

Digitalni master dokument: novi original

OLIVERA PORUBOVIĆ-VIDOVIĆ, Ph.D.

head of IT department, Archives of Yugoslavia, Vase Pelagića 33, Belgrade, Serbia
e-mail: olivera@arhivyu.rs

Digital Master Document: New Original

ABSTRACT

The ultimate objective of every digitization project of archival holdings is to create digital master document - copy that would facilitate searching through archival holdings and eliminate potential damage risks during consulting original documents in archives' reading rooms as well as during handling and using originals while preparing publications and exhibitions of archival documents kept in archival institutions. Therefore the creation of a digital master document implies previous defining of adequate quality parameters and later verifying, by means of careful and rigorous control, whether defined quality has been achieved. The paper deals with digital document quality control as an important element in the chain of creating, managing and long term preservation of e-documents created by digitization, stating that the only worth keeping digital master documents are those that passed through thorough quality control.

Key words: digital master copy, quality control, digitization, quality, sampling

Il documento master: un nuovo originale

SINTESI

L'obiettivo finale di ogni progetto di digitalizzazione dei documenti è quello di creare un documento master digitale - una copia con cui effettuare ricerche nei fondi ed eliminare le operazioni con potenziali rischi di danneggiamento: la necessità di consultazione dei documenti originali nelle sale di lettura degli archivi e la necessità di manipolare ed utilizzare gli originali nella preparazione di pubblicazioni e mostre di documenti d'archivio conservati in istituzioni archivistiche. Quindi, la creazione di un documento master digitale implica la precedente definizione della sua qualità adeguata e la successiva verifica, mediante un controllo attento e rigoroso, del raggiungimento della qualità definitiva. Il presente articolo tratta del controllo di qualità del documento digitale inteso come un elemento importante nella catena di creazione, gestione e conservazione a lungo termine dei documenti elettronici creati dalla digitalizzazione, affermando che solamente vale la pena di conservare quei documenti master digitali che hanno attraversato un accurato controllo di qualità.

Parole chiave: documento master, controllo di qualità, digitalizzazione, qualità

Digitalna master kopija: novi original

IZVLEČEK

Končni cilj vsakega projekta digitalizacije arhivskega gradiva, je ustvariti digitalni master dokument - kopijo, ki bi olajšala iskanje po arhivskem gradivu in odpravljala potencialna tveganja pred poškodbami med uporabo izvirnega gradiva v čitalnici arhiva, kot tudi med rokovanjem in uporabo izvornikov pri pripravi publikacij in razstav arhivskih dokumentov, ki se hranijo v arhivih. Zato je pri ustvarjanju digitalne master kopije potrebno najprej definirati ustrezne parametre kakovosti in kasneje s skrbnim in strogim nadzorom preverjati, ali je bila definirana kakovost tudi dosežena. Članek se ukvarja s kontrolo kakovosti digitalnih dokumentov, kot pomembnim elementom v verigi ustvarjanja, upravljanja in dolgoročne hrambe elektronskih dokumentov, ki so nastali z digitalizacijo. Avtorica izpostavlja, da je vredno hraniti le tiste digitalne master kopije, ki so prestale temeljito kontrolo kakovosti.

Ključne besede: digitalna master kopija, kontrola kakovosti, digitalizacija, kakovost, vzorčenje

Digitalni master dokument: novi original

APSTRAKTUM

Krajnji cilj svakog projekta digitalizacije arhivske građe je kreiranje digitalnog master dokumenta - kopije koja će olakšati pretraživanje arhivske građe i eliminisati rizik oštećenja koji uvek postoji kod korišćenja originala kako u čitaonici arhiva, tako i tokom pripreme publikacija i izložbi arhivskih dokumenata. Stvaranje digitalne master kopije podrazumeva prethodno definisanje parametara kvaliteta, a zatim i pažljivu i rigoroznu kontrolu da li je definisani kvalitet postignut. Ovaj rad se bavi kontrolom digitalnog dokumenta kao važnim elementom u lancu stvaranja, upravljanja i dugoročnog čuvanja elektronskih dokumenata, nastalih kao produkt digitalizacije, i ističe da je vredan trajnog čuvanja samo onaj digitalni dokument koji je prošao temeljnu kontrolu kvaliteta.

Ključne reči: master digitalna kopija, kontrola kvaliteta, digitalizacija, kvalitet, uzorkovanje

1 Uvod

Ogroman broj digitalnih kopija arhivskih dokumenata dostupnih na sajtovima arhiva, potvrđuje da su projekti digitalizacije, odnosno planirane i organizovane aktivnosti na digitalizaciji arhivskih dokumenata, uzeli primat nad ranije sporadičnim i ad hoc sprovedenim digitalizacijama, uglavnom za potrebe pripreme izložbi i/ili publikovanja arhivskih dokumenata, ili kao odgovor na povremene zahteve istraživača za dobijanje digitalnih kopija pojedinih arhivskih dokumenata.

Digitalne kopije koje su dostupne korisnicima, kako putem interneta - na sajtu nekog arhiva, tako i putem intraneta u čitaonici arhiva, kreirane su od master digitalnih dokumenata, tj. od neposrednog produkta nekog projekta digitalizacije. Činjenica da se od digitalnog master dokumenta kreiraju sve druge digitalne kopije koje se koriste za različite namene, nameće potrebu da se, prilikom transformacije u digitalno okruženje informacije koju nosi analogni original¹, zadrže svi bitni elementi originala, odnosno da digitalni master dokument bude njegova verna² kopija, i da se kao takva može koristiti isto kao što se koristi original. Dodatno, budući da je namenjen za trajno čuvanje, i sam digitalni master dokument mora biti dovoljno kvalitetan³, pa se tokom sprovođenja projekata digitalizacije posvećuje posebna pažnja kako bi se potreban kvalitet postigao i kako bi u krajnjoj liniji određena digitalna master kopija bila vredna trajnog čuvanja.

