov v RDFa, čeprav RDFa ne priporoča ali izpostavlja uporabe določene metapodatkovne sheme in bi tako lahko uporabili katerokoli drugo metapodatkovno shemo.

¹ RDFa (*Resource Description Framework in Attributes*) je W3C-specifikacija za attribute, ki omogočajo izražanje strukturiranih podatkov kateregakoli označevalnega jezika (HTML, XML) v strukturi dokumenta (W3C, 2015).

Po zaključku nadgradnje pa je zaradi neuporabe prikaza »Vsi metapodatki« ta prikaz postal tudi nekakšna »vidna različica« RDFa na prikazu – prehod in elementi v DCTERMS, ki so prisotni v razdelku »Vsi metapodatki«, so hkrati prisotni tudi v RDFa-zapisu v izvorni kodi dokumenta (RDFa je v večini primerov skrit uporabniku). Slednje je za primer zapisa Antona Aškercera vidno v Prilogi 6 (RDFa-zapis Anton Aškerc). Če ga primerjamo s Prilogo 5 (natančneje z razdelkom »Vsi metapodatki), lahko vidimo, da gre za enake metapodatke, preslikane v DCTERMS. Pri tem pa je ta zapis pomanjkljiv in bi potreboval dopolnitve, da bi pravilno opravljal funkcije RDFa.

Prehod je torej ostal implementiran v prikazu, vendar pa njegov namen zaradi pomanjkanja uporabe ni bil več olajševanje prikaza in poenostavljanje elementov za uporabo, temveč popolnejše, bogatejše opisovanje informacijskih objektov in strojne berljivosti.

Če torej na koncu naredimo pregled implementacije AP in prehoda, lahko v Tabeli 4 vidimo, ali sta AP in prehod implementirana, na katerih nivojih in kako.

Tabela 4: Pregled implementacije AP in prehoda

	Prikaz (vmesnik)	Sistem (orodje)	XML-prikaz zapisa
AP SIstory	Implementiran, prisotni elementi HOPE.	Implementiran, prisotni elementi HOPE.	Implementiran, prisotni elementi HOPE.
Prehod med AP SIstory in DCTERMS	Implementiran, preslikava »mnogo-v-ena«.	Ni implementiran, elementi za jasnost obdržijo svoja izvorna imena.	Ni implementiran, elementi pridobijo XML-element, ki jih definira enolično.

Če to tabelo prenesemo v Diagram 1, lahko vidimo, da vse izhaja iz SIstory Basic XML, ki je osnova za orodje za vnos metapodatkov. Iz sheme in orodja se generira XML zapis metapodatkov, v katerem je prisoten AP, ne pa tudi prehod. V primeru prisotnosti prehoda bi zaradi ponavljanja imen elementov lahko prišlo do napak zapisa (trije različni elementi sheme ne morejo biti deklarirani z enakim XML elementom).

V orodju je prisoten AP SIstory, prehod pa zaradi tehničnega razloga ni implementiran. Orodje in njegova polja za zapis elementov temeljijo na SIstory Basic XML 2.0, kar pomeni, da vsak element v orodju in zapisu pridobi unikaten XML-označevalec, ki ga enolično označi. Prehod ni dodan, ker uporabniki ne morejo dostopati do orodja, označevalci pa poznajo sestavo in elemente ter hitreje razberejo elemente .

Edini nivo, v katerem sta vključena tako AP kot tudi prehod, pa je prikaz na portalu. Vključenost obeh shem je vidna izključno v prikazu »Vsi metapodatki«, v osnovnem prikazu AP in prehod namreč nista potrebna (vidni so le osnovni metapodatki). Pri tem ponavljanje imen elementov v različnih kontekstih lahko deluje zmedeno za uporabnika, hkrati pa ta imenovanja omogočajo strojno berljivost metapodatkov v RDFa, za katero je DCTERMS najprimernejša. V tem pregledu lahko ugotovimo, da AP SIstory nastopa v vseh nivojih opisovanja informacijskih virov in je temelj celotnega procesa, medtem ko je prehod oblikovan izključno za prikaz na razdelku »Vsi metapodatki«. Njegovo vključevanje v druge nivoje bi bilo nesmiselno, saj bi proces le zakompliciral.

