



PAM Pokrajinski
arhiv
Maribor

Moderna
arhivistika

Časopis arhivske teorije in prakse
Journal of Archival Theory and Practice

Letnik 5 (2022), št. 2 / Year 5 (2022), No. 2

Maribor, 2022

Prejeto / Received: 31. 08. 2022

1.01 Izvirni znanstveni članek

1.01 Original Scientific article

<https://doi.org/10.54356/MA/2022/ORPM2519>

PRETVORNIKI IN VALIDATORJI – PROGRAMSKA ORODJA ZA PRETVARJANJE GRADIVA V FORMATE ZA DOLGOROČNO HRAMBO IN VALIDACIJO TEH PRETVORB

Doc. dr. Tatjana HAJTNIK

Arhiv Republike Slovenije, Ljubljana, Slovenija
tatjana.hajtnik@gov.si

Klavdija KRIVEC

Arhiv Republike Slovenije, Ljubljana, Slovenija
klavdija.krivec@gov.si

Izvleček:

Skladno s slovenskimi predpisi na področju zagotavljanja dolgoročnega ohranjanja gradiva v digitalni obliki (v nadaljevanju digitalnega gradiva) je treba le-to pretvarjati v formate, ki so primerni za dolgoročno hrambo. Za to potrebujemo posebna programska orodja, ki jih je sicer veliko na voljo, tako plačljivih kot brezplačnih. Ker vsaka pretvorba povzroči izgubo določenega dela informacij v okviru posamezne enote digitalnega gradiva, smo v raziskavi želeli poiskati odgovore na ključno vprašanje, katera programska orodja uporabiti, da bo teh izgub čim manj. V raziskavo smo vključili tudi programska orodja, s katerimi je mogoče preveriti ustreznost pretvorb digitalnega gradiva iz enega v drugi format. Ključni cilj raziskave je bil poiskati ustrezna programska orodja za pretvarjanje posameznih vrst digitalnega gradiva in preverjanje teh pretvorb, s tem pa tudi olajšati postopke pretvorbe, predvsem ustvarjalcem arhivskega digitalnega gradiva.

Ključne besede:

projekt e-ARH.si, programska orodja, pretvorniki, validatorji, format, dolgoročno ohranjanje digitalnega gradiva

Abstract:

Converters and Validators: Software Tools for Converting to Formats for Long-term Digital Preservation and Validation

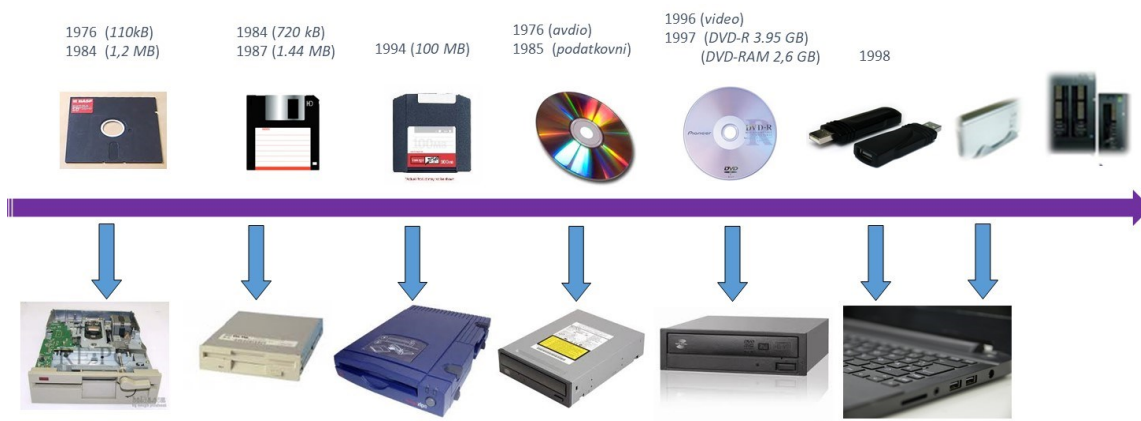
In accordance with Slovenian regulations in the field of ensuring the long-term preservation of records in digital form (hereinafter referred to as digital records), it must be converted into formats suitable for long-term preservation. For this we need special software tools, of which there are many available, both paid and free. Since every conversion results in the loss of a certain part of the information within a single unit of digital records, we wanted to find answers to the key question of which software tools to use in order to minimize these losses. In the research, we also included software tools that can be used to check the suitability of conversions of digital records from one format to another. The key goal of the research was to find suitable software tools for converting individual types of digital records and verifying these conversions, thereby facilitating the conversion process especially for creators of archival digital records.

Key words:

project e-ARH.si, program tools, converter, validator, format, long-term presevation of digital records.

1. Uvod

Potreba po ohranjanju zapisov, tako tistih, ki nastajajo izvirno v digitalni obliki, kot tistih, ki nastanejo z digitizacijo¹, ni več vprašanje, je dejstvo. Ostaja pa vprašanje, kako te zapise ohranjati, da bodo ves potreben čas hrambe dostopni, uporabni, celoviti in avtentični, kar so ključne zahteve oz. načela² za varno hrambo gradiva v digitalni obliki. Seveda ob dejstvu, da sta dostopnost do zapisov v digitalni obliki in njihova predstavitev (uporabnost) tesno povezani in odvisni v prvi vrsti od tehnologije, ta pa se zadnja leta hitro spreminja – tako nosilci zapisa (slika 1) kot tudi oblike zapisa (formati). Slednji so tesno povezani s programsko opremo, prav tako podvrženo hitrim spremembam, s tem pa povečanju tveganja, da zapisi čez čas ne bodo dostopni in uporabni.



Slika 1: Spreminjanje nosilcev digitalnih zapisov skozi čas (vir: lasten)

V Sloveniji smo področje dolgoročnega ohranjanja digitalnih zapisov uredili že leta 2006 s sprejetjem Zakona o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA), ki je bil nato leta 2014 noveliran. Zakon skupaj s pripadajočimi podrejenimi predpisi³ med drugim ureja tudi zajem in hrambo gradiva v digitalni obliki, v nadaljevanju bomo govorili o *digitalnem gradivu*. Omenjeni predpisi določajo, da mora biti digitalno gradivo, kadar ga je treba hraniti dolgoročno⁴, shranjeno na primernih nosilcih in v ustreznih formatih⁵. Slednji namreč običajno ob nastanku digitalnega gradiva niso

¹ Digitizacija je proces pretvorbe podatkov iz analogne v digitalno obliko, ki jo je mogoče enostavno brati in obdelati z računalnikom.

² Načela varne hrambe digitalnega gradiva so določena z ZVDAGA, in sicer morata hramba gradiva v digitalni obliki oz. reproduciranje njegove vsebine zagotavljati njegovo trajnost (4. člen), gradivo pa mora biti ves čas zahtevane hrambe dostopno (6. člen), uporabno (3. člen), celovito (5. člen) in avtentično (27. člen).

³ Ključna sta Uredba o varstvu dokumentarnega in arhivskega gradiva (UVDAG) in Pravilnik o enotnih tehnoloških zahtevah za zajem in hrambo gradiva v digitalni obliki (PETZ).

⁴ Hramba digitalnih zapisov nad pet let se šteje kot dolgoročna hramba (ZVDAGA, 2. člen).

⁵ ZVDAGA uporablja besedno zvezo oblika zapisa, kar je v informatiki in računalništvu bolj znano kot format, opredeljuje pa jo kot »organizacijske in tehnološke značilnosti zapisa, ki določajo, kako je vsebina zapisana, hranjena in prikazana v procesu hrambe« (2. člen). Format je odvisen od vrste zapisa (npr.

primerni za dolgoročno hrambo, zato predpisi določajo, da je treba digitalno gradivo obvezno⁶ pretvarjati v ustrezne formate, ki so primerni za dolgoročno hrambo. Hkrati isti predpisi sicer določajo kriterije⁷, ki jih morajo izpolnjevati formati, da so primerni za dolgoročno hrambo, vendar le na splošno. Zato Arhiv Republike Slovenije (v nadaljevanju Arhiv RS) za posamezne vrste digitalnega gradiva, skladno s *Pravilnikom o enotnih tehnoloških zahtevah za zajem in hrambo gradiva v digitalni obliki*⁸, objavlja na svoji spletni strani⁹ seznam najpogosteje uporabljenih formatov, ki so primerni za njegovo dolgoročno hrambo oz. ki jih Arhiv RS priporoča ali označuje kot sprejemljive.

Da lahko pretvorimo digitalno gradivo iz enega v nek drug format, pa potrebujemo posebno programsko opremo. Ta je lahko vgrajena v samo aplikativno opremo (npr. urejevalnik besedil) ali pa gre za samostojno namensko programsko opremo – **pretvornik** (angl. converter). Ker pa se pri pretvorbah digitalnega gradiva lahko izgubi del njegovih informacij, je treba pravilnost vsake pretvorbe preveriti še s posebno programsko opremo – **validatorjem** (angl. validator). Slovenski predpisi ne opredeljujejo posebej, kateri pretvorniki oz. validatorji so primerni za uporabo, hkrati pa tudi vemo, da jih je danes na spletu dostopnih cel kup¹⁰.

