Arhiviranje digitalnih zapisa

Zmajlović, Marija

SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI
Ak. god. 2019./ 2020.

Marija Zmajlović

Arhiviranje digitalnih zapisa

Završni rad

Mentor: dr.sc. Hrvoje Stančić, red. prof.

Zagreb, rujan 2020.

Izjava o akademskoj čestitosti

Izjavljujem i svojim potpisom potvrđujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenoj i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

Sadržaj

1. Uvod

U današnje vrijeme za sve se više informacija postavljaju zahtjevi, pretražuju se, prenose i isporučuju u elektroničkom obliku. Sve se te informacije nalaze u elektroničkim zapisima. U usporedbi s analognim zapisima, elektronički zapisi sami su po sebi kompleksni za očuvanje zbog stalnog i brzog napretka informatičke tehnologije, standarda i formata. Takvi se zapisi nazivaju složenima ako imaju dodatne zahtjeve tijekom dugoročnog očuvanja, na primjer očuvanje integriteta, autentičnosti ili pouzdanosti, čime se dodaje dodatna razina složenosti. Dodatna razina složenosti pronalazi se i u nekim dodatnim elementima kao što su elektronički potpisi, digitalni certifikati, vremenski žigovi i slično. Također je vrlo važno odabrati kvalitetni sustav medija za pohranu i standarda metapodataka koji će se koristiti u arhiviranju. Arhiviranje i očuvanje predstavljaju jedinstveni izazov zbog dugoročne prirode ovih aktivnosti. Problem dugoročnog čuvanja i održavanja digitalnih informacija može se protumačiti kao očuvanje zapisa i tehnologija na kojoj se temelje unatoč njihovom zastarijevanju. Digitalni objekti zahtijevaju stalno i kontinuirano održavanje i ovise o složenom ekosistemu hardvera, softvera, standarda i zakonskih propisa koji se stalno mijenjaju.

2. Digitalni zapis

Digitalni zapisi mogu nastati na dva načina – digitalizacijom ili izvorno nastaju u digitalnom obliku (engl. *born digital*). Digitalizacija u širem smislu jest proces transformacije analognog signala u odgovarajući digitalni oblik. „U užem je smislu digitalizacija prelazak različitih materijala u digitalni format pri čemu se oni pretvaraju u binarni kod pohranjen u računalnu datoteku.“[1] Stančić dijeli pojam očuvanja na dva dijela - očuvanje sadržaja ili informacije zapisane u dokumentu te očuvanje fizičkog objekta, to jest medija na kojem je sadržaj pohranjen.[2] Ono što je važno jest da je svaki digitalni zapis vjerodostojan. Svaki digitalni zapis mora biti i autentičan, potpun te mora sačuvati dovoljno konteksta. Potpunost se odnosi na onaj zapis kojemu je pridruženo mjesto i vrijeme nastajanja, naslov, sadržaj, predmet i informacije o korisniku. Autentičan je onaj zapis čija je povijest nastanka, korištenja i prijenosa očuvana, a kontekst se odnosi na veze između zapisa i na okolinu u kojoj je zapis nastao. Digitalni objekti zahtijevaju stalno i kontinuirano održavanje te ovise o kompleksnosti hardvera, softvera, standarda i zakonskih propisa koji se neprestano mijenjaju. U usporedbi s analognim zapisima, digitalni se suočavaju s većim rizikom propadanja, upravo zbog brzine razvoja informacijske tehnologije. Ukoliko je objekt očuvan, očuvan je i zapis na njemu. Najveća razlika između analognog i digitalnog zapisa jest u tome što je digitalni zapis ovisan o tehnologiji i kako bi ga se moglo koristiti nužno je imati odgovarajuću kombinaciju hardvera i softvera. Digitalni zapisi s time prestaju biti fizički objekti i postaju rezultat posredovanja podataka i tehnologije.

Iako je pohrana sve jeftinija, to ne znači da svaka datoteka i svaka njena inačica može i treba biti spremljena. Odluka o tome što treba sačuvati postaje složenija s većom količinom podataka i širim rasponom medija za pohranu. To zauzvrat povećava rizik da se ne sačuvaju predmeti koji će se jednog dana pokazati povijesnom vrijednošću. Postoji i veći rizik nemogućnosti pronalaska podataka zbog loših metapodataka. Velika količina podataka ne može se obraditi pojedinačno, kao što je to slučaj s analognim materijalima. Više podataka neizbježno će značiti i veće pouzdanje u automatizaciju i razvoj novih radnih tokova za obradu podataka. Uz to dolazi i potreba za povećanjem snage i pohrane računala. Jedna od trenutnih činjenica koje se odnose na povećanje količine podataka te na očekivanja vezanih

[1] Stančić, H. New Technologies applicable to Document and Records Management: Blockchain // Lligall. Revista Catalana d'Arxivística. Noves perspectives en matèria de gestió documental, 41 (2018), 56.
[2] Ibid., str. 56.

za pohranu jest da se sustav za pohranu možda neće moći nositi s velikim datotekama, većim brojem verzija istih datoteka ili samo s velikim količinama datoteka. Ovo može zahtijevati novu infrastrukturu i softver.[3]

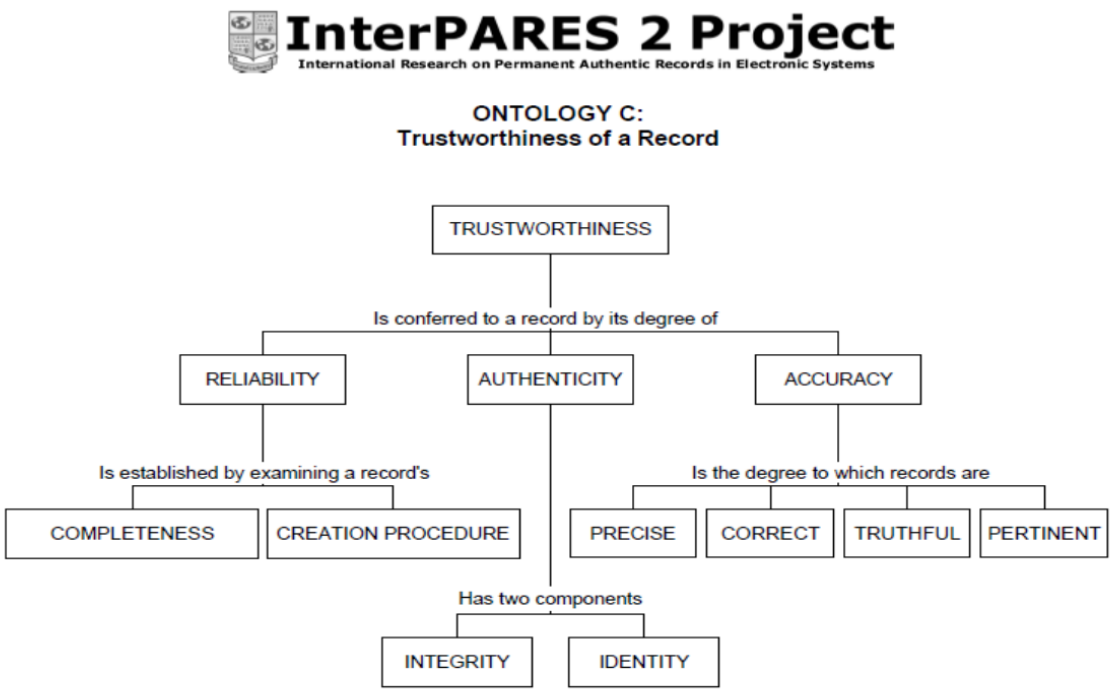
2.1. Vjerodostojnost digitalnog zapisa

„Vjerodostojnost jest cjelina svojstava autentičnosti, pouzdanosti, točnosti, integriteta i upotrebljivosti zapisa.“[4]

- Autentičnost je svojstvo zapisa kojim se potvrđuje da je zapis ono što se tvrdi da jest i da nije bio mijenjan te se najčešće odnosi na unutarnje i vanjske dokaze uzimajući u obzir fizičke značajke, strukturu, sadržaj i kontekst.
- Pouzdanost jest svojstvo zapisa da dokazuje svoju vjerodostojnost, tj. potvrđuje činjenicu o kojoj je u njemu riječ i koja je uspostavljena provjerom cjelovitosti oblika zapisa i razine nadzora tijekom njegova nastanka.
- Točnost je svojstvo onoga što se sastoji od istinitih podataka, koje proizlazi iz pravilnoga postupka; svojstvo onoga što je pravilno postavljeno i izvedeno.
- Integritet je svojstvo čitavosti i neizmijenjenosti; sačuvanost zapisa od gubitka, promjena i oštećivanja.
- Upotrebljivost je svojstvo koje se odnosi na mogućnost korištenja zapisa.

Slika 1 prikazuje koncept vjerodostojnosti i opisanih svojstava.

[3] Houghton, B. Preservation Challenges in the Digital Age. // D-Lib Magazine. 22(7/8), srpanj/kolovoz 2016. Dostupno na: http://www.dlib.org/dlib/july16/houghton/07houghton.html (17.06.2020.).
[4] Mihaljević, M., Mihaljević, M. i Stančić, H. (2015). trustworthiness. U: Arhivistički rječnik: HRVATSKO-ENGLESKI/ENGLESKO-HRVATSKI. Zagreb: Zavod za informacijske studije Odsjeka za informacijske i komunikacijske znanosti Filozofskog fakulteta Sveučilišta u Zagrebu, str. 155.



Slika 1: Vjerodostojnost digitalnog zapisa[5]

Iz slijedećih je razloga važno znati je li neki zapis vjerodostojan:

- „dokaz,
- odgovornost,
- zaštita prava,
- čuvanje identiteta,
- razumijevanje prošlosti,
- oslanjanje na izvore,
- Quod non est in actis, non est in mundo – Ono što nije u zapisu ne postoji.“[6]

Vjerodostojnost je temeljni pojam u arhivskoj znanosti, a izuzetno je važno da se zapisi mogu garantirati vjerodostojnima. Odnos između zapisa, vjerodostojnosti i dokaza predmet je čestih rasprava među arhivskim znanstvenicima pod utjecajem kulturoloških, tehnoloških, pravnih i filozofskih trendova.

[5] InterPARES 2 Project: International Research on Permanent Authentic Records in Electronic Systems. // InterPARES Trust. Dostupno na: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_ontology.pdf (1.09.2020.).
[6] Duranti, L. The Trustworthiness of Digital Records. // Ppt prezentacija : International Congress on Digital Records Preservation, 16. travanj 2010.

3. Mediji za pohranu

Mnogo je načina na koji se elektronička građa pohranjuje i čuva. Izbor sustava za pohranu na dulji period ovisi o nekoliko kriterija:

- „trajnost medija,
- visoki kapacitet,
- dugovječnost medija,
- niska cijena,
- široka prihvaćenost,
- sustav mora biti izravan (engl. *on-line*) ili poluizravan (engl. *near-line*).“[7]

Nekoliko je skupina medija na koje je moguće pohraniti digitalno gradivo i dijele se prema vrsti medija i načinu zapisa:

- „izmjenjivi magnetski diskovi (zastarjeli) – disketa,
- izmjenjivi optički diskovi – CD-ROM, DVD, UDO (engl. *Ultra Density Optical*), BD-Blu-Ray i drugi,
- magnetno-optički diskovi,
- poluvodički mediji – USM prijenosne memorije, SmartMedia, Memory Stick memorijske kartice, CompactFlash,
- magnetske trake – DLT-S4, LTO-4, SAIT-2, T10000,
- optičke trake – TRAAMS, LOTS
- urezivanje ionskom zrakom – HD-Rosetta.“[8]

U procesu arhiviranja potrebno je imati mnogo opcija u odabiru medija za pohranu kako bi se taj posao mogao što kvalitetnije obaviti. Također je potrebno razmišljati hoće li se gradivo pregledavati na zaslonu ili će se ispisivati, hoće li se omogućiti pristup građi lokalno ili putem Interneta i slično. Stoga je posao arhivista situaciju sagledati iz perspektive korisnika kako bi stvorio sustav koji će korisniku biti najpristupačniji.