4. RAZPRAVA

Raziskovalno vprašanje, zastavljeno v uvodnem delu, se nanaša na način implementacije metapodatkov na portalu Sistory. Kot način implementacije sta v ospredje postavljena aplikacijski profil Sistory in pa prehod med njim in metapodatkovno shemo DCTERMS. Namen tega diplomskega seminarja je opisati način in stopnjo odražanja obeh konceptov znotraj različnih nivojev portala – na nivoju XML-zapisa, systemskega orodja za vnos metapodatkov in prikaza na portalu. V tem delu bo na podlagi praktičnega dela prikazana učinkovitost aplikacijskega profila in prehoda kot rešitev problema opisovanja različnih vrst informacijskih objektov na portalu.

Po prikazu implementacije teh dveh konceptov na primeru zapisa v praktičnem delu je bilo ugotovljeno, da je AP Sistory vključen v vseh treh nivojih, saj je temelj celotnega prikaza metapodatkov. Na njem temelji Sistory Basic XML 2.0, v kateri so deklarirani vsi elementi AP Sistory in po kateri je posledično zgrajeno orodje za vnos metapodatkov (ki vsebuje vse elemente in sintakso AP Sistory). AP Sistory pa je viden tudi v prikazu metapodatkov v razdelku »Vsi metapodatki«. Nasprotno od tega pa je prehod med AP Sistory in DCTERMS implementiran le v prikazu metapodatkov v razdelku »Vsi metapodatki« in tako opravlja funkcijo poenostavljanja prikaza metapodatkov na portalu.

Če pogledamo učinkovitost in smiselnost implementacije AP in prehoda, ugotovimo, da ima tako pozitivne kot negativne učinke. Z dodatnimi elementi, ki so se začeli bolj uporabljati, je AP omogočil podrobnejše in bogatejše opisovanje informacijskih virov različnih vrst. Prav tako se je z implementacijo prehoda občutno zmanjšalo število imen elementov v prikazu. Čeprav je to zelo pozitiven aspekt prehoda, pa se je še vedno treba vprašati, kako zelo razumljiva so ta imena elementov uporabniku oziroma ali to še vedno odraža pravilo enostavnosti. Argument, ki je spremljal odločitev za imena elementov v obliki *dcterms:imelementa*, je bilo dejstvo, da je taka oblika pomembna za strojno berljivost elementov v RDFa, tudi za namen kasnejše ponovne nadgradnje.

Hkrati pa se pojavlja tudi problem ponavljanja prevelikega števila elementov z enakimi poimenovanji – DCTERMS je sicer najbolj primerna metapodatkovna shema za povezovanje metapodatkov in strojno berljivost metapodatkov v RDFa. Prehod naj bi po svojem namenu poenostavil prikaz metapodatkov na portalu, kar je bilo delno doseženo, vendar pa problem

leži drugje – sredi procesa oblikovanja prehoda se je poudarek iz poenostavljanja prikaza prenesel na strojno berljivost metapodatkov.

Eden izmed možnih vzrokov za ta zasuk poudarka je najverjetneje potencialna uporaba na portalu, saj primarnim uporabnikom (raziskovalcem na Inštitutu za novejšo zgodovino) zadostujejo osnovni metapodatki za identifikacijo vira, ki ga potrebujejo. Tako je prehod in hkrati aplikacijski profil v prikazu postal »vidna različica« RDFa. Kljub pomanjkljivostim opravlja funkcijo strojne berljivosti metapodatkov, hkrati pa v ločenem prikazu na portalu ponuja bogatejšo opisovanje informacijskega vira, saj v nasprotju z osnovnim prikazom, ki vsebuje le osnovnih pet metapodatkov za identifikacijo vira, vsebuje celoten in podroben opis informacijskega objekta. V retrospektivi je bila ta preusmeritev poudarka ustrezna, saj je uporaba teh razširjenih metapodatkov nizka, viri so še vedno bogato opisani in dovolj poenostavljeni za potencialne uporabnike, ki bi jih potrebovali.