Za odločitev o uporabi ustreznih pretvornikov oz. validatorjev v postopku pretvorbe si je slovenska javna arhivska služba (SJAS) v okviru projekta vzpostavitve slovenskega javnega e-arhiva (Projekt e-ARH.si: ESS 2016 – 2020/2021¹¹; v nadaljevanju projekt) zadala nalogo¹², da bo to področje raziskala in na trgu poiskala pretvornike in validatorje, ki jih bo lahko, v prvi vrsti ustvarjalcem arhivskega gradiva, priporočila za uporabo.

1.1 Teoretična izhodišča in raziskave

Ključno izhodišče pri izvedbi zastavljene naloge sta bili zakonski določili:

1. Pretvarjanje digitalnega gradiva v formate, ki so primerni za dolgoročno hrambo, je obvezno¹³.
2. Pretvorba mora biti zanesljiva¹⁴, kar pomeni, da mora zagotavljati:
 - ohranjanje *celovitosti, uporabnosti, avtentičnosti* izvirnih digitalnih zapisov gradiva,
 - primerno število kontrol pravilnosti in kakovosti pretvorbe z odpravo napak oziroma odstopanj,

dokumenta), njegove vsebine in načina uporabe. Določeni programski opremi omogoča interpretirati podatke v dokumentu; je način predstavitve teh podatkov (Hajtnik, 2016).

⁶ ZVDAGA, 11. člen in 28. člen.

⁷ Uredba o varstvu dokumentarnega in arhivskega gradiva /UVDAG/, 42. člen.

⁸ PETZ, 35. člen.

⁹ Dostopna na naslovu: <https://www.gov.si/assets/organi-v-sestavi/Arhiv-RS/Zakonodaja-2019/Pravilnik-o-enotnih-tehnoloskih-zahtevah-za-zajem-in-hrambo-gradiva-v-digitalni-obliki/Seznam-oblik-zapisa-za-dolgorocno-hrambo-ver-1.0.pdf>.

¹⁰ Če npr. v Googlovo iskalno polje vnesemo iskalni pojem »format converter«, dobimo 119.000.000 zadetkov (17. 8. 2022).

¹¹ Projekt e-ARH.si: ESS 2016 – 2020/2021 je delno (80 %) financiral evropski kohezijski sklad.

¹² Naloga se je izvajala v okviru Kompetenčnega centra 2 (KC2.5 Nabava programske opreme za paketno pretvarjanje e-arhivskega gradiva v format za dolgoročno shrambo (pretvorniki, validatorji)).

¹³ ZVDAGA, 11. člen in 28. člen.

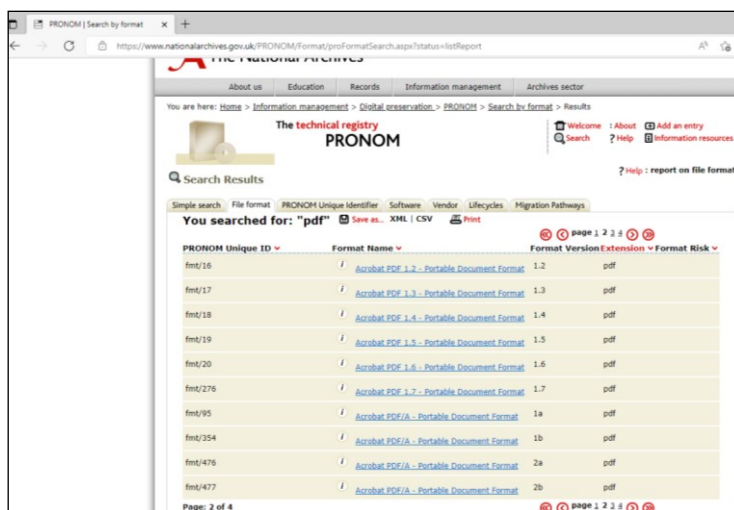
¹⁴ ZVDAGA, 10. člen.

- posebej in jasno ločeno od izvirne vsebine shranitev dodanih vsebin, pomembnih opomb in podatkov glede postopka pretvorbe in glede izvirnega gradiva,
- *dokumentiranost*, pomeni shranitev primerne obsega dokumentacije o pretvorbi, in
- izpolnjevanje drugih pogojev, ki jih dodatno predpiše Vlada Republike Slovenije.

Izhodišče oz. raziskovalna vprašanja:

Obstaja veliko programskih orodij za pretvarjanje formatov – ali so vsa tudi ustrezna? Ali bo rezultat pretvorbe res format, ki je primeren za dolgoročno hrambo?

Izhajali smo iz praktičnih izkušenj pri uporabi različnih pretvornikov v preteklosti, ko smo želeli pretvarjati besedilne dokumente formatov .doc in .docx v arhivski format PDF/A. Na spletu smo poiskali programska orodja, ki so kot vhodni format sprejemala formata .doc oz. .docx, kot izhodni format pa ponujala format .pdf. Že sam vpogled v brezplačno dostopen register formatov PRONOM¹⁵ nam pove, da obstaja format .pdf v 39 unikatnih identifikacijskih številkah fmt (slika 2), različice se razlikujejo glede na razvoj standarda, s tem povezano specifikacijo in namen uporabe. Vsi imajo končnico (angl. extension) .pdf, iz katere pa ni mogoče razbrati, ali je to tudi arhivski format PDF/A. To ugotovimo šele, ko poznamo enolični identifikator formata (PRONOM Unique ID) in preko tega verzijo in opis formata. Te informacije pa nam pretvorniki, ki jih najdemo na spletu, ne ponudijo.



The screenshot shows the PRONOM website search results for the query 'pdf'. The page displays a table of search results with columns for PRONOM Unique ID, Format Name, Format Version, Extension, and Format Risk. The results list various versions of Acrobat PDF and Acrobat PDF/A formats.

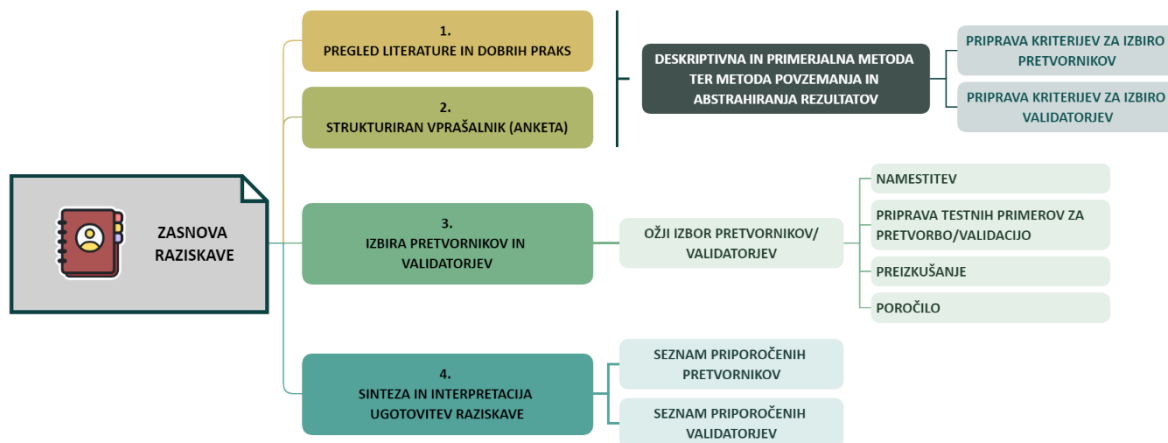
PRONOM Unique ID	Format Name	Format Version	Extension	Format Risk
fmt/16	Acrobat PDF 1.2 - Portable Document Format	1.2	pdf	
fmt/17	Acrobat PDF 1.3 - Portable Document Format	1.3	pdf	
fmt/18	Acrobat PDF 1.4 - Portable Document Format	1.4	pdf	
fmt/19	Acrobat PDF 1.5 - Portable Document Format	1.5	pdf	
fmt/20	Acrobat PDF 1.6 - Portable Document Format	1.6	pdf	
fmt/276	Acrobat PDF 1.7 - Portable Document Format	1.7	pdf	
fmt/95	Acrobat PDF/A - Portable Document Format	1a	pdf	
fmt/354	Acrobat PDF/A - Portable Document Format	1b	pdf	
fmt/476	Acrobat PDF/A - Portable Document Format	2a	pdf	
fmt/477	Acrobat PDF/A - Portable Document Format	2b	pdf	

Slika 2: Informacije o formatu .pdf v registru PRONOM (vir: <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>)

¹⁵ Register formatov PRONOM je razvil in ga vzdržuje Nacionalni arhiv Velike Britanije. Register je javno dostopen in vsebuje podatke o praktično vseh poznanih formatih na svetu ter pripadajoče opise (The National Archives, 2011a). Na voljo je vsem, ki potrebujejo zanesljiv in verodostojen informacijski vir o formatih ter programski opremi, njenih življenjskih ciklih in tehničnih specifikacijah, ki jih podpira. Dostopen je na naslovu <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx> (15. 8. 2022).