Postoji sedam vrsta sustava za pohranu i osiguranje dostupnosti digitalne građe putem mreže. „To su 1. izravni (engl. *online*), 2. poluizravni (engl. *near-line*), 3. hijerarhijski (engl. *Hierarchical Storage Management – HSM*), 4. neizravni (engl. *off-line*), 5. mrežna pohrana

[7] Stančić, H. Digitalizacija. Zagreb : Zavod za informacijske studije, 2009. str. 113.
[8] Ibid., str. 114-115.

(engl. *Network Attached Storage - NAS*), 6. mreža za pohranu (*engl. Storage Area Network – SAN*)"[9] i 7. pohrana u oblaku (engl. *Cloud Storage*).

3.1.1. Magnetske trake

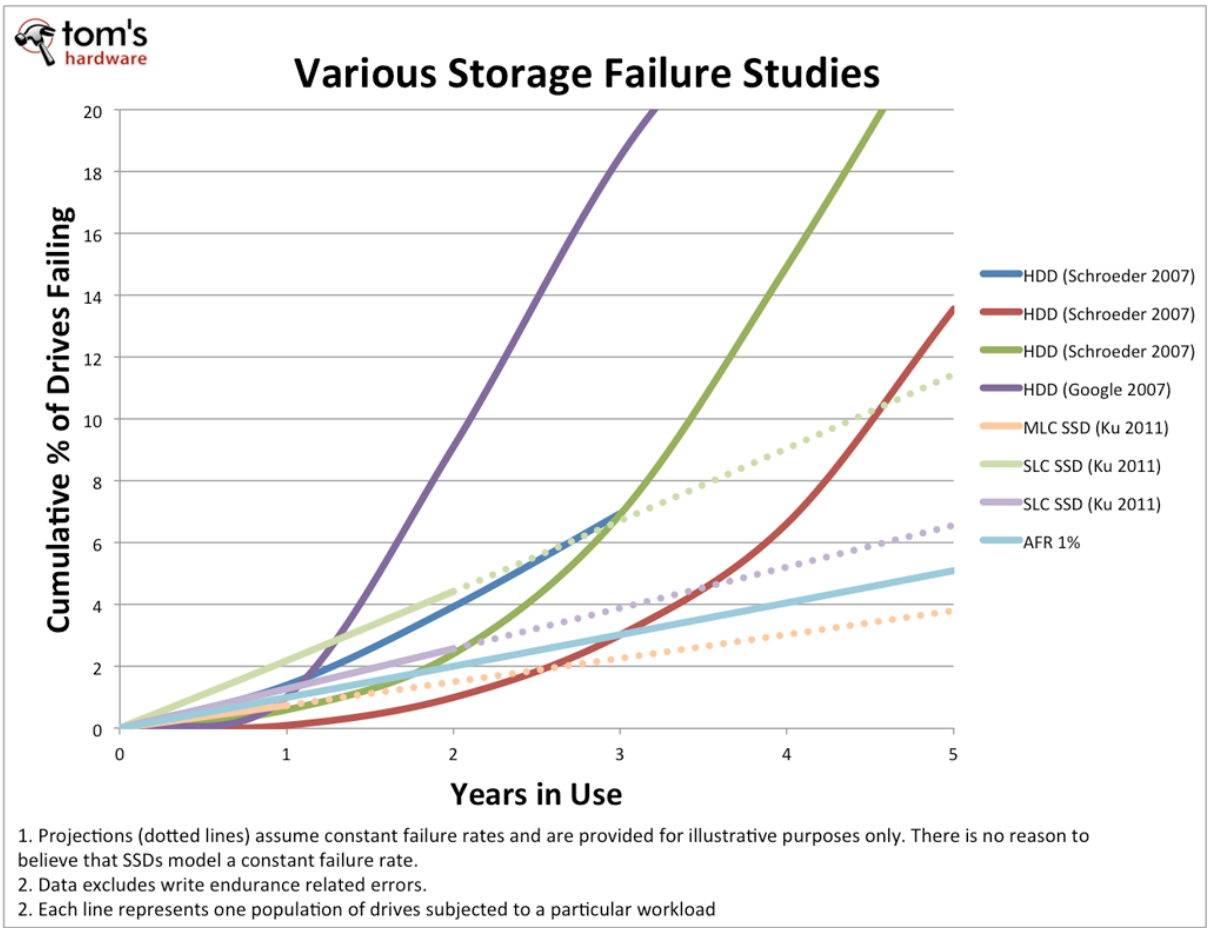
Magnetska traka ima nekoliko prednosti u odnosu na druge medije za pohranu. Magnetska traka može održati svoje performanse više od 30 godina. Nadalje, magnetska traka s premazom je relativno jeftina jer se može brzo proizvesti. Sustav za pohranu trake troši energiju samo tijekom snimanja i pretraživanja, omogućavajući rad koji štedi energiju. Budući da se podaci mogu premjestiti i pohraniti na različito mjesto u odnosu na glavni poslužitelj, slučaj kvara sustava, katastrofa ili druge nesreće manje će ili gotovo uopće neće ugroziti rad sustava. Veliki dio svjetskih podataka i dalje se čuva na magnetskim trakama, uključujući podatke za prirodne znanosti, poput fizike i astronomije, zatim za očuvanje ljudskog nasljeđa i nacionalnih arhiva, kao i u bankarstvu, osiguranju, istraživanju nafte itd. Stručnjaci neprestano rade na povećanju kvalitete i kapaciteta kaseta. "Prvi komercijalni sustav za pohranu digitalnih traka, IBM-ov Model 726, mogao bi pohraniti oko 1,1 megabajta na jedan kolut trake. Danas moderni uložak s trakom može pohraniti 15 terabajta. A jedna knjižnica s robotiziranim sustavom za pohranu i dohvat traka može sadržavati do 278 petabajta podataka. Za pohranu toliko podataka na kompaktnim diskovima trebalo bi ih više od 397 milijuna, koji bi, ako su naslagani, tvorili toranj visok više od 476 kilometara."[10] Arhiviranje je proces pohrane podataka koji zahtijevaju trajno/dugoročno pohranjivanje. Podaci se vraćaju iz arhiva i koriste se prema potrebi. Zahvaljujući visokoj pouzdanosti i proširivosti, pohranjivanje ne magnetskim trakama i danas se koristi kao arhivski sustav. Uz povećani kapacitet i brzinu prijenosa trake za pohranu, proširuje se i raspon njihove uporabe za arhivsku pohranu. Većina prisutnih pogona magnetskih traka uključuje funkciju koja se naziva hardverska kompresija. To omogućuje pogon kompresijom podataka na magnetsku vrpcu. U mnogim se slučajevima ova značajka može pokazati vrlo korisnom. Hardverska kompresija mnogo je brža od softverske jer za razliku od kompresije softverom, ona ne koristi računalni procesor. Hardverska kompresija se ne mora posebno podešavati na razini operativnoga sustava, a podaci se komprimiraju tijekom zapisivanja.

[9] Ibid., str. 115.
[10] Lantz, M. Why the Future of Data Storage is (Still) Magnetic Tape, 28.08.2018. // IEEE Spectrum. Dostupno na: https://spectrum.ieee.org/computing/hardware/why-the-future-of-data-storage-is-still-magnetic-tape (17.08.2020.).

3.2. Izravni sustavi

Izravni sustavi omogućuju izravni pristup podatcima (engl. *Direct Access Storage – DAS*). Gradivo se u ovakvim sustavima pohranjuje na čvrste diskove. Oni služe kao proširenje već postojećem poslužitelju i nisu umreženi kao samostalna mrežna jedinica. Jedan disk svojim kapacitetom ne bi mogao pohraniti sve pa se koristi polje diskova (engl. *Redundand Array of Drives – RAID*) koji zajedno tvore logičku cjelinu te korisniku nalikuju na jedan disk. Dakako, polje diskova ima bolje performanse nego svaki pojedinačni disk. U ovome se sustavu digitalni zapisi dijele u blokove te se svaki blok zapisuje na drugi disk, time se smanjuje vrijeme potrebno za zapisivanje. Kod zahtjeva za čitanjem određenog zapisa se ukupna brzina čitanja povećava tako da nekoliko diskova čita manju količinu podataka. Također se sugerira korištenje polja diskova koje obavlja sigurnosnu kopiju zapisa kako se u slučaju kvara podatci ne bi izgubili, ali to treba razlikovati od izrade pričuvne kopije, jer je ovdje riječ o zalihosti na razini sustava. „Kod čvrstih diskova brisanje uz naknadnu nemogućnost povratka izbrisanih podataka provodi se višestrukim brisanjem fizičke pozicije na disku ili njezino višestruko prepisivanje novim podatcima.“[11] Izravni sustavi su skupi, iako im cijena opada a kapacitet se povećava, te se koriste kada je potrebno osigurati brzi pristup. Prosječna trajnost ovog sustava iskazuje se s prosječnim brojem sati do prvog kvara ili između dva kvara (engl. *Mean Time Between Failure – MTBF*) (slika 2). Kod nekih se diskova to vrijeme kreće oko 1,2 milijuna sati rada što čini 137 godina rada 24 sata na dan, 7 dana u tjednu.

[11] Stančić, H. Digitalizacija. Zagreb : Zavod za informacijske studije, 2009. str. 122.



Slika 2: MTBF različitih diskova[12]

3.2.1. RAID

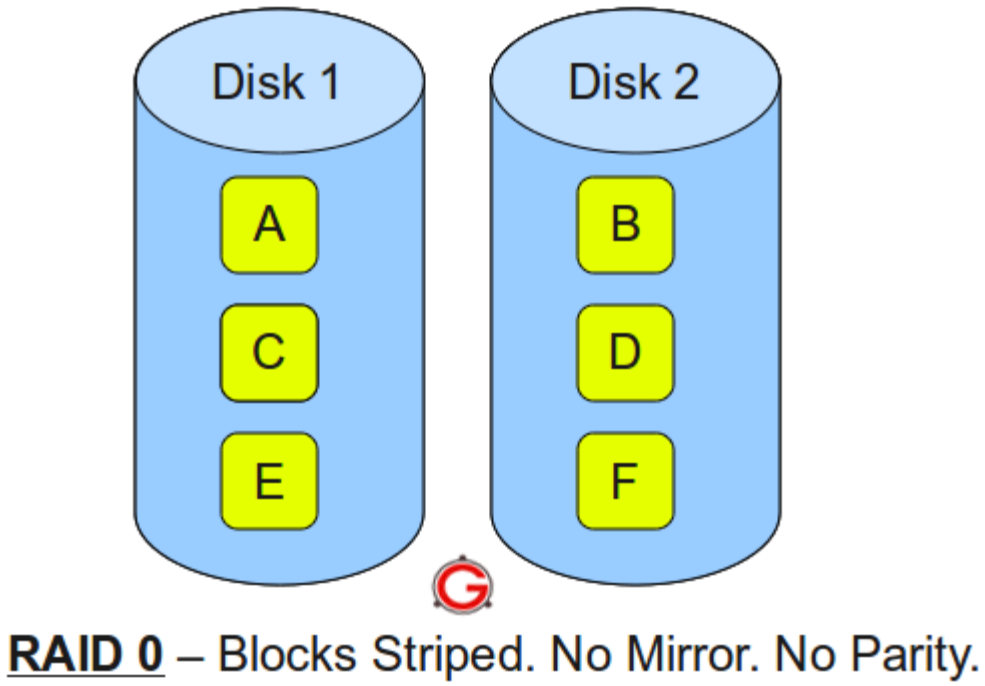
Postoje različite RAID konfiguracije, od kojih svaka primjenjuje različite tehnike upravljanja podacima. Tri najvažnija izraza koja se koriste kod RAID polja su:

1. dijeljenje (engl. *stripping*): dijeli podatke u blokove, zapisujući ih paralelno na diskove unutar RAID niza,
2. zrcaljenje (engl. *mirroring*): sprema identične kopije podataka u različite blokove istovremeno,
3. paritet (engl. *parity*): izračunava blok koji nedostaje kako bi se spriječio gubitak podataka u slučaju neispravnog diska.