Kontrola kvaliteta je važan element svakog projekta digitalizacije. To je skup zahtevnih aktivnosti, koje se sprovode tokom celog životnog ciklusa jednog projekta (FADGI, *Project Planning and Management Outline*, 2009) i kojim se u utiče i meri uspešnost projekta digitalizacije. Težište rada je na kontroli kvaliteta digitalnog master dokumenta, može se reći najdelikatnijoj u nizu kontrola, budući da se tada donosi finalna odluka da li je proizvod digitalizacije zadovoljio zadate parametre kvaliteta. Bitan element ove kontrole je način izbora digitalnih dokumenata koji se kontrolišu, pa je u radu dat nešto širi osvrt na mogući način kreiranja uzorka za kontrolu kvaliteta. Pomenuta kontrola se obavlja u tzv. produkcionoj fazi projekta digitalizacije, jednoj od četiri (planiranje, priprema, produkcija i postprodukcija) kako se u ovom radu posmatra ceo projekt digitalizacije.

2 Kontrola kvaliteta

Kontrola kvaliteta počinje već u fazi planiranja projekta kada se, shodno arhivskoj građi koja je predmet digitalizacije, definišu karakteristike kvaliteta digitalnih master dokumenata koje se žele postići i ustanovljava vrsta, obim, dinamika vršenja kontrole i neophodna prateća dokumentacija. U

1. U ovom radu se kao originalni dokument isključivo posmatra papirni dokument, ne razmatra se originalna arhivska građa na drugim medijumima.

2. Kopija poseduje svojstvo "veran" ukoliko omogućava da se rekonstruišu sve informacije koje pri različitim upotrebama pruža i original. (Ecrire un cahier des charges de numerisation, DAF, 2008).

3. Ne postoji jedinstvena definicija kvaliteta, a za objašnjenje šta se u ovom radu podrazumeva pod terminom "kvalitet" može poslužiti definicija data u standardu ISO 9000 (2005): "Kvalitet je nivo do kojeg skup svojstvenih karakteristika ispunjava zahteve". U standardu se navodi da "svojstven", za razliku od "pridodat", znači da postoji u nečemu, naročito kao trajna karakteristika, kao i da se termin "kvalitet" može koristiti sa pridevima kao što su nedovoljan, dobar ili izvrstan.

pripremnj fazi projekta, kada je već urađena selekcija arhivske građe, vrši se kontrola vrste, redosleda i količine originala, utvrđuje se da li za izabrana dokumenta već postoje opisni metapodaci (La numérisation des documents, 2012), radi se uputstvo za skeniranje. Sprovođenje probnog skeniranja u pripremnj fazi, i kontrola na taj način dobijenih digitalnih kopija, može značajno pomoći u uobličavanju i bližem definisanju kontrole u produkcionj fazi. Produkcionu fazu karakterišu kontrola količine dokumenata u svakoj isporuci za skeniranje; povremene rutinske kontrole skeniranja, što uključuje kontrolu manipulacije originalima, sprovođenja kalibrisanja i čišćenja uređaja, obezbeđivanja adekvatnog osvetljenja. U ovoj fazi vrši se i provera ažurnosti vođenja propisanih evidencija, uključujući i unos tehničkih metapodataka. Nakon skeniranja sledi kontrola redosleda i količine vraćenih originala; kontrola količine preuzetih digitalnih dokumenata i uparenosti broja originala i digitalnih dokumenata, kontrola kvaliteta digitalnih master kopija, kao i kontrola postojanja, tačnosti i konzistentnosti svih predviđenih metapodataka. U postprodukcione kontrole spadaju kontrola sigurnosnog kopiranja digitalnih kopija na odgovarajuće servere, kreiranje dodatnih digitalnih kopija i provera njihovog kvaliteta i redovne periodične provere integriteta medijuma na kome su sve digitalne kopije smeštene⁴.

2.1 Priprema digitalizacije

Adekvatno organizovana, sprovedena i kontrolisana priprema za skeniranje preduslov je dobijanja digitalnog master dokumenta potrebnog kvaliteta. Okosnica pripreme su izabrani originali, budući da je, kako ističe Franziska Frey, (2000), “digitalni master dokument vezan za original od koga je stvoren i za čije predstavljanje je namenjen”. Karakteristike originala, npr. vrsta dokumenta, dimenzije, čitljivost, postojanje oštećenja⁵ itd., direktno utiču na njegov kvalitet. Stoga bliži opis originala, tj. koncizan pregled vrste i stanja selektovane arhivske građe za digitalizaciju, koji se radi u fazi pripreme za skeniranje, s jedne strane pomaže za definisanje elemenata kvaliteta digitalnog master dokumenta koji će se tražiti (očekivati) u predmetnom projektu digitalizacije, a sa druge obezbeđuje bolju pripremu i odvijanje skeniranja.

Tokom pripremnj faze izrađuje se uputstvo za skeniranje, a često se sprovodi i probno skeniranje manjeg skupa raznorodnih originala. Probni skeniranjem simuliraju se produkcioni uslovi što omogućava ispitivanje obuhvatnosti uputstva za skeniranje.