1.2 Zasnova in metodološki pristop k raziskavi

Skladno s projektnim načrtom projekta smo oblikovali delovno skupino KC2.5¹⁶ in pripravili natančen načrt izvedbe raziskave, ki je za posamezno vrsto gradiva vključeval aktivnosti, kot je razvidno iz slike 3:



Slika 3: Zasnova raziskave in metodološki pristop (vir: lasten)

V raziskavi smo uporabili več znanstvenih metod, ključni sta pregled literature in dobrih praks (deskriptivna in primerjalna metoda ter metoda povzemanja in abstrahiranja rezultatov študij v znanstveni in strokovni literaturi predvsem s področja informacijskih znanosti in arhivistike), kakor tudi uporaba metode strukturiranega vprašalnika (ankete). Na podlagi teh metod smo določili kriterije za izbiro programskih orodij za pretvarjanje digitalnih zapisov v formate za dolgoročno hrambo in preverjanje uspešnosti pretvorb, s pomočjo metod preizkušanja, analize in sinteze pa oblikovali končen nabor priporočenih programskih orodij za posamezne vrste digitalnega gradiva.

2. Rezultati raziskave

2.1 Pregled literature in dobrih praks

V prvi fazi raziskave smo pregledali:

- primarno in sekundarno strokovno literaturo s področja informacijskih znanosti in arhivistike, ki obravnava problematiko pretvornikov in validatorjev;
- poročila obiskov konferenc in drugih strokovnih dogodkov, ki smo se jih udeležili v času trajanja projekta e-ARH.si: ESS 2016 – 2020/2021. Naloga članov projekta je bila med drugim tudi, da so na strokovnih dogodkih, ki so se jih udeleževali, iskali informacije o uporabi pretvornikov in validatorjev;

¹⁶ Člani delovne skupine KC2.5: Klavdija Krivec, Tatjana Hajtnik, Tine Završnik, Boris Domajnko, Antun Smerdel, Primož Tanko, Robert Merič (vsi iz Arhiva RS), Damjan Lindental (Zgodovinski arhiv na Ptuj), Urh Kopitar (Zgodovinski arhiv Celje), Erik Pregelj (Pokrajinski arhiv Nova Gorica), Vladimir Drobnjak, Ivančica Sabadin (oba Pokrajinski arhiv Koper), Miroslav Novak, Boštjan Zajšek (oba Pokrajinski arhiv Maribor), Jože Glavič (Zgodovinski arhiv Ljubljana).

- spletne portale kot so COPTR¹⁷, InterPares¹⁸, Digital Preservation Coalition¹⁹, kjer so objavljena različna programska orodja in rešitve s področja dolgoročne hrambe digitalnih zapisov;
- spletno dostopne registre pretvornikov in validatorjev (npr. ConversionSoftwareRegistry²⁰).

Ugotovili smo, da obstaja množica različnih pretvornikov in validatorjev. S problematiko kriterijev za izbiro ustreznih, zaupanja vrednih orodij se je začela ukvarjati Hajtnikova (2016) v okviru doktorske disertacije. Na osnovi pregleda literature, izvedene ankete med nacionalnimi arhivi držav članic EU in študije primerov pretvorb je razvila model kriterijev za izbiro pretvornikov (MkiPretvornik), ki smo ga uporabili kot izhodišče v naši raziskavi. Za razliko od Hajtnikove, ki se je osredotočila na besedilne dokumente, smo v naši raziskavi iskali pretvornike in validatorje še za druge vrste digitalnega gradiva, in sicer za slikovno in avdiovizualno gradivo, podatkovne baze, e-pošto, spletne strani in prostorske podatke.

Pri tem smo upoštevali tudi dejstvo, da je izbira ustreznih programskih orodij za pretvorbo in validacijo na eni strani izjemnega pomena, saj želimo orodja, pri uporabi katerih bodo izgube informacij pretvorjenega digitalnega gradiva ničelne ali kar najmanjše. Po drugi strani pa smo upoštevali tudi, da so programska orodja lahko *lastniška*, *brezplačna* ali *odprtokodna* in omogočajo različne funkcionalnosti ter so različno cenovno dostopna. Ne gre namreč zanemariti, da se v javni upravi, katere pretežen del so tudi ustvarjalci arhivskega gradiva, za izpolnitev zakonskih določil pogosto uporabljajo rešitve, ki cenovno niso zahtevne.

2.2 Strukturiran vprašalnik (anketa)

Za zbiranje podatkov smo pripravili strukturiran anketni vprašalnik z 49 vprašanji znotraj treh različnih področij raziskave. Pripravljen je bil v angleškem jeziku. Proučevano populacijo (reprezentativni vzorec) so predstavljali državni arhivi držav članic EU (28), zaradi siceršnjega sodelovanja na področju dolgoročne e-hrambe pa smo v raziskavo vključili še tri državne arhive izven EU – Islandije, Norveške in Švice. Ankete smo pošiljali po e-pošti na kontaktne naslove državnih arhivov.

Zbrane ankete smo analizirali s pomočjo programa MS Excel in s pomočjo metode interpretacije povzeli ugotovitve o stanju na raziskovalnem področju.

Za potrebe raziskave, ki je predstavljena v tem prispevku, bomo uporabili samo rezultate, ki izhajajo iz III. poglavja ankete in se nanašajo na pretvornike in validatorje.

¹⁷ Community Owned digital Preservation Tool Registry (COPTR), dostopen na naslovu https://coptr.digipres.org/index.php/Main_Page.

¹⁸ Na naslovu <http://interpares.org/>.

¹⁹ Na naslovu: <https://www.dpconline.org/>.

²⁰ Na naslovu <https://www.archives.gov/applied-research/software-downloads/csr.html>.

2.2.1 Metoda analize in interpretacija odgovorov

2.2.1.1 Predstavitev anketirancev

	Države članice EU	Arhiv	DA	NE
1	Avstrija	Austrian State Archives		✓
2	Belgija	Belgian State Archives	✓	
3	Bolgarija	Archives State Agency		✓
4	Čiper*	State Archives of Cyprus		✓
5	Češka	National Archives Czech Republic	✓	
6	Danska	Danish National Archives	✓	
7	Estonija	The National Archives of Estonia	✓	
8	Finska	National Archives Service of Finland		✓
9	Francija	The National Archives of France		✓
10	Grčija	General State Archive		✓
11	Hrvaška	Croatian State Archives	✓	
12	Irska	The National Archives of Ireland		✓
13	Italija	Directorate General of Archives	✓	
14	Latvija	The National Archives of Latvia	✓	
15	Litva	Office of the Chief Archivist of Lithuania	✓	
16	Luksemburg	National Archives Luxembourg	✓	
17	Madžarska	National Archives of Hungary		✓
18	Malta	National Archives Malta		✓
19	Nemčija	Bundesarchiv	✓	
20	Nizozemska	The National Archives of the Netherlands		✓
21	Poljska	Head Office of State Archives	✓	
22	Portugalska	Institute of the National Archives Portugal	✓	
23	Romunija	National Archives of Romania		✓
24	Slovaška	Ministry of the Interior of the Slovak Republic, Archives and Registries Department		✓
25	Slovenija	The National Archives of Slovenia	✓	
26	Španija	Spanish State Archives	✓	
27	Švedska	National Archives of Sweden	✓	
28	Združeno kraljestvo	The National Archives		✓

Tabela 1: Anketiranci v arhivih držav članic EU

* Odgovorili, da Ciper še ni dosegel stopnje tehnološkega napredka, ki je potrebna, da bi vprašanja, raziskana v raziskavi, veljala v njihovi instituciji.

	Države nečlanice EU	Arhiv	DA	NE
29	Islandija	National Archives of Iceland	✓	
30	Norveška	National Archives of Norway	✓	
31	Švica	Staatsarchiv Zürich		✓

Tabela 2: Anketiranci v arhivih izven EU

Anketo smo poslali 28 državnim arhivom držav članic EU (ko smo pošiljali ankete, je bila Velika Britanija še članica EU) in trem drugim državnim arhivom (Islandija, Norveška in Švica) – skupaj torej 31 anket. Na anketo je odgovorilo 17 državnih arhivov, kar predstavlja 54,84 % reprezentativnega vzorca (tabela 1 in tabela 2).

Anketo smo pošiljali 4. marca 2020 z informacijo, da podatke zbiramo do vključno 13. aprila 2020. Ker pa je bila anketa prvič odposlana praktično že v času pandemije covid-19 in smo začeli dobivati odzive o odsotnosti anketirancev, smo naslovnikom še po tem datumu (do 25. junija 2020) pošiljali tudi opomnik s prošnjo za izpolnitev ankete. Z nekaterimi anketiranci smo stopili v stik tudi osebno in preko e-pošte.

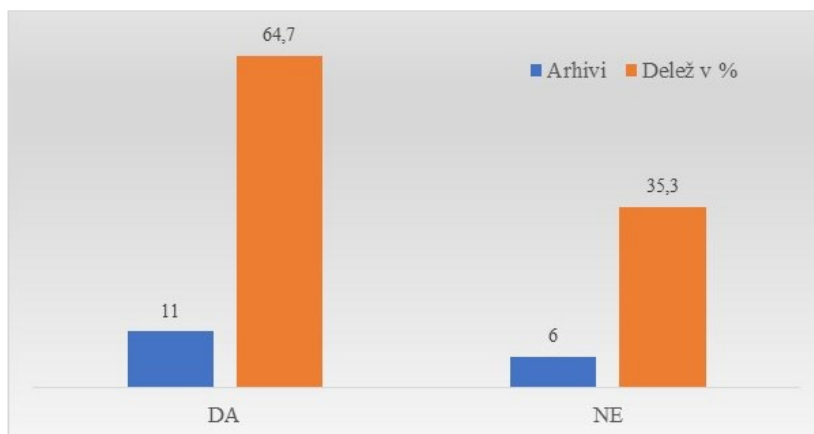
2.2.2 Predstavitev rezultatov ankete

Kot smo omenili, so se vprašanja v anketi nanašala na štiri vsebinska področja, v nadaljevanju pa bomo predstavili rezultate, ki se nanašajo na III. del ankete z naslovom »Pretvarjanje arhivskega gradiva v formate, primerne za dolgoročno hrambo, in validacija pretvorb«.