Najpopularnije RAID konfiguracije koje se koriste (ili njihove kombinacije) da bi se zadovoljile različite potrebe su RAID 0, RAID 1, RAID 5, RAID 6 i RAID 10.

[12] Ordel, S. What is the failure rate of Solid-State Drives (SSD), 1.03.2012. // StackExchange. Dostupno na: https://skeptics.stackexchange.com/questions/7084/what-is-the-failure-rate-of-solid-state-drives-ssd (1.09.2020.).

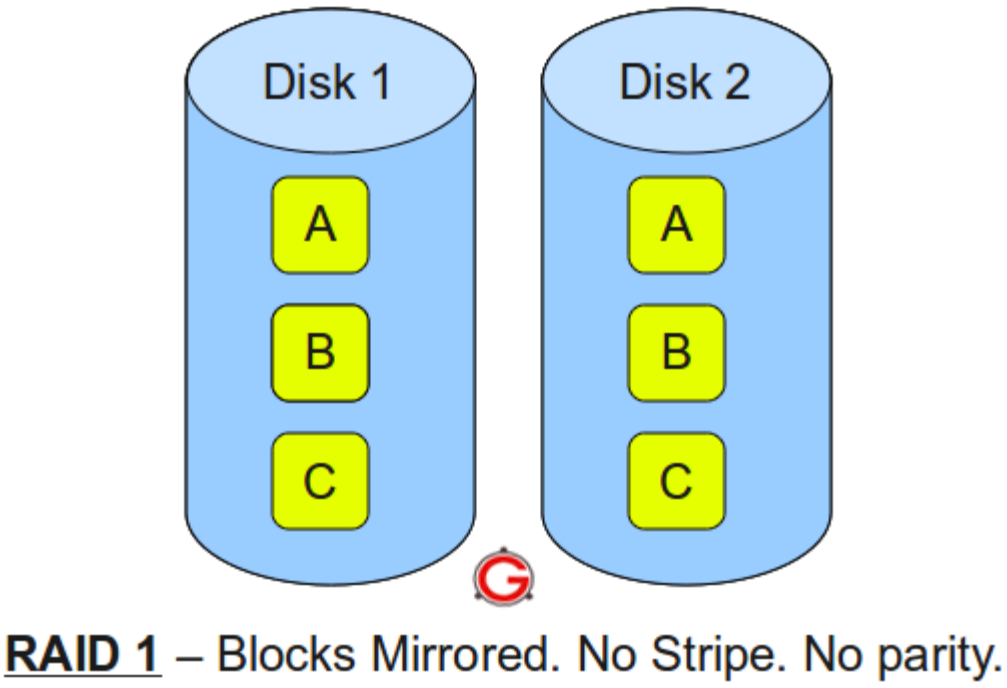
RAID 0 pruža značajno ubrzanje performansi, ali ne uključuje redundantne diskove. Podaci koji se bilježe podijeljeni su u nekoliko blokova, ovisno o broju diskova, a jedan se dio podataka bilježi na svaki disk. Ako jedan od diskova prestane raditi, svi se podaci gube. Ubrzanje performansi je značajno, a to može vrlo lako ukloniti usko grlo današnjih računala - brzinu diska.



Slika 3: RAID 0[13]

RAID 1 omogućuje zrcaljenje diska. To pruža dodatnu sigurnost jer se isti podaci bilježe na dva diska. Čitanje podataka s takvog diska dvostruko je brže jer se jedan dio čita s jednog diska, dok se drugi dio čita s drugog diska. Zapisivanje podataka na disk odvija se istom brzinom kao i kod jednostrukih diskova.

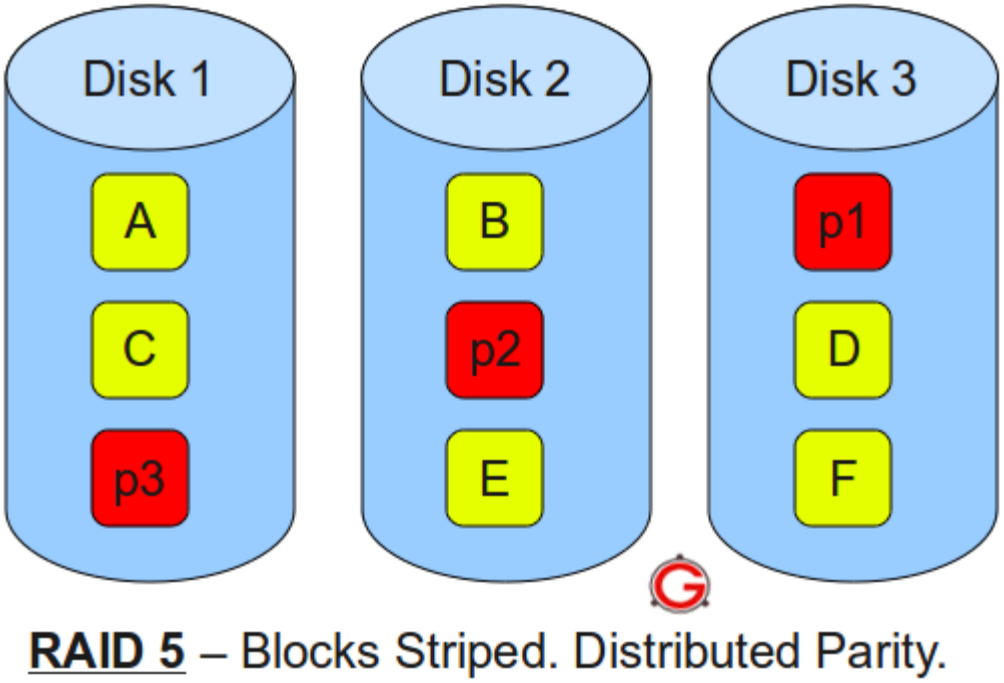
[13] Natarajan, R. RAID 0, RAID 1, RAID 5, RAID 10 Explained with Diagrams, 10.8.2010. // The Geek Stuff. Dostupno na: https://www.thegeekstuff.com/2010/08/raid-levels-tutorial/ (17.08.2020.).



Slika 4: RAID 1[14]

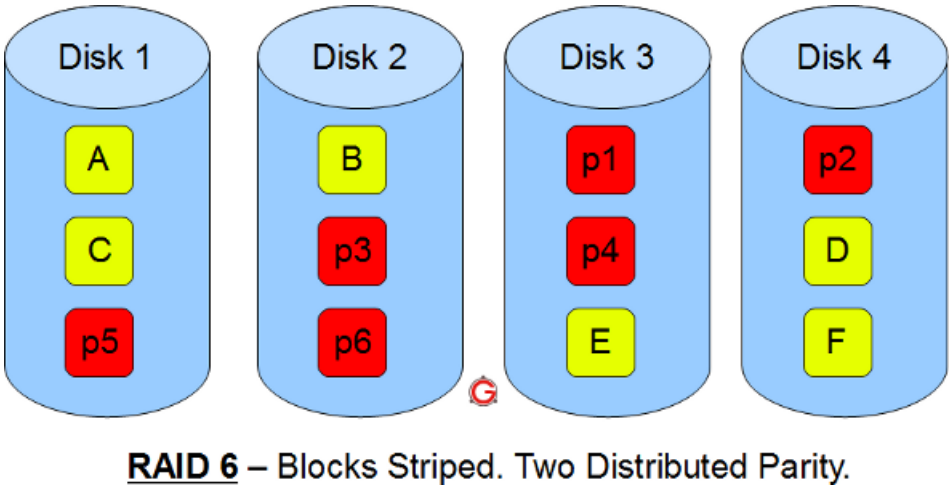
RAID 5 je najčešće korišteni tip RAID polja. Podaci se dijele na razini bajta, a zapisuju se i sigurnosni bitovi za zaštitu od grešaka. Ova vrsta omogućuje vrlo brz pristup podacima i dobru zaštitu od neispravnosti ili grešaka.

[14] Ibid.



Slika 5: RAID 5[15]

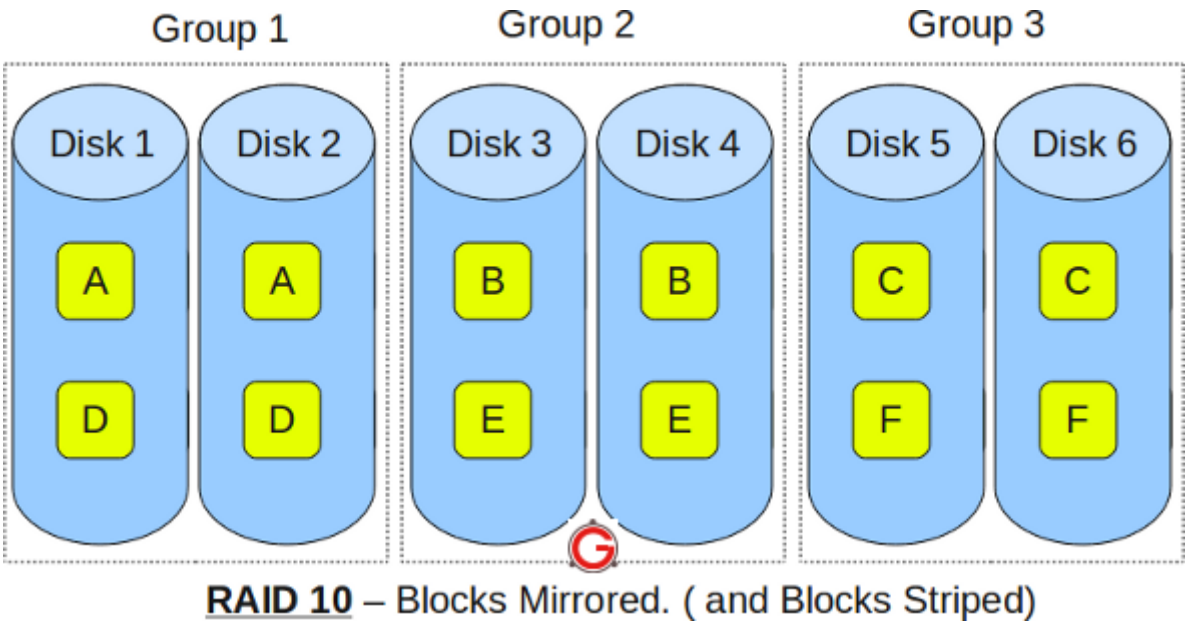
RAID 6 primjenjuje dijeljenje podataka na blokove pri čemu se pariteti zapisuju na sve diskove.



Slika 6: RAID 6[16]

[15] Ibid.
[16] Natarajan, R. RAID 2, RAID 3, RAID 4, RAID 6 Explained with Diagram, 21.11.2011. // The Geek Stuff. Dostupno na: https://www.thegeekstuff.com/2011/11/raid2-raid3-raid4-raid6/ (17.08.2020.).