Če portal in nadgradnjo pogledamo v celoti, lahko rečemo, da slednja kljub določenim pomanjkljivostim opravlja naloge, za katere je bila oblikovana – aplikacijski profil je ustrezneje prikazan kot v prejšnji različici, prehod pa poenostavlja prikaz metapodatkov na portalu. Nadgradnja je bila smiselna, čeprav se je pomen za uporabnika, predvsem kar se tiče prehoda, prehitro preusmeril v poudarek na strojno berljivost v RDFa. In čeprav bi trenutni aplikacijski profil in prehod lahko posodobili ali popravili določene pomanjkljivosti, pa razvijalci sistema načrtujejo prehod obstoječe metapodatkovne sheme (aplikacijski profil s prehodom) na metapodatkovno shemo METS. Kot glavne opisne metapodatke bo vključeval Dublin Core, vendar bo mogoče poljubno dodajati tudi ostale opisne metapodatke po shemah TEI, LIDO, EAD. Vsekakor pa bo pri razvijanju te nove metapodatkovne sheme za SIstory treba paziti na poimenovanja elementov v shemah, se v procesu bolj opirati na različne standarde in celoten proces bolj podrobno dokumentirati – potreben bo bolj strokoven pristop k oblikovanju.

5. ZAKLJUČEK

AP SISTORY je implementiran v vse tri nivoje, saj je temelj celotnega procesa – na njem je osnovana shema SISTORY Basic XML 2.0, slednja je nato predloga za osnovanje orodja za vnos metapodatkov, nazadnje pa se AP SISTORY kaže tudi v prikazu metapodatkov, saj je eden izmed glavnih namenov AP pripomoči k bolj podrobnemu opisovanju informacijskih objektov. Prehod med AP SISTORY in DCTERMS je bil oblikovan izključno samo za prikaz metapodatkov na prikazu »Vsi metapodatki« kot način poenostavljenega načina prikaza.

Gledano v celoti sta način implementacije metapodatkov na portalu SISTORY in nadgradnja koristna oziroma efektivna, čeprav z zasukom poudarka iz uporabnika in poenostavljanja elementov na poudarek strojne berljivosti in bogatejšega opisovanja informacijskih virov. Z aplikacijskim profilom se reši problem bolj podrobnega opisovanja informacijskih virov, medtem ko prehod skrbi za strojno berljivost in bogatejši opis informacijskih virov na portalu. Kljub predanemu in preišljenemu oblikovanju in zasnovi obeh konceptov za portal SISTORY pa seveda obstajajo določene pomanjkljivosti, ki jih imajo razvijalci portala namen popraviti in izboljšati v novi različici SISTORY leta 2018 (SISTORY 4.0), ki napoveduje prehod na metapodatkovno shemo METS, ki bo omogočala še bolj podrobno opisovanje, saj se je v tem času izkazalo, da zaradi novega gradiva potrebujejo izboljšano metapodatkovno shemo, pri kateri bo potreben bolj strokoven pristop k celotnemu procesu.