Vprašanja v okviru III. dela ankete so se nanašala na to, ali anketiranci uporabljajo za posamezne vrste gradiva **programska orodja za pretvarjanje formatov** (pretvornik, angl. *converter*) in **validacijo pretvorbe** (validator, angl. *validator*) in, če da, katera.

A TEKSTOVNO (BESEDILNO) GRADIVO

Odgovorilo je 17 anketirancev. 11 državnih arhivov, tj. 64,70 % (Belgija, Češka, Estonija, Hrvaška, Islandija, Litva, Norveška, Slovenija, Švedska, Luksemburg in Portugalska), uporabljajo programska orodja za pretvorbo ali validacijo tekstovnega (besedilnega) gradiva, 6 državnih arhivov, tj. 35,30 % (Danska, Italija, Latvija, Poljska, Španija, Nemčija), pa takšnih orodij ne uporabljajo (graf 1).

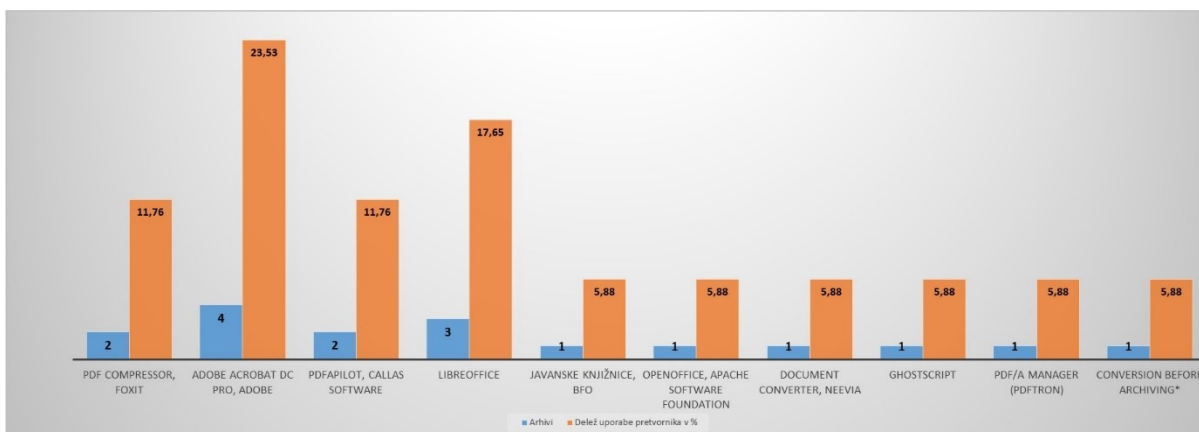


Graf 1: Uporaba pretvornikov in validatorjev za tekstovno (besedilno) gradivo

Tekstovno/besedilno gradivo: PRETVORNIKI

Enajst anketirancev, ki so odgovorili, da uporabljajo programska orodja za pretvarjanje besedilnega gradiva v formate za dolgoročno hrambo, najpogosteje uporablja naslednje pretvornike (graf 2):

- Adobe Acrobat DC Pro (Adobe): 23,53 % (Litva, Norveška, Slovenija, Švedska),
- Libre Office: 17,65 % (Belgija, Estonija, Norveška),
- pdfaPilot (callas software): 11,76 % (Češka, Norveška),
- PDF compressor (Foxit): 11,76 % (Litva, Slovenija).



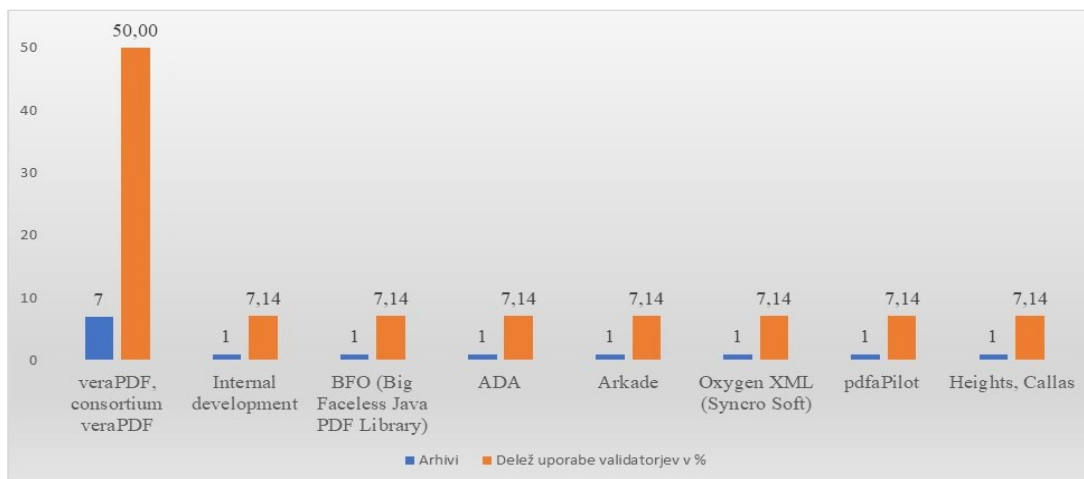
Graf 2: Pretvorniki za tekstovno (besedilno) gradivo

V državnem arhivu Islandije (5,88 %) se pretvorba izvede pred sprejemom gradiva v arhiv.

Ostale programske rešitve uporablja po en državni arhiv (5,88 %): javanske knjižnice od BFO, OpenOffice (Apache Software Foundation), Document Converter (Neevia), Ghostscript in PDF/A Manager (PDFTron).

Tekstovno/besedilno gradivo: VALIDATORJI

Za validacijo besedilnega gradiva je najbolj pogosto (50 %) uporabljen validator v državnih arhivih **veraPDF (consortium veraPDF)**. Uporabljajo ga Estonija, Hrvaška, Slovenija, Švedska, Norveška, Luksemburg, Portugalska (graf 3).



Graf 3: Validatorji za tekstovno (besedilno) gradivo

Ostali državni arhivi uporabljajo za validacijo besedilnega gradiva še naslednje validatorje (delež uporabe posameznega validatorja predstavlja 7,14 %):

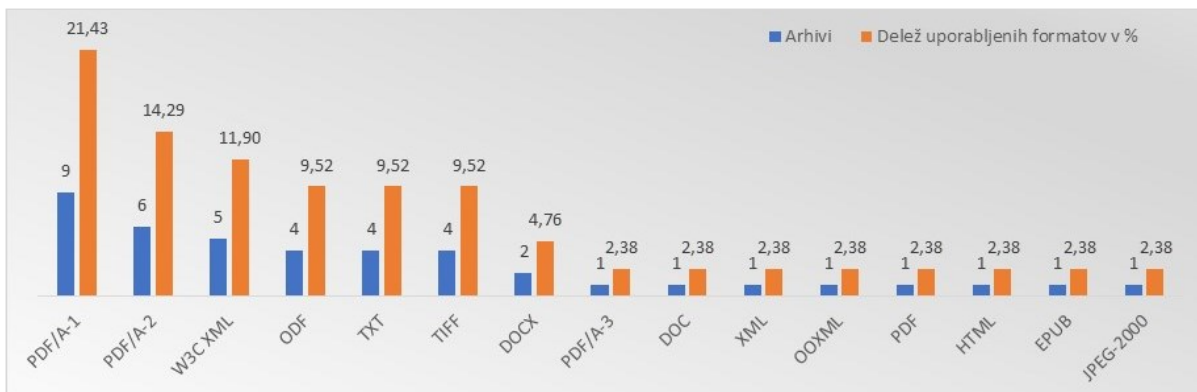
- programsko orodje, ki so ga sami razvili (Belgija),
- javanska knjižnica BFO (Big Faceless Java PDF Library) (Estonija),
- ADA (Islandija),
- Arkade, ki je del lastnega informacijskega sistema (Norveška),
- Oxygen XML (Syncro Soft) (Norveška),
- pdfaPilot (callas software) (Norveška),
- Heights, Callas (<https://digi.nacr.cz/validatorPDF/>) (Češka).

Tekstovno/besedilno gradivo: FORMATI ZA HRAMBO

Vprašanje: **V katerih formatih hranite besedilne dokumente v digitalni obliki?**

Anketiranci uporabljajo za hrambo besedilnih dokumentov različne formate (15) – devet (9) državnih arhivov (Hrvaška, Češka, Litva, Norveška, Slovenija, Švedska, Luksemburg, Nemčija, Portugalska) uporablja format **PDF/A-1**, kar predstavlja 21,43-odstotni delež uporabljenih formatov (graf 4).

Drugi najpogostejši formati za hranjenje besedil so še: 14,29 % **PDF/A-2** (Češka, Estonija, Norveška, Luksemburg, Portugalska, Slovenija), 11,90 % **W3C XML** (Češka, Hrvaška, Norveška, Luksemburg, Slovenija), po 9,52-odstotni delež uporabe pa predstavljajo **TIFF** (Danska, Islandija, Norveška, Luksemburg), **TXT** (Belgija, Estonija, Norveška, Slovenija) in **ODF** (Belgija, Estonija, Norveška, Slovenija).