RAID 10 stvara više RAID 1 polja, dok se među njima stvara RAID 0 polje. RAID 10 se također naziva i RAID 1 + 0. Potrebna su najmanje 4 diska. „Na primjer, ako u RAID 10 postoji ukupno 6 diskova, postojat će tri grupe - grupa 1, grupa 2, grupa 3. Unutar grupe podaci se odražavaju. U primjeru vidljivom na slici 7, Disk 1 i Disk 2 pripadaju grupi 1. Podaci na Disku 1 bit će potpuno isti kao i podaci na Disku 2. Dakle, blok A napisan na Disku 1 bit će zrcaljen na Disku 2. Blok B napisan na Disk 3 će se ogledati na Disku 4. U cijeloj grupi podaci su dijeljeni. tj. blok A je napisan skupini 1, blok B je napisan grupi 2, blok C napisan je grupi 3.“[17]



Slika 7: RAID 10[18]

Današnji kontroleri za RAID polja omogućuju zamjenu diskova (engl. *hot swapping*) pa se diskovi mogu mijenjati i u radu bez potrebe za gašenjem računala što je od kritične važnosti za servere.

3.3. Poluizravni sustavi

Poluizravni sustavi imaju izvrstan odnos cijene i kapaciteta. U tim se sustavima podatci čuvaju na jeftinijim medijima kao što su optički diskovi, magnetske trake i sl. „Poluizravni sustavi za pohranu i prijenos podatka sastoje se od smještajnog dijela, nekoliko čitača i

[17] Natarajan, R. RAID 10 Vs RAID 01 (RAID 1+0 Vs RAID 0+1) Explained with Diagram, 24.10.2011. // The Geek Stuff. Dostupno na: https://www.thegeekstuff.com/2011/11/raid2-raid3-raid4-raid6/ (17.08.2020.).
[18] Ibid.

robotske ruke koja služi za automatski prihvat medija i njegov prijenos od smještajnog dijela do čitača i natrag.“[19] Diskovni automati (engl. *jukebox*) su robotski sustavi koji imaju manji smještajni kapacitet, a oni većeg kapaciteta se nazivaju silosi. Stančić tumači da postoje mnoge prednosti kod korištenja robotskih sustava naspram ručnog prihvata medija:

- manji troškovi korištenja – ne plaća se poslužiteljsko osoblje,
- manji troškovi osvježavanja i migracije zapisa,
- brži pristup gradivu,
- moguć pristup 24 sata u danu i 7 dana u tjednu,
- manja mogućnost da se medij izgubi ili vrati na krivo mjesto,
- mogućnost rada u hijerarhijskom sustavu za pohranu.

Najveća mana poluizravnih sustava je spor pristup do kojeg dolazi zbog vremena koje je potrebno za dohvat medija i pronalazak traženih podataka, npr. za premotavanje trake do nekog mjesta te slijednog načina čitanja podataka. Ukoliko dođe do nagomilavanja zahtjeva za čitanjem podataka moguća su duga čekanja. Poluizravni sustavi u kojima se koriste magnetske trake još se nazivaju i kontaktni sustavi jer za vrijeme čitanja dolazi do fizičkog kontakta između čitača i medija. Ovakav način čitanja nakon duljeg perioda rezultira u trošenju materijala te njegovog oštećenja. Magnetske su trake osjetljive na promjene u temperaturi i vlažnosti pa je trajnost ovog medija potrebno provjeravati, a po potrebi se zapisi moraju kopirati na novu traku. Taj se postupak naziva postupkom osvježavanja medija. Također, kada su u pitanju sustavi s magnetskim trakama, nije moguće brisanje određenih podataka nego samo trake u cijelosti ili svih podataka od zadanog mjesta pa do kraja trake.

3.4. Hijerarhijski sustav

Hijerarhijski je sustav spoj izravnog i poluizravnog sustava. Ovaj se sustav sastoji od čvrstog diska ili polja diskova te magnetskih traka ili nekih drugih izmjenjivih medija. Tehnika sustava bazira se na statistici korištenja digitalnih zapisa. Noviji zapisi češće se koriste pa ih se najprije pohranjuje na tvrdi disk. Nakon nekog vremena zapisi koji su se najmanje koristili premještaju se na izmjenjive medije. Hijerarhijski se sustavi opisuju kao virtualni diskovi jer cijela digitalna zbirka izgleda kao jedna cjelina. Jedina razlika što se tiče fizičkog smještaja jest što je potrebno dulje čekati kako bi se dohvatilo gradivo koje je

[19] Stančić, H. Digitalizacija. Zagreb : Zavod za informacijske studije, 2009. str. 117.

pohranjeno na izmjenjivom mediju. Kod ovih je sustava moguće učiniti cjelovito uništenje podataka uništenjem medija ili brisanjem podataka.

3.5. Neizravni sustavi

U neizravnim sustavima nije moguće izravno pristupiti gradivu bez obzira na to što su metapodatci o njemu dostupni putem mreže. Od iznimne je važnosti da je metabaza (baza metapodataka) usklađena s podatcima koji su pohranjeni u digitalnom arhivu. Iako ovi sustavi ograničavaju pristup gradivu, oni ubrzavaju pronalaženje gradiva i pretraživanje. Vrlo se često koriste za izradu i održavanje sigurnosnih kopija (engl. *backup*). Potreba za ljudskom intervencijom kako bi se pristupilo podatcima bitna je karakteristika neizravnih sustava.

3.6. Mrežna pohrana i mreža za pohranu

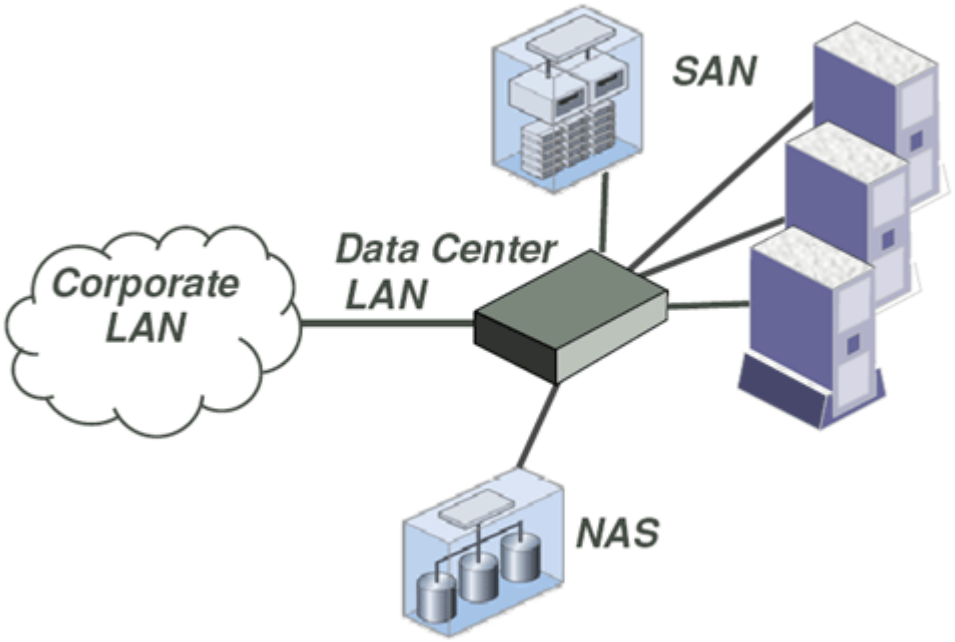
Mrežna pohrana (engl. *Network Attached Storage – NAS*) je koncept pohrane velike količine podataka koji je sličan izravnim sustavima, ali se od njih razlikuje u nekoliko pogleda. Ponajprije, NAS se priključuje na mrežu kao samostalna i nezavisna jedinica, a u izravnim se sustavima slični uređaju spajaju na poslužitelj i djeluju kao proširenje njegovog kapaciteta. Za potrebe konfiguracije, ovim se sustavima pristupa mrežnim preglednikom tako da se spaja izravno na njihovu mrežnu adresu. Tu se najčešće radi o RAID sustavu diskova za pohranu podataka.

Mreža za pohranu (engl. *SAN*) je koncept koji je namijenjen pohranjivanju podataka na razini institucije i koji je usklađen s činjenicom da su institucije fizički distribuirane, ali međusobno povezane globalnom informacijskom infrastrukturom. Mreža za pohranu i računalna mreža međusobno su povezane, ali mreža za pohranu čini podmrežu koja povezuje sve uređaje za trajnu pohranu i izradu sigurnosnih kopija. „Do nedugo je pohranjivanje bilo organizirano tako da su se na više mjesta nalazila poslužiteljska mjesta koja imaju implementirana neka od rješenja za pohranu. SAN objedinjuje sve sustave pa se njima može upravljati s jednog mjesta.“[20] Ovim se rješenjem optimalnije iskorištava prostor za pohranjivanje.

Mrežna pohrana i mreža za pohranu nerijetko se uspoređuju, a glavna razlika među njima je činjenica da SAN kod pohranjivanja podatke organizira u blokove, a NAS s druge strane kao

[20] Stančić, H. Digitalizacija. Zagreb : Zavod za informacijske studije, 2009. str. 128.

osnovnu organizacijsku jedinicu koristi datoteke. Iako se razlikuju, SAN i NAS se mogu koristiti kombinirano uslijed čega se stvori hibridno SAN-NAS rješenje (slika 8).



Slika 8: SAN-NAS hibridni sustav[21]

3.7. Pohrana u računalnome oblaku

Sve popularniji način pohrane podataka danas je računalni oblak. Iako oblak nije dizajniran za arhiviranje, realnost je da ga u tu svrhu koriste neke institucije. Očuvanje arhivskih zapisa putem trećih strana predstavlja veći rizik od gubitka ako nešto pođe po zlu. „Treća strana može prestati poslovati - Nirvanix i Megacloud dva su takva primjera - tako da je potrebno uspostaviti strategiju prelaska drugom pružatelju usluga u računalnome oblaku (engl. *Cloud Service Provider – CSP*) za takve scenarije.“[22] Dodatne probleme mogu predstavljati autorska prava, licenciranje i sigurnost u vezi s pohranom u oblaku, a privatnost osobnih podataka je također sve više ugrožena.

„Prije su arhivisti bili viđeni i predstavljani kao čuvari vjerodostojnih zapisa, ali posve je prirodno da se ta slika može i mora mijenjati. Jedan od arhivista rekao je: 'Arhivist je postao

[21] Convergence of Networking and Storage. // Chelsio Communications. Dostupno na: https://www.chelsio.com/convergence_of_networking_and_storage/ (1.09.2020.).
[22] Houghton, B. Preservation Challenges in the Digital Age. // D-Lib Magazine. 22(7/8), srpanj/kolovoz 2016. Dostupno na: http://www.dlib.org/dlib/july16/houghton/07houghton.html (17.06.2020.).

više čuvar arhivskih interesa cijele organizacije nego samo čuvar zapisa."“[23] Možda će također trebati izraditi pravila i propise koji podupiru rad organizacije s raznim uslugama oblaka. Potreba arhivista da budu proaktivni i proširuju svoje odgovornosti izvan čuvara zapisa mijenja svoju ulogu. Arhivist je sada odgovoran za upravljanje informacijama i njihovo upravljanje - oni više nisu samo čuvari, već više kontrolori odgovorni za upravljanje informacijama.