2.1.1 Uputstvo za skeniranje

U uputstvu za skeniranje se navode “pravila” i parametri skeniranja, anticipiraju mogući posebni slučajevi koji se mogu desiti tokom skeniranja i daju predlozi za njihovo razrešenje. Na taj način se smanjuje rizik kreiranja nekvalitetne master kopije. Između ostalog, operateru se navode elementi procedure u skeniranju koje ne sme da zanemari npr. naglašava se obaveza kalibrisanja opreme⁶, korišćenja stakla skenera kojim se pritiska dokument, da ne bi došlo do izvijanja dokumenta, podizanja ivica, stvaranja dodatne senke. Traži se obezbeđivanje prave orijentacije za čitanje odnosno pregledanje dokumenta; zahteva se održavanje originalnog redosleda arhivskih dokumenata; ne dozvoljava se nikakvo doterivanje dobijene digitalne kopije, jer svaka akcija, makar i minimalna je “smanjivanje kvaliteta” - odnosno gubitak postojećih informacija (Ecrire un cahier des charges de numérisation - Annexe 1, DAF); traži se ostavljanje uske margine oko skeniranog dokumenta (kao potvrda da je skeniran ceo dokument i da nije došlo do odsecanja), itd.

U uputstvu se navodi i način imenovanja pojedinačnih dokumenata i grupe dokumenata. (Ecrire un cahier des charges de numérisation du patrimoine, DAF, 2008)

4. Kontrola postojanja, tačnosti i konzistentnosti metapodataka, kalibrisanja opreme, sigurnosnog kopiranja svih digitalnih kopija na odgovarajuće servere i provera integriteta medijuma na kome su digitalne kopije smeštene nisu predmet ovog rada.

5. “Kada je originalni dokument veoma lošeg fizičkog kvaliteta, tada će i digitalna kopija takođe biti lošeg kvaliteta. Kopije koje će se generisati od takvih originala moraju biti identifikovane” (Digitization Quality Assurance - Information Management, 2013).

6. Kalibracija opreme koja se koristi pri skeniranju u projektu digitalizacije utiče na kvalitet digitalnog dokumenta. Ona se mora sprovoditi, a uz to, neophodno je koristiti testne šeme za boju, tonalitet i rezoluciju, jer se samo na taj način garantuje dugotrajno korišćenje digitalnih dokumenata i isplativost investicija uloženi u digitalizaciju. (Frey, 2000).

2.1.2 Probno skeniranje

Sprovođenje probnog skeniranja manjeg skupa arhivskih dokumenata, različitih po dimenzijama, tipu, vrsti otiska, kvalitetu papira, opštem fizičkom stanju, može pomoći da se preciznije odrede elementi potrebnog kvaliteta digitalnog master dokumenta. (Digitization Quality Assurance - Information Management, 2013 i La numérisation des documents, 2012).

Ovo skeniranje omogućava obavljanje kontrole ne samo kvaliteta dobijenih digitalnih kopija u uslovima koji odgovaraju produkcijom, već i kontrolu opreme, softvera, radnog okruženja, načina manipulacije, brzine realizacije, pa i kontrolu predviđenog načina kontrole. Probno skeniranje je i prilika da se operateru koji obavlja skeniranje predoče “kvalitetne” i “nekvalitetne” digitalne kopije⁷, da mu se pojasni način inicijalne kontrole, da se proveru u kojoj meri je već kreirano uputstvo za skeniranje jasno i lako primenjivo, da se oceni da li oprema za skeniranje odgovara potrebama i da li je urađena dobra kalibracija.

2.2 Produkciona faza

U produkcionoj fazi projekta digitalizacije radi se kontrola usaglašenosti količine i redosleda originalnih dokumenata pre isporuke za skeniranje i nakon skeniranja, čime se utvrđuje da li je broj i redosled arhivskih dokumenata ostao nepromenjen, i kontrola količine preuzetih digitalnih dokumenata i uparenosti njihovog broja sa brojem originala, čime se potvrđuje celovitost skeniranja pojedine isporuke. (La numérisation des documents, 2012)

Kontrola kvaliteta digitalnog master dokumenta, koja se takođe odvija u ovoj fazi, može se razmatrati sa dva aspekta. Jedan se odnosi na obim digitalizacije i broj isporuka (preuzimanja) digitalnih kopija, a drugi se tiče ispunjenosti definisanih karakteristika digitalnog dokumenta koje govore o njegovom kvalitetu.

2.2.1 Obim dokumenata za kontrolu

Shodno broju digitalnih dokumenata koje treba pregledati, tj. u zavisnosti od ukupnog obima digitalizacije, kontrola se može obaviti na dva načina. Prvi način podrazumeva pregled svih dobijenih digitalnih dokumenata, a najčešće je to slučaj kada je ukupan broj digitalnih kopija manji od 1000 ($N \leq 1000$) ili kada svojim specifičnim karakteristikama to nameće sama arhivska građa, a drugi način je pregled samo dela dokumenata, tj. uzorka od n dokumenata. U drugom slučaju, u zavisnosti od ishoda kontrole uzorka, prihvataju se ili odbijaju sve kopije čiji je uzorak bio predmet kontrole. Kriterijum za prihvatanje/odbijanje svih digitalnih kopija N (tj. cele populacije) je definisan maksimalnim dozvoljenim brojem neodgovarajućih stavki c u uzorku veličine n . (Nikolaidis i Nenes, 2008)⁸.