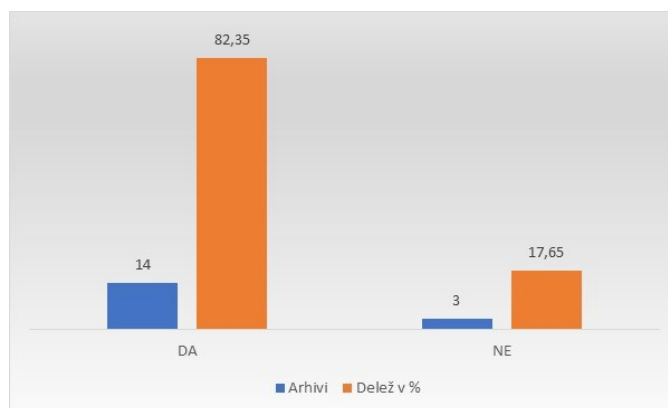
Graf 4: Formati za hrambo tekstovnega (besedilnega) gradiva

Državni arhivi za hranjenje besedilnih dokumentov uporabljajo še format **DOCX** (4,76 % – Estonija, Hrvaška), po 2,38-odstotni delež uporabe pa predstavljajo formati **PDF/A-3** (Slovenija), **EPUB** (Slovenija), **CSV** (Slovenija), **DOC** (Hrvaška), **XML** (Islandija), **OOXML** (Norveška), **PDF** (Hrvaška), **HTML** (Hrvaška) in **JPEG-2000** (Islandija).

B SLIKOVNO GRADIVO

Vprašanje: **Ali uporabljate pretvornike in validatorje za slikovno gradivo?**

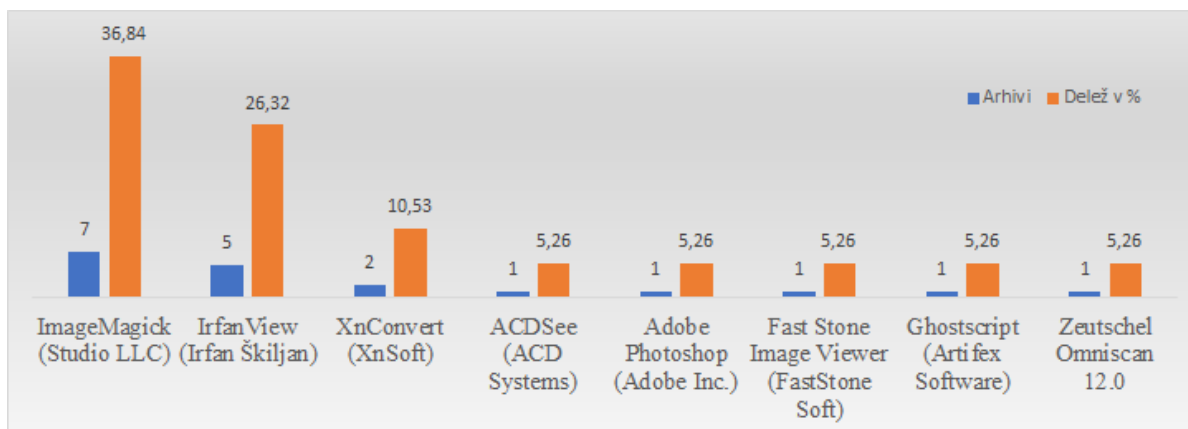
Odgovorilo je 17 anketirancev. Za pretvorbo ali validacijo slikovnega gradiva uporablja programska orodja 14 državnih arhivov (82,35 %). To so Belgija, Češka, Danska, Estonija, Hrvaška, Islandija, Latvija, Litva, Norveška, Slovenija, Švedska, Luksemburg, Nemčija in Portugalska. Takšnih orodij ne uporabljajo v treh (3) državnih arhivih (17,65 %), in sicer v Italiji, Poljski in Španiji (graf 5).



Graf 5: Uporaba pretvornikov in validatorjev za slikovno gradivo

Slikovno besedilo: PRETVORNIKI

Za pretvorbo slikovnega gradiva uporabljajo brezplačno in odprtokodno programsko orodje **ImageMagick (Studio LLC)** v sedmih (7) državnih arhivih (36,84 % – Belgija, Estonija, Slovenija, Švedska, Luksemburg, Nemčija, Portugalska). Programsko opremo **IrfanView (razvijalec Irfan Škiljan)** uporablja pet (5) državnih arhivov (26,32 % – Belgija, Hrvaška, Norveška, Slovenija, Islandija le v namene testiranja). Programsko opremo **XnConvert (XnSoft)** uporabljata dva (2) državna arhiva (10,53 %) – v Latviji in Sloveniji (graf 6).

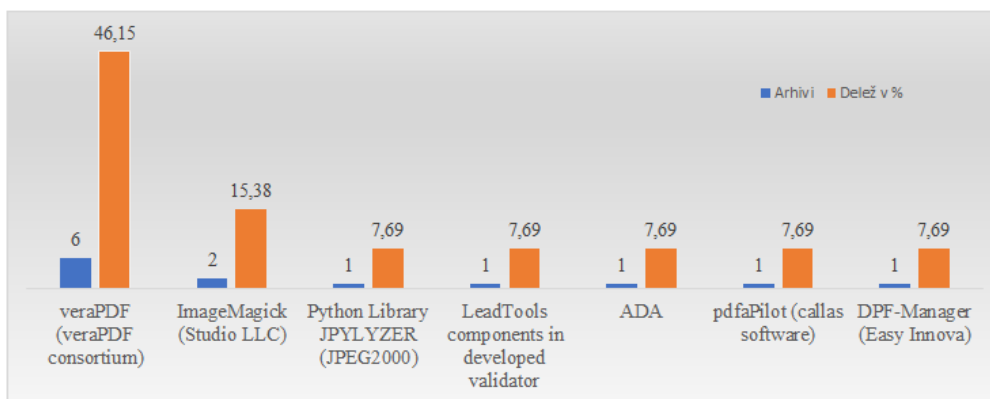


Graf 6: Pretvorniki za slikovno gradivo

Ostale programske rešitve uporablja po en državni arhiv (delež uporabe posameznega pretvornika za slikovno gradivo je 5,26 %): **ACDSee (ACD Systems)** (Litva), **Adobe Photoshop** (Adobe Inc.) (Litva), **Fast Stone Image Viewer** (FastStone Soft) (Litva), **Ghostscript (Artifex Software)** (Luksemburg), **Zeutschel Omniscan 12.0** (Slovenija).

Slikovno besedilo: VALIDATORJI

Za validacijo slikovnega gradiva državni arhivi najpogosteje uporabljajo programsko orodje **veraPDF**, in sicer v šestih (6) državnih arhivih (46,15 % – Češka, Estonija, Norveška, Švedska, Luxemburg, Portugalska). To prikazuje graf 7.



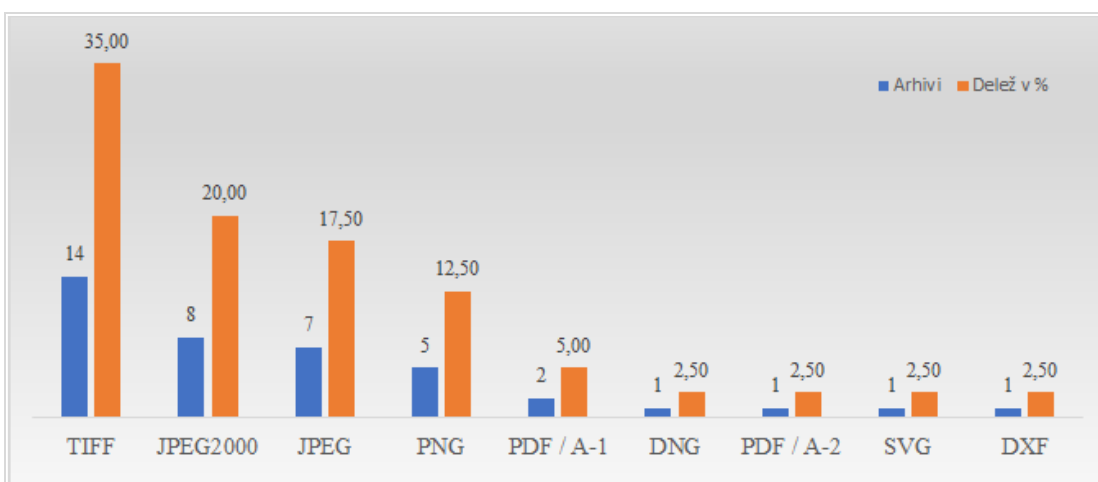
Graf 7: Validatorji za slikovno gradivo

Vsak od ostalih validatorjev ima v državnih arhivih 7,69-odstotni uporabni delež:

- **ImageMagic (ImageMagick Studio LLC)** – Belgija, Slovenija,
- **LeadTools** komponente v okviru lastnega razvoja, le za validacijo datotek TIFF – Danska,
- **ADA** – Islandija,
- **pdfaPilot** (callas software) – Norveška,
- **DPF-Manager (Easy Innova)** – Švedska,
- **Python Library JPYLYZER (JPEG2000)** – Češka.