3.8. Formati datoteka

Oblici datoteka odavno se smatraju jednim od najvećih rizika digitalne sačuvanosti. Međutim, to se nije pokazalo pretjeranom opasnošću kako se to u početku percipiralo. Dobrim dijelom to je zbog dostupnosti otvorenih formata datoteka, što rezultira podrškom za više različitih softvera. Vlasnički formati datoteka i dalje predstavljaju izazov jer je manje vjerojatno da će njihove specifikacije biti javno dostupne. Pretvaranje softvera da postane kompatibilan s takvim formatima ili pretvaranje datoteka u otvoreniji format može se izvršiti samo uz dozvolu vlasnika patenta. „To komplicira dugoročno očuvanje takvih datoteka jer ih možda neće biti moguće konvertirati u pristupačniji format.“[24] Da bi takve datoteke bile dostupne, možda je potrebno sačuvati i softver što donosi novi set problema. Naime, posve je sigurno da softver koji radi s trenutnim hardverom i operativnim sustavom nakon nekog vremena više neće biti moguće pokrenuti na tada suvremenom hardveru i operativnom sustavu. Uz to, nisu svi formati datoteka prikladni za dugoročno očuvanje, čak i ako imaju otvorenu specifikaciju. Dodatno, komprimirani formati datoteka predstavljaju veći rizik od potpunog gubitka ako se izgubi čak i jedan bit. Neki formati datoteka, zbog svojih karakteristika, predstavljaju općeprihvaćeni arhivski format. TIFF je, primjerice, prihvaćeni takav format za slike. Međutim, nemaju sve vrste medija optimalni arhivski format. Tako se na primjer videozapisi uvijek komprimiraju, iako najčešće bez gubitaka, jer bi u nekomprimiranom obliku, kao što je to TIFF za slike, naprosto bili preveliki. „Iako će se ovaj problem s vremenom riješiti, institucije za očuvanje moraju u međuvremenu iskoristiti svoje najbolje procjene o tome koji formati datoteka za pohranu trebaju koristiti u takvim slučajevima ili se osloniti na preporučene primjere dobre prakse koje su izradile druge

[23] Duranti, L., et al. Records and Archives in the Cloud. // The Canadian Journal of Information and Library Science. 2015. str.122.
[24] Houghton, B. Preservation Challenges in the Digital Age. // D-Lib Magazine. 22(7/8), srpanj/kolovoz 2016. Dostupno na: http://www.dlib.org/dlib/july16/houghton/07houghton.html (17.06.2020.).

institucije. Dobar primjer toga je američka Kongresna knjižnica i njezin popis preporučenih arhivskih formata.“[25]

[25] Library of Congress Recommended Formats Statement 2019-2020, Dostupno na: https://www.loc.gov/preservation/resources/rfs/TOC.html

4. Metapodaci

Ozbiljnije zanimanje za metapodatke se među arhivistima pojavilo u vezi s problemom zaštite, preuzimanja i upravljanja elektroničkim dokumentima. Metapodaci su vjerojatno najbitniji aspekt očuvanja i pohranjivanja digitalnih zapisa. Najopćenitije su definirani kao podaci o podacima. To su strukturni podatci koji opisuju, objašnjavaju, pomažu u pronalaženju ili na drugi način olakšavaju pronalazak, korištenje i upravljanje informacijama. Zapisi s jednostavnim metapodacima ili bez njih se teško pronalaze, ne mogu se verificirati te njihov sadržaj čak može biti i nejasan. Također je vrlo vjerojatno da ih se neće moći koristiti u istoj mjeri kao kada imaju kvalitetne metapodatke. Pohranjivanje zapisa bez valjanih metapodataka je ekvivalentno bacanju tih objekata u otpad. Ako je moguće, važno je prikupiti metapodatke u trenutku stvaranja s obzirom na to da se neki sadržaj s vremenom može izgubiti. Važnost toga prepoznata je u polju istraživanja, s tim da se sada ulaže više napora u poticanje istraživača na stvaranje planova za upravljanje podacima na početku projekata. Tako na primjer, PDF zapisima često prijeti opasnost da sadrže loše ili netočno ugrađene metapodatke. Prema Houghtonu, nerijetko se događa da je "autor" ime osobe koja je datoteku pretvorila u PDF, a ne stvarni autor sadržaja.[26] Loši metapodaci poput ovoga mogu imati štetan utjecaj na dugoročno očuvanje datoteke i njezinog konteksta. Važno je da se bilo kakve promjene u zapisu ili njegovim derivatima dobro dokumentiraju u metapodacima. „Nekoliko je vrsta metapodataka:

- opisni metapodaci specificiraju izvor informacija s ciljem njegovog pronalaženja i identifikacije. Mogu sadržavati elemente kao što su naslov, sažetak, autor i ključne riječi,
- strukturalni metapodaci pokazuju kako su složeni digitalni objekti konstruirani,
- administrativni metapodaci sadrže informacije koje pridonose upravljanju izvorima informacija - kada i kako su nastali, tip datoteke te ostali tehnički podaci i tko ih može koristiti. Postoji nekoliko podskupova administrativnih metapodataka, a dva koja su ponekad navedena kao zasebna vrsta metapodataka su:
    - metapodaci za upravljanje pravima, koji se bave intelektualnim vlasništvom,

[26] Houghton, B. Preservation Challenges in the Digital Age. // D-Lib Magazine. 22(7/8), srpanj/kolovoz 2016. Dostupno na: http://www.dlib.org/dlib/july16/houghton/07houghton.html (17.06.2020.).

- metapodaci za očuvanje, koji sadrže podatke potrebne za arhiviranje i čuvanje izvora informacija.“[27]

Metapodaci se od ostalih vrsta podataka razlikuju po strukturiranosti i dosljednosti u načinu opisivanja podataka jer koriste utvrđene standarde. Kada se oblikuju kvalitetni metapodaci vrlo je važno obuhvatiti opisne podatke koji su neophodni za pronalaženje, razumijevanje i korištenje opisanih digitalnih zapisa. „Metapodatke se smatra kvalitetnima ako:

- su u skladu sa sadržajem zbirke, prikladni korisnicima zbirke, i namijenjeni aktualnom i budućem korištenju digitalnog objekta,
- podržavaju interoperabilnost,
- koriste standardne kontrolirane rječnike za opis teme, mjesta, vremena i autora u sadržaju zbirke
- sadrže jasnu izjavu o uvjetima korištenja digitalnog objekta,
- podržavaju dugoročno upravljanje objektima u zbirkama,
- ako imaju kvalitete dobrih objekata, uključujući mogućnost pohrane, dosljednost, jedinstvenu identifikaciju, itd.,
- su vjerodostojni i provjerljivi.“[28]

Standarde metapodataka najčešće razvijaju korisničke zajednice kako bi se omogućio najbolji način opisa nekog informacijskog zapisa za potrebe te zajednice. Sheme i standardi stvoreni su kako bi se nametnula struktura i dosljednost u zapisivanju podataka. Time se osigurava točnost i pouzdanost te se korisnicima omogućava unakrsno pretraživanje različitih zbirki što promovira interoperabilnost. Standard za metapodatke jest specifikacija koja sadrži skup polja ili elemenata, od kojih je svako polje ili element oblikovano tako da sadrži podatke o nekom aspektu izvora informacija. Standard definira značenje svakog elementa i pruža smjernice za njegovu primjenu. Neki od standarada koji se koriste u arhivskim institucijama su ISAD(G) (engl. *General International Standard Archival Description*), EAD (engl. *Encoded Archival Description*), ISAAR(CPF) (engl. *International Standard Archival Authority Record for Corporate Bodies, Persons and Families*), ISDIAH (engl. *International Standard for Describing Institutions with Archival Holdings*) i ISDF (engl. *International Standard for Describing Functions*).

[27] National Information Standards Organization. Understanding metadata. 2004. str. 1. Dostupno na: https://www.lter.uaf.edu/metadata_files/UnderstandingMetadata.pdf (17.06.2020.).
[28] Cole, W. T. Creating a Framefork of Guidance for Building Good Digital Collections, 2002. str. 10. Dostupno na: http://firstmonday.org/ojs/index.php/fm/article/view/955/876 (17.06.2020.).

5. Metode osiguranja integriteta

Sve je više dokumenata koji nastaju izvorno u digitalnom obliku ili su digitalizirani, stoga problem njihovog arhiviranja i dugotrajnog očuvanja postaje sve veći. Postavlja se pitanje kako riješiti probleme i zahtjeve, koju tehnologiju koristiti te na koje se norme osloniti kako bi se što kvalitetnije arhivirao i očuvao neki digitalni zapis. Iznimno je važno da digitalni zapis u procesu arhiviranja i dugotrajnog očuvanja ostane autentičan, pouzdan, očuvanog integriteta, iskoristiv te sačuvanog konteksta. „Autentičan je onaj spis za koji se može dokazati:

a) da je ono što tvrdi da jest,

b) da ga je izradila ili odaslala osoba za koju se tvrdi da ga je izradila ili poslala, i

c) da je izrađen ili odaslan u vrijeme za koje se to tvrdi.

Pouzdan je onaj spis za čiji se sadržaj može vjerovati da potpuno i točno predstavlja transakcije, aktivnosti ili činjenice koje potvrđuje i o kojima može biti ovisan tijekom sljedećih transakcija i aktivnosti. Integritet spisa se odnosi na to da je potpun i neizmijenjen. Iskoristiv je onaj spis kojemu je moguće utvrditi smještaj, dohvatiti ga, predočiti i interpretirati.“[29] „Kontekst se odnosi na međusobne veze pojedinih zapisa te okolinu u kojoj je zapis stvoren.“[30] U procesu očuvanja digitalnoga gradiva nije riječ samo o očuvanju datoteka, već se teži omogućavanju pristupa sadržaju i osiguravanju očuvanja vjerodostojnosti dokumenta.

Kako bi se osigurao integritet digitalnog zapisa, zapisima se dodaju digitalni potpis, digitalni certifikat, digitalni vodeni žig, digitalni vremenski žig, a u novije vrijeme se koristi i princip ulančanih blokova (engl. *blockchain*).

5.1. Digitalni potpis

Ono što razlikuje analogni od digitalnog potpisa jest da je digitalni potpis različit za svaki dokument koji autor potpiše. Koncept na kojem se digitalni potpis temelji jest infrastruktura

[29] ISO 15489-1: Informacije i dokumentacija - Upravljanje spisima - 1. dio: Općenito. Prijevod za internu upotrebu, str. 11.
[30] Stančić, H. Arhiviranje digitalnih dokumenata. // 4. seminar Arhivi, knjižnice i muzeji. Mogućnosti suradnje u okruženju globalne informacijske infrastrukture. Zagreb, 2001., str. 210 - 211.

javnog ključa (engl. *Public Key Infrastructure - PKI*) u kojem autor generira par ključeva - privatni i javni ključ. Privatni ključ uvijek ostaje kod autora, a javni ključ postaje javno dostupan. Postoje dvije vrste digitalnog potpisa - osnovni i napredni. Uredba (EU) br. 910/2014 Europskog parlamenta i vijeća od 23. srpnja 2014. o elektroničkoj identifikaciji i uslugama povjerenja za elektroničke transakcije na unutarnjem tržištu i stavljanju izvan snage Direktive 1999/93/EZ kaže da elektronički potpis treba ispuniti sljedeće zahtjeve da postane napredni elektronički potpis:

- „jedinstveno je povezan s potpisnikom,
- sposoban je identificirati potpisnika,
- stvoren je korištenjem sredstava koje potpisnik može održavati pod svojim isključivim nadzorom,
- povezan je s podacima na koje se odnosi na takav način da se otkrije svaka naknadna promjena podataka.“[31]

„Digitalni potpisi mogu se realizirati u više formata – XMLDSig (engl. *XML Digital Signature*), XAdES (engl. *XML Advanced Electronic Signature*), CAdES (engl. *Cryptographic Message Syntax Advanced Electronic Signature*) i PAdES (engl. *PDF Advanced Electronic Signature*).“[32] Kad autor želi digitalno potpisati dokument, on primjenjuje svoj privatni ključ na dokument i šalje ga primatelju. Primatelj, primjenjujući javni ključ autora na primljeni dokument, može potvrditi je li autor potpisao dokument. Rezultat provjere je jednostavno „Da“ ili „Ne“. Ako je dokument izmijenjen, primjena autorskog javnog ključa za izmijenjeni dokument rezultira pobijanjem. Važno je da je matematički nemoguće izračunati privatni ključ iz javnog ključa. Autor također može izračunati takozvani sažetak dokumenata (engl. *document digest*) primjenom algoritma za izradu kriptografskog sažetka (engl. *hash*). Rezultat je alfanumerički niz koji je jedinstven za svaki dokument. Ako se dokument promijeni, promijenit će se i njegova rezultirajuća *hash* vrijednost. Brži način potpisivanja i provjere je da kad autor želi digitalno potpisati dokument, on primjenjuje svoj privatni ključ na izrađeni sažetak dokumenta i šalje ga primatelju. Primatelj ponovno izračunava sažetak dokumenata i primjenom autorovog javnog

[31] Uredba (EU) br. 910/2014 Europskog parlamenta i Vijeća od 23. srpnja 2014. o elektroničkoj identifikaciji i uslugama povjerenja za elektroničke transakcije na unutarnjem tržištu i stavljanju izvan snage Direktive 1999/93/EZ. Dostupno na: https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX%3A32014R0910 (14.08.2020.).