Par vrednosti koji sačinjavaju veličina uzorka i broj neadekvatnih kopija koji omogućava prihvatanje uzorka (n, c), poznat je i kao plan uzorkovanja. Navedene vrednosti preporučuju odgovarajući međunarodni standardi kao što su ISO / TR 8550-1 (2007) i ISO / TR 8550-3 (2007), i međunarodni standardi iz porodice ISO 2859 koji uključuje standarde ISO 2859-10 (2006), 2859-1 (1999), ISO 2859-2 (1985), ISO 2859-3 (2005), ISO 2859-4 (2002) i ISO 2859-5 (2005), pogodne za različite slučajeve uzorkovanja. Pored ovih standarda, zbog izuzetne praktične primenjivosti i tehnički izveštaj ANSI/AIIM TR34 (1996) može poslužiti kao odličan vodič za uzorkovanje i uspešnu kontrolu kvaliteta skeniranih dokumenata.

Metode za kontrolu kvaliteta kojim se određuje veličina uzorka n i prihvatljiv broj neadekvatnih kopija c se nazivaju metode uzorkovanja. Osnovna ideja metoda uzorkovanja je da se u uzorku n pronađe broj neadekvatnih kopija X , i da se u slučaju kada je $X > c$ čitava isporuka odbaci, odnosno da se prihvati ukoliko je $X \leq c$. Takođe, podrazumeva se prethodni dogovor između onoga ko skenira arhivska dokumenta (proizvođač) i onoga ko skeniranje inicira (potrošač) oko prihvatljivog i neprihvatljivi-

7. Na ovaj način će se operateru praktično pojasniti kakav kvalitet se traži i koji od dobijenih digitalnih dokumenata ne bi prošao finalnu kontrolu tj. za koje dokumente bi se zbog nezadovoljavajućeg kvaliteta njihove digitalne kopije tražilo ponovno skeniranje.

8. U ovom radu, maksimalni broj neodgovarajućih stavki je maksimalan broj digitalnih kopija nezadovoljavajućeg kvaliteta.

vog nivoa kvaliteta skeniranih dokumenata. U tom smislu definišu se prihvatljivi nivo kvaliteta (**AQL** - Acceptable Quality Level) tj. nivo kvaliteta koji se želi što je onda i nivo kvaliteta koji proizvođač nastoji da ostvari (Krajewski et al., 2009), odnosno neprihvatljivi nivo kvaliteta (**UQL** - Unacceptable Quality Level), što je donji nivo kvaliteta skeniranih dokumenata, tj. najniži nivo koji potrošač još uvek toleriše⁹.

Za razliku od kontrole celokupne količine, korišćenje uzorka implicira rizik da se napravi greška u odlučivanju. Naime, imajući nakon kontrole uzorka digitalnih dokumenata saznanja samo o dokumentima u uzorku veličine **n**, ali ne i u celoj populaciji **N** (u celoj isporuci), neznajući dakle ništa o dokumentima van uzorka, može se desiti da se na osnovu pregledanih digitalnih dokumenata odbaci čitava populacija **N** (cela isporuka), iako bi u stvari trebalo da se prihvati. Ovaj rizik je poznat kao rizik proizvođača **PR** ili Tip greške I (obično se označava kao α). Iako su tipične vrednosti za ovu vrstu rizika između 1% i 20%, najčešće se rizik proizvođača postavlja na 5%, tj. 0.05. Slično, postoji i rizik da će na osnovu pregledanog uzorka veličine **n** populacija **N** (cela isporuka) biti prihvaćena, iako je njen kvalitet loš i trebalo bi je odbaciti. Ovaj rizik je poznat kao rizik potrošača **CR** ili Tip greške II (obično se označava kao β). Tipične vrednosti za ovaj tip rizika su takođe između 1% i 20%, ali uobičajena vrednost sa kojom se operiše je 10%, tj. 0.1. Ukoliko su za neku grupu dokumenata ovi rizici neprihvatljivi, tada jedino zadovoljavajuće rešenje može biti kontrola cele populacije **N**.

Nakon što su prihvatljiv i neprihvatljiv nivo kvaliteta skeniranih dokumenata, **AQL** i **UQL** respektivno dogovoreni, kao i rizici **PR** i **CR**, primenom nekog od raspoloživih pristupa može se definisati plan uzorkovanja (**n**, **c**). U tu svrhu, kako je već ranije i istaknuto, može koristiti tabelarni proračun (Tabela 1) prezentovan u tehničkom izveštaju ANSI / AIIM TR34 (1996), na osnovu vrednosti koje je originalno utvrdio Grubbs (1949). Vrednosti u tabeli proračunate su pod pretpostavkom da je rizik proizvođača **PR=5%**, a rizik potrošača **CR=10%**, dakle imaju tipične vrednosti koje se navode u standardima.

Tabela 1: vrednosti potrebne za sprovođenje tabelarnog proračuna

VREDNOSTI POTREBNE ZA SPROVOĐENJE TABELARNOG PRORAČUNA																
c	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n x UQL	2.303	3.890	5.322	6.681	7.994	9.274	10.532	11.771	12.995	14.207	15.407	16.598	17.782	18.958	20.128	21.292
n x AQL	0.0513	0.355	0.818	1.366	1.970	2.613	3.286	3.981	4.695	5.426	6.169	6.924	7.690	8.464	9.246	10.035
UQL/AQL	44.84	10.96	6.51	4.89	4.06	3.55	3.21	2.96	2.77	2.62	2.50	2.40	2.31	2.24	2.18	2.12

Interesantno je videti kako promena vrednosti prihvatljivog i neprihvatljivog nivoa kvaliteta skeniranih dokumenata (**AQL** i **UQL**) utiče na plan uzorkovanja, a to je u nastavku prikazano na dva numerička primera. Vrednosti rizika su u oba slučaja iste, tj. **PR=0.05**, **CR=0.1**, budući da je pomenu-ta tabela kreirana upravo za te vrednosti.