6. LITERATURA

1. Chan, L.M., Lei Zeng, M. (2006). *Metadata Interoperability and Standardization - A Study of Methodology Part 1*. Pridobljeno s:
<http://www.dlib.org/dlib/june06/chan/06chan.html>
2. Heery, R, Patel, M. (2000). *Application profiles: Mixing and Matching Metadata Schemas*. Pridobljeno s: <http://www.ariadne.ac.uk/issue25/app-profiles>
3. *Heritage of People's Europe*.(2013). Pridobljeno s: <http://www.peoplesheritage.eu/>
4. Lei Zeng, M., Qin, J.(2008). *Metadata*. London
5. Lemmens et al.(2011).*HOPE - Heritage of the People's Europe - The Common HOPE Metadata Structure, including the Harmonisation Specification*.
6. Pančur, A. (2013). *Metapodatki portala Zgodovina Slovenije – Sistory - Navodila za uporabo orodja za vnos metapodatkov*. Pridobljeno s:
<http://www.sistory.si/cdn/publikacije/36001-37000/36151/index.html>
7. St. Pierre, M., LaPlant, W.P. (1998). *Issues in Crosswalking Content Metadata Standards*. Pridobljeno s: http://www.niso.org/publications/white_papers/crosswalk/
8. Vilar, P.(2015/2016). *Metapodatki - Predavanje za predmet Organizacijo informacijskih zbirk 2015/2016*.
9. Vilar, P.(2015/2016). *Prehodi med metapodatki - Predavanje za predmet Organizacijo informacijskih zbirk 2015/2016*
10. W3C. (2016). *XML- eXtensible Markup Language*. Pridobljeno s:
<https://www.w3.org/XML/>
11. W3C. (2015). *RDFa.- Resource Description Framework in Attributes*. Pridobljeno s:
<https://www.w3.org/TR/rdfa-syntax/>
12. *Zgodovina Slovenije - Sistory.si*.(2016). Pridobljeno s: www.sistory.si

7. PRILOGE

Priloge na CD-ROM-u:

Priloga 1: Aplikacijski profil Sistory

Priloga 2: Semantična preslikava elementov

Priloga 3: Sistory Basic XML 2.0

Priloga 4: XML-zapis Anton Aškerc (MGML)

Priloga 5: Osnovni in vsi metapodatki Anton Aškerc (MGML)

Priloga 6: RDFa-zapis Anton Aškerc (MGML)

Priloga 7: Seznam kratic

- AP – aplikacijski profil
- DARIAH-SI – Digital Research Infrastructure for the Arts and Humanities
- DC – Dublin Core
- DCMES – Dublin Core Metadata Element Set
- DCMI – Dublin Core Metadata Initiative
- DCTERMS – DCMI Metadata Terms
- EAD – Encoded Archival Description
- HOPE – Heritage of People's Europe
- HTML – Hyper Text Markup Language
- ID – Identifier
- ISBD – International standard bibliographic description
- LCSH – Library of Congress Subject Heading
- LIDO – Lightweight Information Describing Objects
- MARC – MACHine Readable Cataloging
- METS – Metadata Encoding and Transmission Standard
- MGML – Muzeji in Galerije Mesta Ljubljane
- MySQL – My Structured Query Language
- PREMIS – Preservation Metadata: Implementation Strategies
- RDF – Resource Description Framework
- RDFa – Resource Description Framework in Attributes
- SGML – Standard Generalized Markup Language
- SPECTRUM – Standard ProcEdures for CollecTions Recording Used in Museums
- TEI – Text Encoding Initiative
- UDC – Universal Decimal Classification
- VRA Core – Visual Resources Association
- W3CDTF – Word Wide Web Date and Time Formats
- XML – eXtensible Markup Language

Priloga 8: Izjava o avtorstvu

IZJAVA

Izjavljam, da je diplomski seminar v celoti moje avtorsko delo ter da so uporabljeni viri in literatura navedeni v skladu z mednarodnimi standardi in veljavno zakonodajo.

Katja Meden

V Ljubljani, 25. avgusta 2017

Priloga 9: Dovoljenje za objavo

IZJAVA

Dovoljujem, da Oddelek za bibliotekarstvo, informacijsko znanost in knjigarstvo Filozofske fakultete Univerze v Ljubljani moj diplomski seminar objavi v digitalni zbirki diplomskih del.

Katja Meden

V Ljubljani, 25. avgusta 2017