[32] Rajh A., Stančić H., Romčević B., Vitaljić M. Koncept rješenja za osiguranje i očuvanje vjerodostojnosti zapisa u upravnim organizacijama prilikom razvoja državnog računalnog oblaka i državnog digitalnog arhiva. Arhivski vjesnik br. 61 (2018), str. 73.

ključa provjerava je li autor potpisao dokument. Ovo je brži postupak s istim učinkom, jer se potpis potvrđuje ili odbacuje provjerom potpisa kriptografskog sažetka dokumenta, a ne cijelog dokumenta, koji može biti dugotrajan. Korištenjem osnovnog digitalnog potpisa nemoguće je sa sigurnošću znati tko je prava osoba koja stoji iza digitalnog potpisa. Brojne su web usluge koje nude stvaranje kombinacije javnih i privatnih ključeva i svatko može tvrditi da je bilo tko drugi ako to želi. Za dokazivanje identiteta autora potrebni su digitalni certifikati.

5.1.1. Digitalni certifikat

Ukoliko se želi potvrditi identitet autora digitalnog potpisa potrebna je pouzdana treća strana koja se zove certifikacijsko tijelo. Usporedba digitalnog certifikata i virtualne ID kartice izdane od strane pouzdanog tijela pokazuje da infrastruktura javnog ključa uključuje organizacije zvane certifikacijska tijela koja izdaju, upravljaju i opozivaju digitalne certifikate. Ta tijela mogu stvoriti zasebno tijelo za registraciju koje će se baviti identifikacijom pojedinaca koji se žele prijaviti za certifikat. Na primjer, odjeljak za pomoć za Google Chrome objašnjava da je potvrda javnog ključa, koja se obično naziva samo certifikat, digitalno potpisana izjava koja veže vrijednost javnog ključa s identitetom osobe, uređaja ili usluge koja ima odgovarajući privatni ključ.[33] S obzirom na to da certifikat odgovara javnom ključu određenog pojedinca, a izdavač jamči autentičnost certifikata, digitalni certifikat pruža rješenje problema kako pronaći javni ključ korisnika i znati da je valjan. Te probleme rješava primatelj dobivajući javni ključ autora iz digitalnog certifikata. Primatelj tada zna da je valjan jer je provjereno certifikacijsko tijelo izdalo certifikat. Nadalje, digitalni se certifikati oslanjaju na kriptografiju javnog ključa za vlastitu provjeru autentičnosti. Kad se izda digitalni certifikat, tijelo koje ga izdaje potpisuje certifikat vlastitim privatnim ključem. Kako bi se potvrdila autentičnost digitalnog certifikata, primatelj može dobiti javni ključ tog certifikacijskog tijela i upotrijebiti ga kako bi utvrdio je li certifikat potpisan od strane certifikacijskog tijela.

Digitalni certifikati mogu biti valjani ili se zbog određenog razloga mogu opozvati. Popis opoziva certifikata (engl. *Certificate Revocation List - CRL*) je popis pretplatnika koji su upareni sa statusom digitalnog certifikata i kojima su opozvani certifikati. Popis navodi i

[33] Googleova najsnažnija zaštita za one kojima je najpotrebnija.// Google Program Napredne Zaštite. Dostupno na: https://landing.google.com/advancedprotection/?utm_source=Chrome&utm_medium=ChromeSecuritySettings&utm_campaign=ChromeSettings (14.08.2020.).

razlog za opoziv. Uključen je i datum izdavanja certifikata te pouzdano tijelo koje ga je izdalo te datum i vrijeme opoziva. Glavno ograničenje CRL-a je činjenica da se ažuriranja moraju često preuzimati kako bi se popis održao aktualnim. Internetski protokol o statusu certifikata (engl. *Online Certificate Status Protocol - OCSP*) nadilazi ovo ograničenje provjerom statusa certifikata u stvarnom vremenu. U hrvatskom zakonodavstvu koncept neporecivosti je povezan s naprednim digitalnim potpisom koji se temelji na kvalificiranom certifikatu. „Da bi se postigla i sačuvala karakteristika neporecivosti nekog potpisanog dokumenta, potrebno je osigurati:

- digitalni identitet potpisnika,
- informaciju o opozivu prava na digitalni potpis u stvarnom vremenu,
- prisutnost digitalnog vremenskog žiga kojim se osigurava valjanost digitalnog potpisa u trenutku potpisivanja,
- dugoročno očuvanje.“[34]

5.2. Digitalni vremenski žig

U kontekstu digitalnog potpisa digitalna vremenska oznaka, to jest digitalni vremenski žig, igra važnu ulogu. Predstavlja digitalno potpisanu potvrdu izdavača vremenske oznake koja potvrđuje postojanje podataka, dokumenata ili zapisa na koje se vremenska oznaka odnosi. Digitalni vremenski žig daje pouzdan dokaz da su podaci, dokument ili zapis nastali ranije ili neposredno prije vremena naznačenog u digitalnom vremenskom žigu. Sve naknadne promjene podataka, dokumenata, zapisa ili vremenske oznake nisu dopuštene i mogu se lako otkriti. Prema tome, „digitalni vremenski žig potvrđuje:

- da su podaci, dokument ili zapis koji su trenutno postojali u tom obliku u vrijeme naznačeno u vremenskom žigu,
- da podaci, dokument ili zapis nisu promijenjeni nakon vremena naznačenog vremenskim žigom,

[34] Brzica, Hrvoje; Herceg, Boris; Stančić, Hrvoje Long-term Preservation of Validity of Electronically Signed Records // Information Governance / Gilliland, Anne; McKemmish, Sue; Stančić, Hrvoje; Seljan, Sanja; Lasić-Lazić, Jadranka (ur.). Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2013. str. 150-151. Dostupno na: https://infoz.ffzg.hr/INFuture/2013/papers/403%20Brzica,%20Herceg,%20Stancic,%20LTP%20of%20Validit%20of%20Electronically%20Signed%20Records.pdf

- da se provjera digitalnog potpisa može pouzdano obaviti i nakon opoziva ili isteka certifikata (u tom se slučaju može provjeriti da podaci, dokument ili zapis nisu promijenjeni, ali valjanost potpisanog certifikata ne može se provjeriti),
- da su podaci, dokument ili zapis poslani ili primljeni u vrijeme naznačeno u vremenskom žigu.“[35]

Tijelo za izdavanje digitalnih vremenskih žigova (engl. *Time-Stamping Authority - TSA)* digitalno potpisuje *hash* vrijednost podataka, dokumenta ili zapisa zajedno s informacijom o vremenu izdajući tako digitalni vremenski žig koji se kasnije kombinira s podacima, dokumentom ili zapisom i privatnim ključem potpisnika za stvaranje digitalnog potpisa i koja označava vrijeme potpisivanja.

5.2.1. Arhivski digitalni vremenski žig

Arhivski digitalni vremenski žig jest posebna vrsta vremenskog žiga koji je namijenjen za dugoročno očuvanje. Takav se arhivski žig od običnoga razlikuje po obuhvatu. On obuhvaća velik broj *hash* vrijednosti, to jest svih vrijednosti koje označavaju podatak koji je potrebno dugoročno očuvati. Nakon što autor digitalno potpiše dokument, te se tom potpisu dodaju potrebni certifikati generira se digitalni vremenski žig koji obuhvaća podatke u cijelosti te potpis nadopunjuje vremenskim žigom. Kako bi se gradivo dugoročno očuvalo, bitno je pravovremeno obnavljati arhivski digitalni vremenski žig. Stoga je „glavna namjena arhivskoga digitalnoga vremenskog žiga da produlji valjanost vremenskoga žiga, digitalnog potpisa kao i OCSP i CRL odgovora na upit o valjanosti e-potpisa i nakon isteka njihovih potpisnih certifikata.“[36]

5.3. Digitalni vodeni žig

Digitalni vodeni žig (engl. *watermark*) je signal dodan elektroničkom gradivu kako bi prenio neku informaciju. Prisutnošću, to jest neprisutnošću, digitalnog vodenog žiga utvrđuje se autentičnost elektroničkog gradiva. Ovi se žigovi najčešće koriste za obilježavanje

[35] Stančić, H. New Technologies applicable to Document and Records Management: Blockchain // Lligall. Revista Catalana d'Arxivística. Noves perspectives en matèria de gestió documental, 41 (2018), 58.
[36] Volarević, I., Stančić, H. Norme za elektroničke vremenske žigove i mogućnosti njihove primjene u arhivskoj struci. // Arhivi i domovinski rat / Babić, Silvija (ur.). Zagreb: Hrvatsko arhivističko društvo, 2016. str. 429.

zvučnog, slikovnog ili video gradiva. „Nekoliko je različitih primjena digitalnih vodenih žigova, a najčešće su:

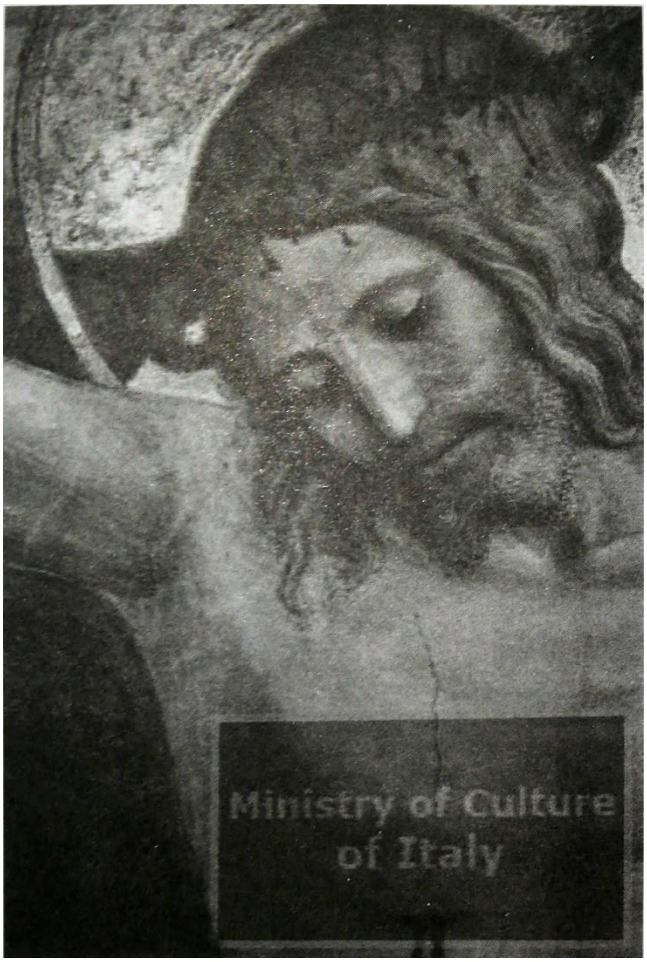
- dokazivanje vlasništva nad nekim sadržajem,
- umetanje podataka o primatelju (engl. *fingerprinting*) kako bi se moglo ustanoviti odakle je potekla eventualna nelegalna kopija,
- provjera autentičnosti i integriteta,
- opisivanje sadržaja (engl. *content labeling*),
- kontrola korištenja,
- zaštita sadržaja.“[37]