- **korak A** - Definisane ulaznih vrednosti

PLAN UZORKOVANJA 1: **AQL=0.01**, **UQL=0.02**, **PR=0.05**, **CR=0.1**

PLAN UZORKOVANJA 2: **AQL=0.02**, **UQL=0.06**, **PR=0.05**, **CR=0.1**

- **korak B** - Određivanje odnosa **UQL/AQL**

PLAN UZORKOVANJA 1: **UQL/AQL=2.0**

PLAN UZORKOVANJA 2: **UQL/AQL=3.0**

9. **AQL** i **UQL** se ugovaraju i za jednokratne i za višestruke isporuke.

- **korak C** - Pronalaženje najbliže vrednosti odnosa UQL/AQL u Tabeli 1

PLAN UZORKOVANJA 1

Najbliža vrednost dobijenoj vrednosti 2.0 je 2.12

PLAN UZORKOVANJA 2

Najbliža vrednost dobijenoj vrednosti 3.0 je 2.96

- **korak D** - Identifikovanje prihvatljivog broja neadekvatnih kopija

PLAN UZORKOVANJA 1: **c=15**

(vrednost 2.12 pronađena u prethodnom koraku nalazi se u koloni u kojoj je c=15)

PLAN UZORKOVANJA 2: **c=7**

(vrednost 2.96 pronađena u prethodnom koraku nalazi se u koloni u kojoj je c=7)

- **korak E** - Određivanje veličine uzorka
(deljenjem odgovarajućih vrednosti iz tabele i zaokruživanjem vrednosti)

PLAN UZORKOVANJA 1: $n \times UQL / UQL = 21.292 / 0.02 = 1064.6$, nakon zaokruživanja 1065
 $n \times AQL / AQL = 10.035 / 0.01 = 1003.50$, nakon zaokruživanja 1004

PLAN UZORKOVANJA 2: $n \times UQL / UQL = 11.771 / 0.06 = 196.18$, nakon zaokruživanja 197
 $n \times AQL / AQL = 3.981 / 0.02 = 199.05$, nakon zaokruživanja 200

korak F - Definisane plana uzorkovanja
(korišćenjem dobijenog prihvatljivog broja neadekvatnih kopija i većeg broja od dva dobijena za veličinu uzorka)

PLAN UZORKOVANJA 1: $c=3$, $n = \max(1065, 1004) = 1065$

PLAN UZORKOVANJA 2: $c=7$, $n = \max(197, 200) = 200$

Iz navedenih primera vidi se da je određivanje plana uzorkovanja prilično jednostavno, kao i da se sa strožijim kriterijumima zahtevanog kvaliteta dramatično povećava potrebna veličina uzorka. U prvom slučaju uzorkovanja, kada je nivo prihvatljivog kvaliteta 1%, što znači da se ne očekuje više od jednog neodgovarajućeg dokumenta na 100 skeniranih, a nivo neprihvatljivog kvaliteta i iznosi 2% potreban je uzorak veličine preko 1000 dokumenata¹⁰. U drugom slučaju, kada je nivo prihvatljivog kvaliteta 2%, a nivo neprihvatljivog kvaliteta iznosi 6% potreban je uzorak veličine 200 dokumenata.

Potrebno je naglasiti da će, uz dobro definisanu veličinu uzorka, kontrola kvaliteta na osnovu uzorka dati validne rezultate samo u slučaju kada se koristi tzv. reprezentativan uzorak, tj. onaj koji odražava čitavu populaciju. Ukoliko je populacija homogena po karakteristikama, reprezentativni uzorak se može dobiti slučajnim izborom, a apsolutna slučajnost obezbeđuje se korišćenjem generatora slučajnih brojeva. Kod nehomogenih populacija formira se tzv. stratifikovan uzorak, odnosno uzorak koga čini više slojeva (stratuma) homogenih uzoraka.

Izložen način uzorkovanja odgovara slučajevima kada u projektu digitalizacije ima jedna ili nekoliko isporuka digitalnih dokumenata i služi kao ilustracija sa jedne strane pomoći koju uspostavljeni standardi pružaju u kreiranju uzorka za pregled, a sa druge načina kontrolisanja rizika greške koje će se neizbežno javiti pri kontroli korišćenjem uzoraka. U tehničkom izveštaju ANSI / AIIM TR34 (1996) i standardima porodice ISO 2859 su obrađena i uzorkovanja za slučaj velikog broja isporuka homogenog materijala (npr. 100 i više) i ukazano je na dodatnu mogućnost definisanja tri stepena "strogosti" kontrole (normalna, slabija, jača) koje se primenjuju u zavisnosti rezultata kontrole više uzastopnih prethodnih isporuka. Kontrola prve isporuke je uvek normalna sve dok se ne ispune uslovi da se pređe na drugi režim: slabija ili jača kontrola. Na primer, ukoliko je određen broj sukcesivnih isporuka zadovoljio uslove kontrole (nema odbijanja isporuke) nadalje se, do pojave unapred dogovorenog broja sukcesivno odbačenih isporuka, radi blaža kontrola. Ukoliko se tokom blaže kontrole sukcesivno pojavi unapred definisan nedozvoljen broj odbačenih isporuka, kontrola se vraća na tzv. "normalnu".