Slikovno besedilo: FORMATI ZA HRAMBO

Vprašanje: **V katerih formatih hranite slikovno gradivo v digitalni obliki?**



Graf 8: Formati za hrambo slikovnega gradiva

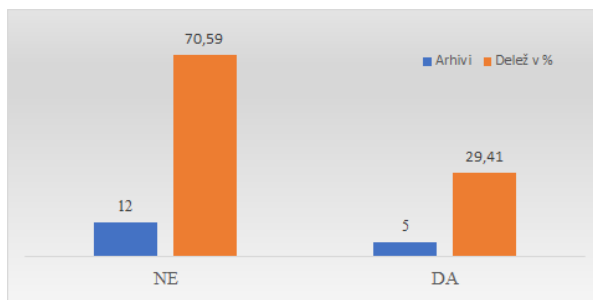
Za hrambo slikovnega gradiva se v največji meri uporablja format **TIFF**, in sicer v 14 državnih arhivih, kar predstavlja 35-odstotni delež (Belgija, Češka, Danska, Estonija, Hrvaška, Islandija, Latvija, Litva, Norveška, Slovenija, Švedska, Luksemburg, Nemčija, Portugalska). Pogosto uporabljena sta še formata **JPEG2000** (20 % – Češka, Danska, Estonija, Islandija, Latvija, Norveška, Nemčija, Slovenija) in **JPEG** (17,50 % – Češka, Hrvaška, Latvija, Litva, Norveška, Nemčija, Slovenija). Deleži so prikazani v grafu 8.

Za hrambo slikovnega gradiva državni arhivi uporabljajo še formate **PNG** (12,50-odstotni delež – Češka, Latvija, Norveška, Estonija, Slovenija), **PDF/A-1** (5-odstotni delež – Češka, Norveška). Po 2,5-odstotni delež uporabe pa predstavljajo formati **DNG** (Litva), **PDF/A-2** (Norveška), **SVG** (Slovenija) in **DXF** (Slovenija).

C PROSTORSKI PODATKI

Za pretvorbo ali validacijo prostorskih podatkov pet (5) državnih arhivov uporablja pretvornike in validatorje za prostorske podatke (29,41 % – Belgija, Danska, Estonija, Islandija, Slovenija), to je prikazano v grafu 9.

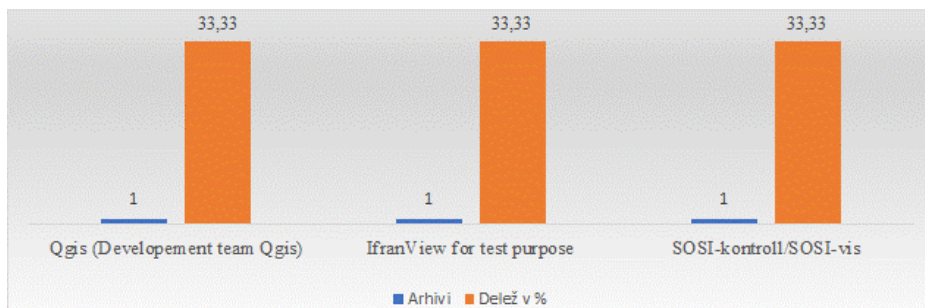
Programskih orodij za prostorske podatke v 12 državnih arhivih ne uporabljajo (70,59 % – Češka, Hrvaška, Italija, Latvija, Litva, Norveška, Poljska, Španija, Švedska, Luksemburg, Nemčija in Portugalska).



Graf 9: Uporaba pretvornikov in validatorjev v državnih arhivih

Prostorski podatki: PRETVORNIKI

Za pretvorbo prostorskih podatkov državni arhiv v Belgiji (33,3 %) uporablja programsko opremo **Qgis** (Development team QGIS). Na Islandiji ustvarjalci sami poskrbijo za pretvorbo v arhivski format pred izročitvijo državnemu arhivu (National Archives of Iceland), v testne namene pa uporabljajo programsko opremo **IrfanView** (razvijalec Irfan Škiljan). Programsko opremo **SOSI-kontroll/SOSI-vis** (Norwegian Mapping Authority) uporablja norveški državni arhiv (33,3%). Vse to prikazuje graf 10.



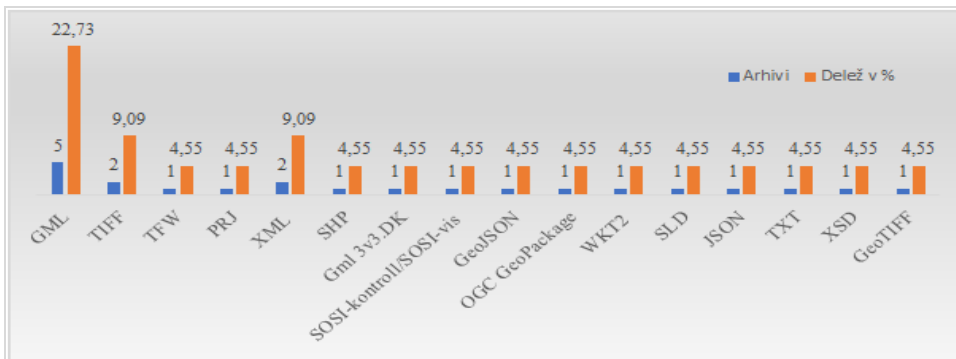
Graf 10: Pretvorniki za prostorske podatke

Prostorski podatki: VALIDATORJI

Za validacijo prostorskih podatkov le državni arhiv v Islandiji uporablja programsko orodje **ADA**, vendar samo v testne namene.

Prostorski podatki: FORMATI ZA HRAMBO

Kakor prikazuje graf 11, uporablja za hrambo prostorskih podatkov pet (5) državnih arhivov v največji meri format **GML** (22,73 % – Danska, Estonija, Islandija, Norveška, Slovenija).



Graf 11: Formati za hrambo prostorskih podatkov

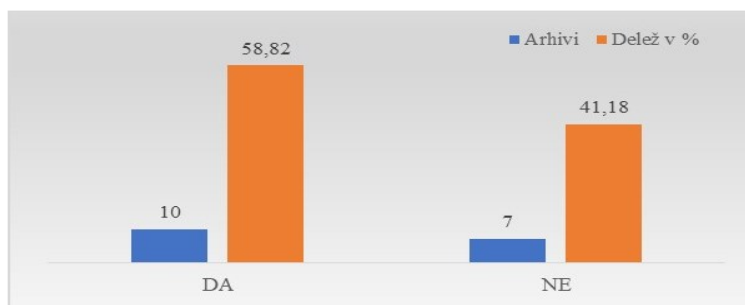
Za hrambo prostorskih podatkov državni arhivi v 9,09-odstotnem deležu uporabljajo še formata **TIFF** (Norveška, Slovenija) in **XML** (Belgija, Slovenija). Vsak od ostalih formatov je v uporabi v 4,55-odstotnem deležu, in sicer: **SOSI-kontroll/SOSI-vis** (Norveška) in **Gml 3v3.DK** (Islandija).

V Sloveniji se dodatno za hrambo prostorskih podatkov uporabljajo še formati **TFW**, **PRJ**, **WKT2**, **SLD**, **JSON**, **TXT**, **XSD**, **GeoTIFF**, v Estoniji pa **SHP**, **GeoJSON**, **OGC GeoPackage**. V obeh državah je uporaba posameznega formata v 4,55-odstotnem deležu glede na ostale formate.

D AVDIO-VIDEO GRADIVO

Vprašanje: **Ali uporabljate pretvornike in validatorje za A/V gradivo?**

Za pretvorbo ali validacijo avdio-video gradiva uporablja programska orodja 10 državnih arhivov (58,82 % – Danska, Estonija, Hrvaška, Islandija, Latvija, Litva, Norveška, Nemčija, Portugalska, Slovenija). Preostalih 7 državnih arhivov (41,18 % – Belgija, Češka, Italija, Poljska, Španija, Švedska, Luksemburg) takšnih orodij ne uporablja (graf 12).



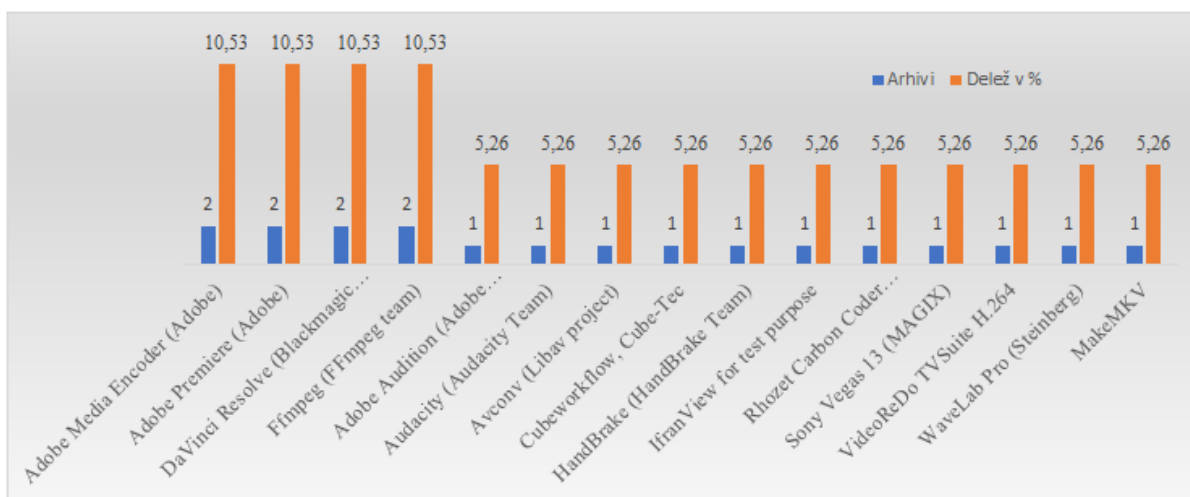
Graf 12: Uporaba pretvornikov in validatorjev za avdio-video gradivo

Avdio-video gradivo: PRETVORNIKI

Za pretvorbo avdio-video gradiva v državnih arhivih najpogosteje uporabljajo pretvornike, vsakega v 10,53-odstotnem deležu (graf 13):

- **DaVinci Resolve** (Blackmagic Design, da Vinci Systems) – Latvija, Litva,
- **FFmpeg** (FFmpeg team) – Norveška, Slovenija,
- **Media encoder** (Adobe) – Estonija, Latvija,
- **Adobe Premiere** (Adobe) – Latvija, Litva.