Sustav rada digitalnih vodenih žigova ima dva dijela: funkcija umetanja žiga i funkcija detekcije žiga. Prvo pošiljatelj dokumenta pokreće funkciju umetanja žiga koja za ulazne vrijednosti ima izvorni digitalni zapis, žig i korisnički ključ, a za rezultat digitalni zapis s umetnutim digitalnim vodenim žigom. Dalje pošiljatelj prosljeđuje taj dokument, s ključem, primatelju koji je u mogućnosti detektirati žig. Kako bi ga detektirao, primatelj mora pokrenuti funkciju detekcije žiga. Ta funkcija kao ulaznu vrijednost ima korisnički ključ kao i primljeni dokument, a rezultat jest digitalni vodeni žig. Funkcija detekcije potvrđuje ili opovrgava autentičnost dokumenta. Tijekom procesa umetanja podataka o primatelju u žig postoji dodatna ulazna vrijednost – šifra korisnika. Tom se šifrom na jedinstveni način određuje primatelj kojemu je dopušteno koristiti dokument. U tom se slučaju kao izlazna vrijednost funkcije umetanja žiga dobiva digitalni dokument s umetnutim vodenim žigom i podatcima o primatelju. Funkcija detekcije tada kao dodanu vrijednost ima šifru korisnika te opet rezultira potvrđivanjem ili opovrgavanjem prava na korištenje pojedinom primatelju. Ako se detektira postojanje elektroničkog zapisa kod osobe kojoj ono nije izvorno isporučeno ovim se postupkom može utvrditi koji je korisnik ilegalno proslijedio taj zapis. Tim je postupkom korisnik prekršio vlasnička prava institucija u čijem se vlasništvu zapis nalazi.

Digitalni vodeni žigovi mogu biti vidljivi i nevidljivi. Vidljivi žigovi se pojavljuju u obliku logotipa ili poruke na vidljivom ili čujnom području gradiva, a korisnici ih koriste kao informaciju vlasništvu ili dopuštenom korištenju. Pojedine institucije koriste ove žigove za slobodnu distribuciju materijala nedovoljne kvaliteta za profesionalnu upotrebu, a naplaćuju gradivo visoke kvalitete. S druge strane, nevidljivi se žigovi koriste kao dokaz ilegalnog korištenja digitalnih dokumenata. „Po oblikovanju digitalni vodeni žigovi mogu biti krhki

[37] Stančić, H. Digitalizacija. Zagreb : Zavod za informacijske studije, 2009. str. 106.

(engl. *fragile*) i robusni (engl. *robust*).“[38] Krhki nisu postojani kod obrade digitalnih dokumenata, stoga su korisni samo ako nije došlo do izmjene dokumenta. No, ako je, na primjer, postavljeni logotip na nekoj fotografiji izrezan tada više nije moguće dokazati podrijetlo dokumenta. Robusni se žigovi provlače kroz cijeli digitalni zapis i njegovi se dijelovi mogu pronaći i nakon izmjene dokumenta i ako je on uključen u neki novi dokument.



Slika 9: Digitalni vodeni žig[39]

5.4. Ulančani blokovi

Tehnologija ulančanih blokova (engl. *blockchain*) je distribuirana baza podataka zapisa koja pohranjuje *hash* vrijednosti podataka, informacija, transakcija, dokumenata ili zapisa i povezana je s konceptom tehnologije distribuirane glave knjige (engl. *Distributed Ledger Technology - DLT*). Engleski naziv *blokchain* se sastoji od dva izraza - "blok", koji se odnosi na kompletan skup sadržaja, i "lanac", koji se odnosi na međusobno povezivanje blokova. Taj

[38] Ibid., str. 107.
[39] Ibid., str. 107.

lanac raste linearno, a izračun novog bloka u kontekstu kriptovaluta naziva se rudarenjem. Ulančani blokovi se realiziraju putem distribuirane (engl. *peer-to-peer*) mreže u kojoj svako spojeno računalo (čvor) pohranjuje podatke o svim transakcijama (*blockchain* ne pohranjuje podatke, već samo njihove *hash* vrijednosti).

5.4.1. Kriptografski sažetak

Hash ili kriptografski sažetak poruke je jednosmjerna funkcija koja brzo izračunava jedinstveni niz fiksne duljine iz bilo kojih podataka, informacija, dokumenata ili zapisa bilo koje veličine. Karakteristika jednosmjernosti znači da nije moguće obnoviti izvorni dokument znanjem njegova *hash*-a. Izuzetno je teško i gotovo nemoguće stvoriti "preklapanja", tj. imati dva ili više različitih zapisa s istom *hash* vrijednošću. Zbog toga se dobivena *hash* vrijednost naziva i digitalnim otiskom prsta. Ako netko primi docx datoteku i odgovarajuću *hash* vrijednost on može generirati *hash* primljene datoteke i usporediti je s primljenom *hash* vrijednošću. Ako su vrijednosti iste, datoteka nije promijenjena, to jest njezin integritet nije ugrožen. Najpoznatije *hash* funkcije su one koje se temelje na MD5 i SHA algoritmima.

5.4.2. Merkleovo stablo

Hash vrijednosti mogu se grupirati zajedno u jedan *hash*. Stančić daje primjer tvrtke koja kreira brojne dokumente. *Hash* vrijednost se izračunava za svaki dokument. Svakog sata, sve *hash* vrijednosti iz svih dokumenata grupiraju se i na temelju njih se izračunava jedna hash vrijednost tog sata. Na kraju osmosatnog radnog dana, na primjer u ponedjeljak, sve osmosatne vrijednosti raspršivanja se kombiniraju zajedno kako bi se dobila jedna *hash* vrijednost za ponedjeljak.[41] Taj se *hash* zove korijenski hash (engl. root hash ili top hash). Ovaj je pristup prvi put uveo 1980. godine Ralph C. Merkle. Budući da struktura nalikuje stablu naopačke, takva struktura se naziva Merkleovim stablom.

5.4.3. Distribuirani konsenzus

Blockchain koristi distribuiranu mrežu. Distribuirana mreža nema središte jer se sva međusobno povezana računala tretiraju jednako. Ova vrsta mreže nema jedinstvenu kontrolu

[41] Stančić, H. New Technologies applicable to Document and Records Management: Blockchain // Lligall. Revista Catalana d'Arxivística. Noves perspectives en matèria de gestió documental, 41 (2018), 62.

pa stoga, niti jednu jedinu točku napada. *Blockchain* koristi princip distribuiranog konsenzusa (engl. d*istributed consensus*) u kojem svaki sudionik (čvor) bilježi svaki događaj u svojoj knjizi. Konsenzus se koristi kako bi se osiguralo da su sve knjige točne kopije, to jest da su sinkronizirane, i da se utvrdi istina. „Događaj (npr. novčana transakcija ili registracija dokumenta) vrijedi samo ako to potvrdi kvalificirana većina (50% + 1 čvor).“[42]

5.4.4. Povezivanje blokova

Blockchain stvara lanac povezanih blokova. Ranije spomenuti primjer tvrtke može ponoviti postupak napravljen u ponedjeljak za dokumente kreirane svakog sata u utorak. Ovo će rezultirati u dvije *hash* vrijednosti - po jedna za svaki dan. Te dvije vrijednosti bi se dalje mogle povezati kako bi se stvorio novi korijenski *hash* koji objedinjuje pojedinačne *hash* vrijednosti od ponedjeljka i utorka. Svaki novi korijenski hash izračunava se iz *hash* vrijednosti tekućeg dana i prethodnog korijenskog *hasha,* povezujući na taj način korijenske *hash* vrijednosti u jednu. Svaki novi blok označava se digitalnim vremenskim žigom u trenutku stvaranja. „To jamči da su *hash* vrijednosti, to jest podaci, dokumenti ili zapisi postojali u trenutku registracije u *blockchain*.“[43] Postoji nekoliko prednosti koncepta ulančanih blokova. Prije svega, samo se *hash* vrijednosti spremaju u *blockchain.* Stvarni podaci, dokumenti ili zapisi pohranjeni su u institucionalnom repozitoriju ili sustavima za upravljanje dokumentima ili zapisima. Nadalje, svaki dodatni blok pojačava prethodne, budući da su blokovi povezani i svaki novi blok ovisi o vezama prethodnih blokova. Konačno, izmjena bilo kojeg bloka u lancu poništava sve naredne blokove.

5.4.5. Korištenje ulančanih blokova u očuvanju digitalnih zapisa

Upravljanje digitalnim zapisima poboljšava produktivnost poslovanja i organizacijsku učinkovitost. Najčešće korištene funkcije upravljanja dokumentima su praćenje koraka u poslovnom procesu, provjera promjena, strukture dokumenta i sadržaja te pojednostavljena i pouzdana razmjena dokumenata. *Blockchain* bi mogao biti koristan u nekoliko aspekata procesa upravljanja dokumentima. Na primjer, kad god se stvori nova verzija dokumenta, ona se može registrirati u *blockchain*. Na taj način, budući da se svaki novi blok vremenski označava, postaje jasno koja je verzija dokumenta kada nastala, jesu li izvršene promjene

[42] Ibid., str. 62.
[43] Ibid., str. 64.

strukture dokumenta, a sadržaj može biti praćen i provjeren ako je to potrebno. Nadalje, tijekom poslovanja dokumenti se često šalju drugim strankama. Registracija u *blockchain* mogla bi pružiti potreban dokaz da neki dokument nije bio mijenjan.

S druge strane, dokumenti su često digitalno potpisani ili im je dodan digitalni pečat. Jednom kada dokumenti postanu zapisi više se ne smiju mijenjati, a tijekom upravljanja zapisima i arhiviranja njihova bi autentičnost, integritet, pouzdanost i upotrebljivost trebala ostati netaknuta, dok bi neki od njih također trebali sačuvati karakteristike sigurnosti i povjerljivost. Certifikati koje se koriste u digitalnom potpisu istječu za dvije do pet godina, što arhivistima donosi situaciju u kojoj se valjanost digitalnog potpisa više ne može potvrditi. U sklopu projekta InterPARES Trust razvilo se TRUSTER (engl. *Model for Preservation of Trustworthiness of the Digitally Signed, Timestamped and/or Sealed Digital Records*) VIP (engl. *Validity Information Preservation*) rješenje pod nazivom TrustChain.[44] Istražuju se mogućnosti uporabe brisanja vremenskog označivanja i *blockchain* tehnologije za dugoročno čuvanje digitalno potpisanih zapisa. TrustChain je model temeljen na *blockchainu* koji se može koristiti za registriranje podataka o valjanosti digitalnih vremenskih certifikata iz digitalnih potpisa u *blockchain* u trenutku zaprimanja digitalno potpisanih ili zapečaćenih zapisa u arhiv, dok su digitalni certifikati još uvijek važeći. Kasnije, „kad istekne rok valjanosti digitalnih certifikata, moguće je:

1) potvrditi da je digitalni certifikat bio valjan u trenutku zaprimanja,
2) potvrditi da se zapis nije promijenio (ponovnim izračunom *hash* vrijednosti i usporedbom s registriranim i onim koji je pronađen u digitalnom potpisu),
3) kada je prije navedeno (pod 1 i 2) točno, na zapis se gleda kao da digitalni certifikat još uvijek vrijedi.“[45]

[44] Stančić, H. Model for Preservation of Trustworthiness of the Digitally Signed, Timestamped and/or Sealed Digital Records (TRUSTER Preservation Model) (EU31). InterPARES Trust. 2018. Dostupno na: https://interparestrust.org/assets/public/dissemination/TRUSTERPreservationModel(EU31)-Finalreportv_1_3.pdf

[45] Ibid., str. 65.