Posmatranjem rizika koji postoje kod kontrole kvaliteta na osnovu uzorka, može se pomisliti da je kontrola čitave populacije sigurna 100% i da nakon ovako izvršene kontrole nema sačuvanih digitalnih master dokumenata sa greškom. Praksa međutim to demantuje. Naime, kontrola velikog broja

10. Ukoliko se proračunom dobije uzorak veći od čitave populacije, tada se vrši kontrola čitave populacije.

dokumenata je zamorna i dugotrajna i upravo je umor generator nastanka grešaka. Na ovu činjenicu se ukazuje u dokumentaciji o završenim projektima digitalizacije, pa se, u cilju sprečavanja i/ili smanjivanja procenta grešaka, savetuje da se u toku jednog radnog dana na kontroli ne radi puno radno vreme već maksimalno 4 sata (Rahel Bacher et al., 2011). Ograničavanjem dnevnog angažovanja na kontroli dodatno se usporava poželjan ritam odvijanja projekta i produžava njegovo trajanje.

Razmatranje kontrole metodom uzorkovanja odnosi se i na slučaj kada arhiv angažuje specijalizovano preduzeće da obavi skeniranje arhivske građe i u slučaju kada se cela digitalizacija radi korišćenjem sopstvenih kadrovskih i tehničkih resursa arhiva. U dostupnoj literaturi nailazi se na podatak da je digitalizacija u kući dobra za mali obim arhivske građe, do 1000 dokumenata. Ipak, moguće je sopstvenim resursima arhiva ostvariti digitalizaciju većeg obima, ukoliko adekvatni resursi postoje, kada najčešće postoji i posebna organizaciona celina koja je zadužena za skeniranje. U tom slučaju, jedna organizaciona celina arhiva radi skeniranje, a druga ga inicira i kontroliše¹¹, pa se jedna od njih može posmatrati kao proizvođač, a druga kao potrošač i odnosi između njih, odnosno definisanje rizika, mogu biti ustanovljeni na način kako je to u prethodnom delu teksta izloženo.

2.2.2 Način kontrole

Pregled digitalnih kopija obavlja se automatizovanim procedurama i vizuelno. U arhivima se koriste i gotova rešenja¹² i razvijaju sopstvene automatizovane procedure koje značajno ubrzavaju kontrolu digitalnih dokumenata. Međutim, iako dobro osmišljene automatizovane procedure mogu biti veoma efikasne, praktično je nemoguće da se na taj način uoče (izdvoje) svi mogući "nedostaci" u kvalitetu digitalnog dokumenta, pa je dodatna vizuelna kontrola neophodna. Za nju je pak potrebno obezbediti dobro kalibrisan monitor, softver za pregled digitalnih dokumenata (slika) i adekvatno osvetljenje. Slučaj vizuelnog poređenja originala i njegove digitalne kopije na ekranu računara, obrađuje standard ISO 12646:2008, a standard ISO 3664:2009 daje specifikaciju kalibrisanja monitora za pregled digitalnih kopija bez upoređivanja sa originalom.

Vizuelna kontrola na ekranu se obavlja pri uvećanju digitalne kopije 100%, a u okviru nje može se koristiti i opcija pregleda krive histograma koju uobičajeno sadrže programi za pregled slika.

2.2.3 Karakteristike master dokumenta koji se kontrolišu

Da bi neki digitalni dokument označio kao kvalitetan i prihvatljiv za trajno čuvanje, neophodno je da su zadovoljeni prethodno definisani kriterijumi. Atributi digitalnog dokumenta koji se uobičajeno kontrolišu: oštrina, celovitost i kompletnost, orijentacija, format i kompresija, rezolucija, dubina boja, kolorni mod, imenovanje prikazani su u Tabeli 2.

11. Opšteprihvaćeno je da finalnu kontrolu digitalne kopije ne vrši onaj ko je radio na skeniranju.

12. U radu "Practical quality control procedures for digital imaging projects" (Jenn Riley i Kurt Whits, 2005) ponuđen je alat koji je svoju primenu već našao na Univerzitetu Indijana, u sklopu programa Digitalna biblioteka, koji omogućava prikupljanje podataka iz zaglavlja tif datoteka, formata koji se najčešće koristi za čuvanje master digitalnih kopija arhivske i bibliotečke građe. Pogledati: www.libtiff.org.