Sicer pa so uporabljeni zelo raznoliki pretvorniki za avdio-video gradivo, vsak izmed navedenih je v državnih arhivih prisoten v 5,26-odstotnem deležu: **Adobe Audition** (Adobe Systems) – Latvija, **Audacity** (Audacity) – Latvija, **Avconv** (Libav project) – Portugalska, **HandBrake** (HandBrake Team) – Latvija, **IrfanView**²¹ (razvijalec Irfan Škiljan) – Islandija, **Rhomet Carbon Coder** (Harmonic Inc), **Sony Vegas 13** (MAGIX) – Latvija, **VideoReDo TVSuite H.264** – Litva, **WaveLab Pro** (Steinberg) – Litva in **MakeMKV** – Slovenija.

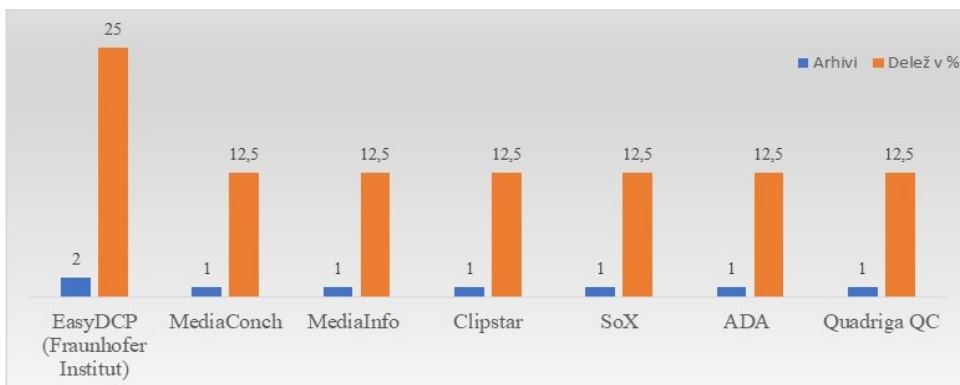


Graf 13: Pretvorniki za avdio-video gradivo

²¹ Opomba: V islandskem državnem arhivu se pretvorba formatov izvede pred sprejemom v arhiv, programsko opremo IrfanView (razvijalec Irfan Škiljan) uporabljajo v testne namene.

Avdio-video gradivo: VALIDATORJI

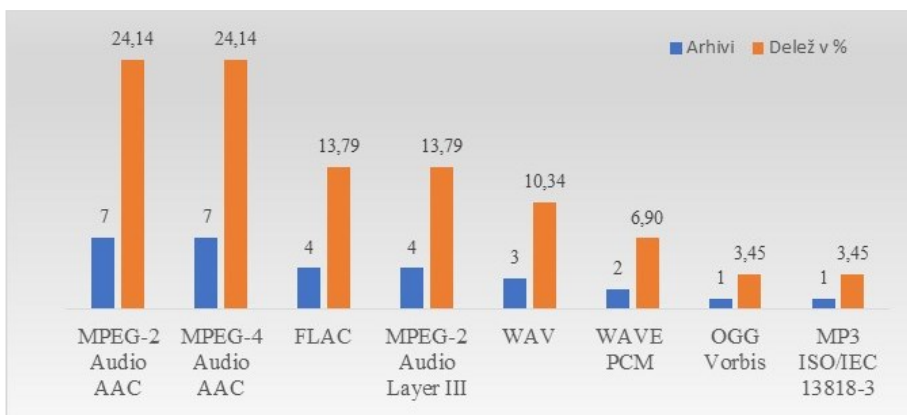
Za validacijo avdio-video gradiva državni arhivi uporabljajo zelo različne validatorje (graf 14), in sicer je v 25-odstotnem deležu v uporabi **EasyDCP** (Fraunhofer Institut), tj. v Nemčiji in Sloveniji, vsak od ostalih validatorjev pa se pojavlja v 12,5-odstotnem deležu: **MediaConch** – Estonija, **MediaInfo** – Danska, **Clipstar**²² – Nemčija, **SoX** – Portugalska in **ADA**²³ – Islandija.



Graf 14: Validatorji za avdio-video gradivo

Avdio-video gradivo: FORMATI ZA HRAMBO

Graf 15 prikazuje pogostost uporabe formatov za hrambo zvočnega, tj. **avdio gradiva**. Državni arhivi v 24,14-odstotnem deležu uporabljajo formata **MPEG-4 Audio AAC** (Češka, Hrvaška, Islandija, Litva, Portugalska, Nemčija, Slovenija) in **MPEG-2 Audio AAC** (Češka, Danska, Hrvaška, Islandija, Litva, Nemčija, Slovenija). S 13,8-odstotnim deležem sledita še formata **FLAC** (Litva, Latvija, Norveška, Slovenija) in **MPEG-2 Audio Layer III** (Češka, Latvija, Nemčija, Slovenija).



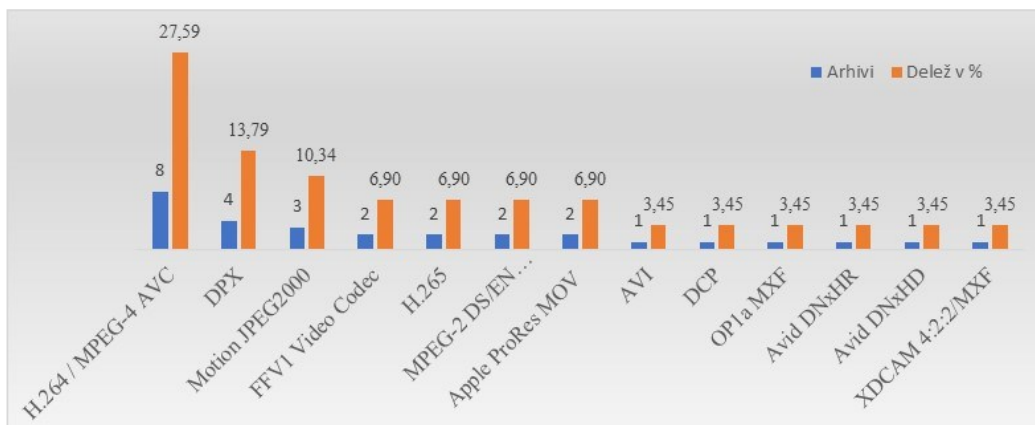
Graf 15: Formati za hrambo avdio gradiva

²² Opomba: za formate kina in dokumentarne filme je validacija del pogodb s ponudniki storitev digitalizacije.

²³ Opomba: samo za testiranje.

Za hrambo **audio** gradiva državni arhivi uporabljajo še formata **WAV** (10,34 % – Litva, Portugalska, Slovenija) in **WAVE PCM** (6,90 % – Latvija, Norveška). Formata **MP3 ISO/IEC 13818-3** (Norveška) in **OGG Vorbis** (Latvija) sta za hrambo audio gradiva prisotna v 3,45-odstotnem deležu.

Za hrambo **video** gradiva (graf 16) državni arhivi najpogosteje uporabljajo format **H.264 / MPEG-4 AVC**, in sicer v 27,59-odstotnem deležu (Danska, Islandija, Latvija, Litva, Norveška, Nemčija, Slovenija). Format **DPX** predstavlja 13,79-odstotni delež uporabe (Hrvaška, Litva, Nemčija in Slovenija), medtem ko je format **Motion JPEG2000** v uporabi v 10,3-odstotnem deležu (Latvija, Nemčija, Slovenija).



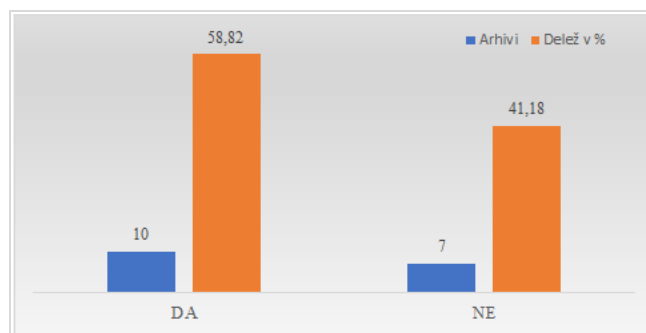
Graf 16: Formati za hrambo video gradiva

Ostali formati predstavljajo 6,9-odstotni delež, in sicer: **MPEG-2 DS/EN ISO/IEC 13818-2** (Islandija, Norveška), **Apple ProRes MOV** (Estonija, Latvija), **H.265** (Litva, Slovenija) in **FFV1 Video Codec** (Litva, Slovenija).