6. Zaključak

Sve informacijske ustanove kao jednu od glavnih zadaća imaju pohranjivanje i očuvanje gradiva. Kako bi u tome bile što uspješnije, informacijski su stručnjaci razvili niz tehnologija, a pritom se oslanjaju na relevantne norme. Kada je riječ o arhiviranju digitalnih zapisa zahtijeva se kontinuirano i pravovremeno očuvanje gradiva iz razloga što je digitalno gradivo podložnije propadanju zbog brzog razvoja informacijske tehnologije. Arhivski stručnjaci suočavaju se s izazovom dugoročnog pohranjivanja i zaštite integriteta. Mnogo je dilema oko izbora medija za pohranu, standarda koji će se koristiti za metapodatke te koja je metoda zaštite integriteta najbolja opcija za pojedinu jedinicu građe i zašto. Kada je riječ o izboru medija za pohranu, stručnjaci mogu odabrati između sljedećih sustava: izravni, poluizravni, hijerarhijski, neizravni, mrežna pohrana, mreža za pohranu te pohrane u računalnome oblaku. Pritom građu moraju opisati nekim od metapodatkovnih standarda za opis kao što su ISAD(G), EAD, ISAAR(CPF), ISDIAH ili ISDF. Također moraju odabrati potrebne metode za osiguranje integriteta među kojima se nalaze digitalni potpis, digitalni vodeni žig, digitalni vremenski žig i tehnologija ulančanih blokova. Svi su ovi postupci potrebni kako bi se zapisi mogli arhivirati i očuvati.

Jedan od prvih problema s kojima će se susresti arhivisti u procesu arhiviranja digitalnih zapisa jest vjerodostojnost. Iznimno je važno utvrditi je li zapis koji se pohranjuje vjerodostojan kako bi se za njega sa sigurnošću moglo tvrditi je li bio mijenjan, kako je i gdje nastao, tko mu je autor itd. Također je važno zapis kvalitetno opisati, to jest pridodati mu metapodatke. Kako bi oni bili kvalitetni nužno je koristiti neki od dostupnih standarada metapodataka. Iako živimo u dobu gdje je Internet vrlo veliki dio svakodnevnog života to ne znači da je on jedino rješenje za pohranu zapisa. Iako je među najstarijim medijima za pohranu, magnetska traka i danas je vrlo popularni izbor. Ne postoji točan ili netočan odabir kada se radi o izboru sustava za pohranu. Svaki od njih služiti će svrsi za koju je potreban, a na arhivistima je da odluče koji je sustav njihovoj instituciji najoptimalniji i najisplativiji. Nadalje, kako bi uopće došlo do čina pohrane zapisa potrebno mu je osigurati integritet. Metodama osiguranja integriteta teži se omogućavanju pristupa sadržaju i osiguravanju očuvanja vjerodostojnosti dokumenta. Tada je korisnik siguran da je zapis ono što se tvrdi da jest te da nije bio mijenjan ili kompromitiran.

Proces arhiviranja digitalnih zapisa detaljan je te zahtijeva praćenje određenih pravila, a sve to kako bi se korisniku pružila što bolja i kvalitetnija usluga. Ovaj proces nikada ne završava jer nije bitno samo pohraniti zapis, bitno ga je i dugoročno čuvati.

Literatura

Brzica, Hrvoje; Herceg, Boris; Stančić, Hrvoje Long-term Preservation of Validity of Electronically Signed Records // Information Governance / Gilliland, Anne; McKemmish, Sue; Stančić, Hrvoje; Seljan, Sanja; Lasić-Lazić, Jadranka (ur.). Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2013. str. 147-158. Dostupno na: https://infoz.ffzg.hr/INFuture/2013/papers/403%20Brzica,%20Herceg,%20Stancic,%20LTP%20of%20Validity%20of%20Electronically%20Signed%20Records.pdf

Cole, W. T. Creating a Framefork of Guidance for Building Good Digital Collections, 2002. Dostupno na: http://firstmonday.org/ojs/index.php/fm/article/view/955/876 (17.06.2020.).

Convergence of Networking and Storage. // Chelsio Communications. Dostupno na: https://www.chelsio.com/convergence_of_networking_and_storage/ (1.09.2020.).

Duranti, L. The Trustworthiness of Digital Records. // Ppt prezentacija : International Congress on Digital Records Preservation, 16. travanj 2010.

Duranti, L., et al. Records and Archives in the Cloud. // The Canadian Journal of Information and Library Science. 2015.

Googleova najsnažnija zaštita za one kojima je najpotrebnija.// Google Program Napredne Zaštite. Dostupno na: https://landing.google.com/advancedprotection/?utm_source=Chrome&utm_medium=ChromeSecuritySettings&utm_campaign=ChromeSettings (14.08.2020.).

Houghton, B. Preservation Challenges in the Digital Age. // D-Lib Magazine. 22(7/8), srpanj/kolovoz 2016. Dostupno na: http://www.dlib.org/dlib/july16/houghton/07houghton.html (17.06.2020.).

InterPARES 2 Project: International Research on Permanent Authentic Records in Electronic Systems. // InterPARES Trust. Dostupno na: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_ontology.pdf (1.09.2020.).

ISO 15489-1: Informacije i dokumentacija - Upravljanje spisima - 1. dio: Općenito. Prijevod za internu upotrebu.

Ivanović, J. Sheme metapodataka u upravljanju dokumentima. // Arivski vjesnik. god. 44 (2001), str. 103-121.

Lantz, M. Why the Future of Data Storage is (Still) Magnetic Tape, 28.08.2018. // IEEE Spectrum. Dostupno na: https://spectrum.ieee.org/computing/hardware/why-the-future-of-data-storage-is-still-magnetic-tape (17.08.2020.).

Library of Congress Recommended Formats Statement 2019-2020, Dostupno na: https://www.loc.gov/preservation/resources/rfs/TOC.html

Mihaljević, M., Mihaljević, M. i Stančić, H. (2015). trustworthiness. U: Arhivistički rječnik: HRVATSKO-ENGLESKI/ENGLESKO-HRVATSKI. Zagreb: Zavod za informacijske studije Odsjeka za informacijske i komunikacijske znanosti Filozofskog fakulteta Sveučilišta u Zagrebu, str. 155.

Natarajan, R. RAID 0, RAID 1, RAID 5, RAID 10 Explained with Diagrams, 10.8.2010. // The Geek Stuff. Dostupno na: https://www.thegeekstuff.com/2010/08/raid-levels-tutorial/ (17.08.2020.)

Natarajan, R. RAID 2, RAID 3, RAID 4, RAID 6 Explained with Diagram, 21.11.2011. // The Geek Stuff. Dostupno na: https://www.thegeekstuff.com/2011/11/raid2-raid3-raid4-raid6/ (17.08.2020.).

National Information Standards Organization. Understanding metadata. 2004. Dostupno na: https://www.lter.uaf.edu/metadata_files/UnderstandingMetadata.pdf (17.06.2020.).

Ordel, S. What is the failure rate of Solid-State Drives (SSD), 1.03.2012. // StackExchange. Dostupno na: https://skeptics.stackexchange.com/questions/7084/what-is-the-failure-rate-of-solid-state-drives-ssd (1.09.2020.).

Rajh A., Stančić H., Romčević B., Vitaljić M. Koncept rješenja za osiguranje i očuvanje vjerodostojnosti zapisa u upravnim organizacijama prilikom razvoja državnog računalnog oblaka i državnog digitalnog arhiva. Arhivski vjesnik br. 61 (2018), str. 69-88.

Stančić, H. Arhiviranje digitalnih dokumenata. // 4. seminar Arhivi, knjižnice i muzeji. Mogućnosti suradnje u okruženju globalne informacijske infrastrukture. Zagreb, 2001.

Stančić, H. Digitalizacija. Zagreb : Zavod za informacijske studije, 2009.

Stančić, H. Long-term Preservation of Digital Signatures. // Technical and field related problems of traditional and electronic archiving / Gostenčnik, Nina (ur.). Maribor: Pokrajinski arhiv, 2016. str. 481-491.

Stančić, H. Model for Preservation of Trustworthiness of the Digitally Signed, Timestamped and/or Sealed Digital Records (TRUSTER Preservation Model) (EU31). InterPARES Trust. 2018. Dostupno na: https://interparestrust.org/assets/public/dissemination/TRUSTERPreservationModel(EU31)-Finalreportv_1_3.pdf

Stančić, H. New Technologies applicable to Document and Records Management: Blockchain // Lligall. Revista Catalana d'Arxivística. Noves perspectives en matèria de gestió documental, 41 (2018), str. 56-72

Uredba (EU) br. 910/2014 Europskog parlamenta i Vijeća od 23. srpnja 2014. o elektroničkoj identifikaciji i uslugama povjerenja za elektroničke transakcije na unutarnjem tržištu i stavljanju izvan snage Direktive 1999/93/EZ. Dostupno na: https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX%3A32014R0910 (14.08.2020.).

Volarević, I., Stančić, H. Norme za elektroničke vremenske žigove i mogućnosti njihove primjene u arhivskoj struci. // Arhivi i domovinski rat / Babić, Silvija (ur.). Zagreb: Hrvatsko arhivističko društvo, 2016. str. 425-435.

Popis slika

Arhiviranje digitalnih zapisa

Sažetak

U današnje vrijeme sve više dokumenata izvorno nastaje u digitalnom obliku. Digitalni informacijski objekt predstavlja onaj objekt koji je nastao uz pomoć informacijske tehnologije, bez obzira je li to njegov izvorni oblik ili je riječ o gradivu u klasičnom obliku koje je preneseno u elektroničku okolinu postupkom digitalizacije. Problemi se pojavljuju kod očuvanja digitalnih informacijskih objekata i njihove vjerodostojnosti. U ovome radu pobliže se objašnjava proces arhiviranja digitalnih zapisa sagledan s tehnološkog aspekta. U radu se analiziraju mediji na kojima se zapisi pohranjuju, metapodaci za opis gradiva te metode osiguranja integriteta arhiviranih zapisa. U tom kontekstu se opisuju digitalni potpis, digitalni vodeni žig, digitalni vremenski žig i tehnologija ulančanih blokova. Metapodaci nose glavnu ulogu u opisu arhivirane jedinice gradiva, dok se integritet zapisa, cjelovitost sadržaja i vjerodostojnost arhiviranih zapisa mogu osigurati uporabom digitalnog potpisa i digitalnog vremenskog žiga.

Ključne riječi: digitalni zapis, digitalni vremenski žig, digitalni potpis, ulančani blokovi, metapodaci, medij za pohranu

Archiving digital records

Summary

Nowadays, more and more documents are digitally born. A digital information object is an object created with the help of information technology, regardless of whether it is its original form or it is a material in the analogue form that has been transferred to the digital environment by the process of digitization. Problems arise with the preservation of digital information objects and their trustworthiness. This thesis explains in more detail the process archiving digital records. It covers the media on which records are stored, their metadata and methods of ensuring the integrity of archived records. Integrity assurance methods include digital signature, digital watermark, digital timestamp, and blockchain. Metadata plays a major role in description of the archived records, while their integrity and authenticity may be ensured by using digital signature and digital timestamp applied to each archived record.

Key words: digital record, digital timestamp, digital signature, blockchain, metadata, storage medium