Tabela 2: Atributi digitalnog dokumenta i njihova kontrola

ATRIBUTI DIGITALNOG DOKUMENTA I NJIHOVA KONTROLA		
Atribut	Kontrola (komentar)	O b a v e z n o p o n o v n o skeniranje
OŠTRINA	Pregled digitalne kopije na ekranu mora biti rađen pri uvećanju 100%. Manje uvećanje može prouzrokovati pogrešan zaključak da je oštrina zadovoljavajuća.	DA
CELOVITOST	Da li je svaki dokument u potpunosti (ceo) skeniran i kao takav predstavlja "pravu" kopiju originala.	DA
KOMPLETNOST	Da li su svi elementi grupe prisutni (npr: da li su skenirane sve stranice pisma, knjige, novine itd.).	DA
FORMAT I KOMPRESIJA	Da li oznaka ekstenzije dokumenata odgovara stvarnom formatu i da li je poštovan princip da se digitalni master dokument kreira bez primene bilo kog algoritma za kompresiju.	DA
REZOLUCIJA, DUBINA BOJA, KOLORNI MOD	Razmatranje kvaliteta digitalne kopije podrazumeva proveru definisane rezolucije, pri čemu bi, kao kontrolni element opredeljenja za neku rezoluciju, trebalo imati u vidu da pravilo veća rezolucija - bolji kvalitet digitalne kopije važi samo do nekog nivoa, a da dalje povećanje rezolucije samo povećava vreme skeniranja i stvara nepotrebno veće zahteve za potrebnim prostorom na serveru na kome se kopija čuva. (La numérisation des documents, 2012), (Ian Bogus at al., 2013). Kolorni mod RGB sa 24-bitnom dubinom boja uobičajena opcija u digitalizaciji bibliotečkog i arhivskog materijala (Ian Bogus at al., 2013). Neki autori naglašavaju da u projektima digitalizacije unutar ustanova kulture, za ditalna master dokumenta nikako ne treba koristiti CMYK kolorni mod (FADGI, <i>Creation of Raster Image Master Files</i> , 2010).	DA
IMENOVANJE	Provera poštovanja prethodno zadate šeme imenovanja, odnosno strukture lako se može sprovesti korišćenjem automatskih procedura. Uobičajeno, deo šeme čini redni broj digitalnog dokumenta unutar grupe, pa se može lako pripremiti softversko rešenje koje će pored provere šeme moći da identifikuje postojanje preskoka u nizu, ali i da posluži za izračunavanje ukupnog broja dokumenata unutar više grupa.	DA
ARTEFAKTI	Dokumenti koji sadrže artefakte moraju se ponovo skenirati, jer je njihovim postojanjem kvalitet digitalne kopije značajno degradiran. (John McIlwaine at al., 2002).	DA
ORIJENTACIJA	Neusaglašenost orijentacije digitalne kopije i originala ne mora biti razlog za ponovno skeniranje, ali se pri pregledu registruje kao greška, jer je znanje o neadekvatnoj orijentaciji potrebno kod kreiranja kopija za korišćenje.	NE
NEADEKVATNO POSTAVLJANJE ORIGINALA	Ukoliko su originali neznatno iskošeni pre skeniranja i to ne utiče na čitljivost i preglednost, greška se evidentira, ali se skeniranje ne ponavlja.	NE
	Pogrešno postavljanje originalnog dokumenta na skener koje je izazvalo skeniranje kroz poleđinu (dešava se kod transparentnih podloga) zahteva ponovno skeniranje originala.	DA

3 Zaključna razmatranja

Tehnička poboljšanja opreme i stalni razvoj softvera doprinose smanjenju troškova izrade digitalnih kopija, a prednosti njihovog korišćenja u odnosu na korišćenje drugih kopija (mikrofilm i fotokopija) odavno su proverene i dokazane u praksi. Sa porastom masovnosti projekata digitalizacije arhivske građe, samim tim i broja kreiranih digitalnih master kopija, pitanje kontrole njihovog kvaliteta postaje sve važnije, jer digitalne master kopije nastavljaju da "žive" na namenskim serverima i nakon završetka projekta digitalizacije. Imajući u vidu troškove njihovog kreiranja, a zatim i trajnog održavanja, nameće se potreba njihove kontrole odmah nakon nastanka, kako bi se obezbedilo trajno čuvanje samo onih kopija koje to i zaslužuju.

Kontrola svih digitalnih master kopija je u uslovima masovne digitalizacije neisplativa sa stanovišta utrošenog vremena, a i nekonformna jer ju je zbog vremena koje zahteva nemoguće sprovesti u produkcionoj fazi projekta. Zato je uputno razmišljati o kontroli uzorka, a ne čitave populacije digitalnih dokumenata.

Dobru kontrolu digitalne kopije čine precizno definisni i atributi kvalitetnog digitalnog dokumenta i način sprovođenja kontrole, uz odgovarajuću prateću dokumentaciju, pri čemu oslonac svakako treba tražiti u primerima dosadašnje dobre prakse, a prevashodno u ISO i nacionalnim standardima i tehničkim preporukama koji se bave nekim od segmenata kontrole digitalnog dokumenta. Činjenica da su sistemi, procesi, proizvodi i usluge nesavršeni, iznova će problem kontrole činiti aktuelnim, a ovaj rad je prilog nastojanju da od kontrole kvaliteta digitalnog dokumenta, ma kako zahtevna bila, ne treba odustajati, jer samo kvalitetan digitalni master dokument, onaj koji ima sva svojstva originala, treba trajno čuvati.

Literatura

ANSI/AIIM TR34-1996: *Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) and Micrographics Systems*. Dostupno na: <https://law.resource.org/pub/us/cfr/ibr/001/aimm.tr-34.1996.pdf> (posećeno: 14.05.2015).

Bacher, Rahel et al (2011). EUROPEANA REGIA, CIP- ICT- PSP 2009 - 250560, *D5.3 Quality Management*.

Bogus, Ian et al (2013). *Minimum Digitization Capture Recommendations*. The Association for Library Collections and Technical Services Preservation and Reformatting Section. Dostupno na: <http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations> (posećeno: 14.05.2015).

Digitization Quality Assurance - Information Management. Dostupno na: http://www.im.gov.ab.ca/documents/imtopics/RIM0004_Digitization_Quality_Assurance_February_2013.pdf (posećeno: 14.05.2015).

Ecrire un cahier des charges de numérisation du patrimoine- GUIDE TECHNIQUE, documents reliés - manuscrits - plans-dessins - photographies - microformes (2008). Paris : Direction des Archives de France. Dostupno na: <http://www.archivesdefrance.culture.gouv.fr/static/4132> (posećeno: 14.05.2015).

Federal Agencies Digitization Guidelines Initiative - Still Image Working Group (2009). *Digitization activities. Project Planning and Management Outline*. Dostupno na: <http://www.digitizationguidelines.gov/guidelines/DigActivities-FADGI-v1-20091104.pdf> (posećeno: 14.05.2015).

Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group (2010). *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*, Dostupno na: http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf (posećeno: 30.04.2015).

Fernandez, Edmond, Perrot, Patrick. *Ecrire un cahier des charges de numérisation - Recommandations sur les caractéristiques techniques des fichiers images - Annexe 1*. Paris : Direction des Archives de France. Dostupno na : <http://www.archivesdefrance.culture.gouv.fr/static/1309> (posećeno: 14.05.2015).

Frey, Franziska (2000). *Measuring quality of digital masters*. Guides to Quality in Visual Resource Imaging. Council on Library and Information Resources. Dostupno na: <http://www.oclc.org/research/publications/library/visguides/visguide4.html>, (posećeno: 30.04.2015).

Grubbs, F. E. (1949). On designing single sampling inspection plans. *Annals of mathematical statistics*, Vol. XX (1949), pp. 242-256.

Internet 1: <http://www.europeanaregia.eu/sites/www.europeanaregia.eu/files/europeana-regia-digitization-quality-management.pdf> (posećeno: 14.05.2015).

ISO 9000 (2005) *Quality management systems - Fundamentals and vocabulary*. Dostupno na: <https://law.resource.org/pub/in/bis/S07/is.iso.9000.2005.pdf> (posećeno: 30.04.2015).

Krajewski L.J., Ritzman L.P., Malhotra M.K. (2009). *Operations Management: Processes and Supply chains*. Prentice Hall, 9th Edition, 2009 - Supplement G: Acceptance sampling plans. Dostupno na: http://wps.prenhall.com/wps/media/objects/7117/7288732/65767_28_SuppG.pdf (posećeno: 16.05.2015).

La numérisation des documents. Méthodes et recommandations. Ottawa: Bibliothèque et Archives nationales Québec. Dostupno na: http://www.banq.qc.ca/documents/services/archivistique_ged/Numerisation_des_documents.pdf (posećeno: 30.04.2015).

McIlwaine, John et al (2002). *Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives*. Dostupno na: <http://www.ifa.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines.pdf> (posećeno: 14.05.2015).

Nikolaidis, Yiannis, Nenes, George (2008). Economic Evaluation of ISO 2859 Acceptance Sampling Plans Used with Rectifying Inspection of Rejected Lots. *Quality Engineering*, Vol.21, No.1, pp.10-23. DOI:10.1080/08982110802355877.

Riley, Jenn, Whitsel, Kurt (2005). Practical quality control procedures for digital imaging projects, *OCLC Systems & Services: International Digital Library Perspectives* Vol. 21 No. 1, 2005, pp 40-48. Dostupno na: <http://www.dlib.indiana.edu/~jenrile/publications/imageqc/qc.pdf> (posećeno: 14.05.2015).

SUMMARY

Technical improvements of equipment and continuous development of adequate software contribute to the cost reduction of making digital copies, whose advantages compared to the use of other copies (microfilm and photocopies), have been long tested and proven in archives practices all over the world. With the increase of large scale digitization projects of archival holdings and the augmentation of created digital master copies, the question of their quality control is becoming increasingly important. Digital master copies continue "to live" on dedicated servers after the completion of the digitization project. During their "life" they serve as a source for derivative files for a variety of uses: web presentations, publications, exhibitions, database search and so on. Having in mind the costs of the creation and later permanent maintenance of digital master copies, it is necessary to carry out quality control check immediately after their creation in order to permanently preserve those high-quality copies worthy of preservation. In this paper, digitization project is presented through four stages: planning, preparation, production and post-production. In each of them quality control is an important component. In the planning phase, according to the nature of archival holdings which is the subject of digitalization, the quality parameters of the digital master documents are defined, the scope and dynamics of quality control is established, as well as all the necessary documentation. In the preparatory phase of the project, the control of the type, sequence and quantity of the original archival documents is performed and the existing data are copied or descriptive metadata are created and instruction manual for scanning is prepared. Also, at this stage, test scanning can be performed, and the results can be of significant help in the formulation and better definition of the control needed in the production phase. A production phase is characterized by the control of the quantity and sequence of original archival documents, before and after each delivery for scanning; periodic checks of compliance with all previously defined scanning procedures; the quantity control of digital documents is performed, as well as the verification of the existence, accuracy and consistency of all metadata and the quality control of the digital master copies. Post-production controls include creating backup of digital master documents to the appropriate servers, creating additional digital copies as well as checking their quality and regular periodic integrity checks of the medium on which all digital copies are stored. In this paper, a special attention is given to the quality control of digital master document itself. Firstly, the way in which the control is performed is discussed and afterwards, some of the essential characteristics of digital documents for quality assurance are presented. Depending on the volume of scanned pages, quality control can be done in two ways: by checking each page, for the low volume case, or by checking items from the sample. In the former case a lot of scanned pages may be accepted or rejected depending on the outcome of the control and the acceptance-rejection criterion, i.e., the maximum allowable number of defective items in the sample. Based on technical report ANSI / AIIM TR34 (1996) the risks of using samples are presented and it is shown how the sample size increases drastically with the increase of the required quality level. On the other hand, it was pointed out that even the control of all documents is not 100% safe, because it takes a long time, and tired employees working on control make mistakes. As for the essential characteristics of digital documents for quality assurance, a short

list is presented: sharpness, completeness, resolution, bit depth, colour space, orientation, file format, file name and compression. The paper finishes with conclusion that as systems, processes and products are not and could not be perfect the problem of quality control would always be present and that archives should stick to ISO standards and examples of good practice of cultural heritage community in the domain of quality control.

Typology: 1.02 Review Article

Submitting date: 15.03.2015

Acceptance date: 09.04.2015