Preostale formate uporablja po en državni arhiv, zato njihov delež uporabe predstavlja 3,45 %: **AVI** (Hrvaška), **DCP** (Nemčija), **OP1a MXF** (Estonija), **Avid DNxHR** (Latvija), **Avid DNxHD** (Latvija) in **XDCAM 4:2:2/MXF** (Latvija).

E PODATKOVNE BAZE

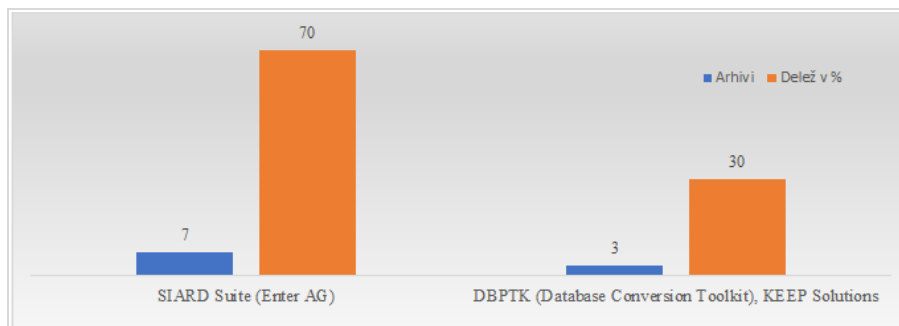
Za pretvorbo ali validacijo podatkovnih baz uporablja programska orodja deset (10) državnih arhivov (58,82-odstotni delež – Belgija, Češka, Danska, Estonija, Latvija, Norveška, Slovenija, Švedska, Nemčija, Portugalska). Preostali državni arhivi (41,18 % – Hrvaška, Islandija, Italija, Litva, Poljska, Španija, Luksemburg) ne uporabljajo programskih orodij za podatkovne baze (graf 17).



Graf 17: Uporaba pretvornikov in validatorjev za podatkovne baze

Podatkovne baze: PRETVORNIKI

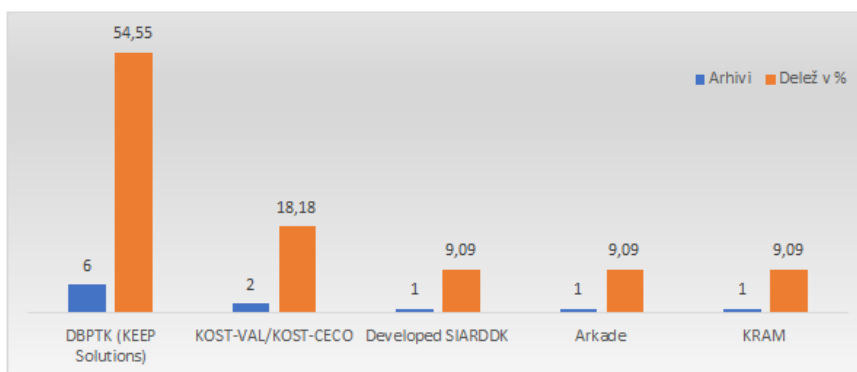
Od tistih, ki so odgovorili, da uporabljajo pretvornike za podatkovne baze, jih sedem (7) uporablja pretvornik **SIARD Suite (Enter AG)**, kar predstavlja 70-odstotni delež (Češka, Estonija, Latvija, Norveška, Nemčija, Portugalska in Slovenija). Pretvornik **DBPTK (Database Conservation Toolkit from KEEP Solutions)** pa je v uporabi pri treh (3) državnih arhivih (30-odstotni delež), in sicer na Češkem, v Latviji in Sloveniji (graf 18).



Graf 18: Pretvorniki za podatkovne baze

Podatkovne baze: VALIDATORJI

Najpogosteje uporabljeno programsko orodje za validacijo formatov podatkovnih baz je **Database Conservation Toolkit (DBPTK)**, ki ga je razvilo podjetje KEEP Solutions (graf 19); uporabljajo ga v šestih (6) državnih arhivih (54,55-odstotni delež – Češka, Estonija, Islandija, Latvija, Portugalska, Slovenija). Validator **KOST-VAL/KOST-CECO** ima 18,18-odstotni delež uporabe (Češka, Nemčija).



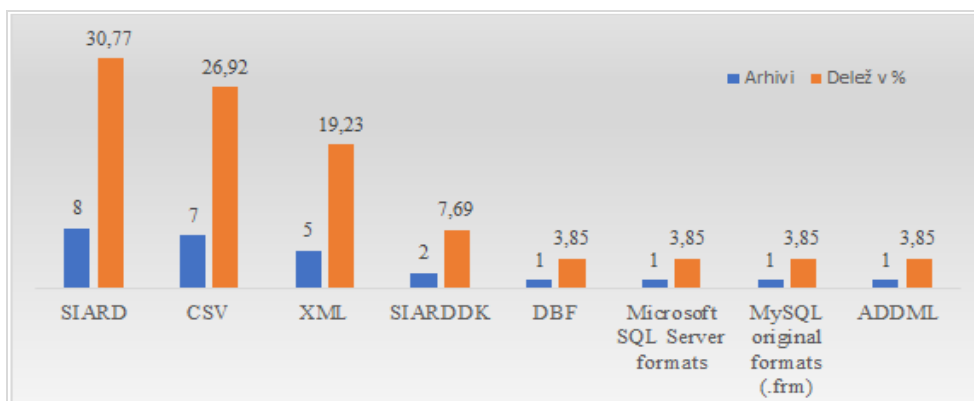
Graf 19: Validatorji za podatkovne baze

Ostali validatorji so v uporabi v 9,09-odstotnem deležu, in sicer gre za **SIARDDK** (Danska), **Arkade** (Norveška) in **KRAM** (Švedska). Vsi trije validatorji so produkt lastnega razvoja.

Podatkovne baze: FORMATI ZA HRAMBO

Za dolgoročno hrambo podatkovnih baz osem (8) državnih arhivov najpogosteje uporablja format **SIARD** (30,77 % – Češka, Danska, Estonija, Latvija, Norveška, Nemčija, Portugalska, Slovenija). Sledi format **CSV**, ki je v uporabi pri sedmih (7) državnih arhivih (26,92 % – Češka, Belgija, Latvija, Norveška, Švedska, Nemčija, Slovenija). Format **XML** uporabljajo v petih (5) državnih arhivih (19,23 % – Belgija, Češka, Latvija, Norveška, Švedska).

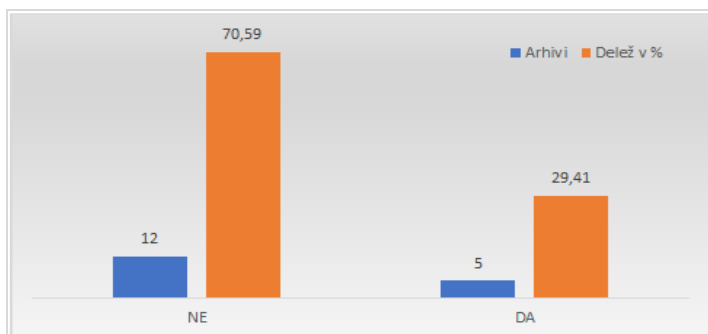
Format **SIARDDK** uporabljata dva (2) državna arhiva (7,69 % – Danska, Islandija). Ostali uporabljeni formati za hrambo podatkovnih baz so prisotni v 3,85-odstotnem deležu, in sicer: **DBF** (Latvija), **Microsoft SQL Server formate** (Latvija), **MySQL formate** (Latvija) in **ADDML**, ki ga uporablja Švedski državni arhiv za hrambo metapodatkov (graf 20).



Graf 20: Formati za hrambo podatkovnih baz

F ELEKTRONSKA POŠTA

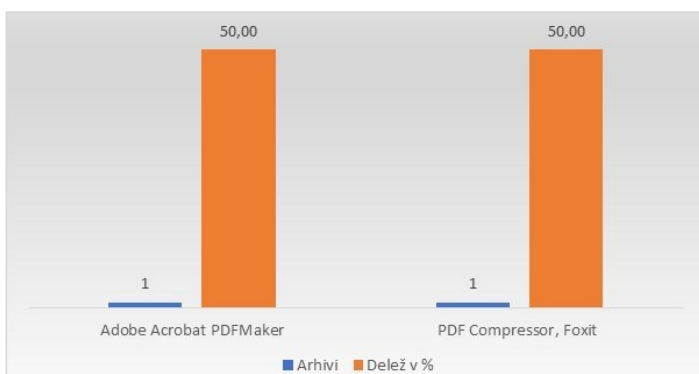
Odgovorilo je 17 anketirancev. Za pretvorbo ali validacijo e-pošte uporabljajo programska orodja v petih (5) državnih arhivih (29,41 % – Belgija, Estonija, Islandija, Litva, Slovenija). Preostalih 12 državnih arhivov (70,59 % – Češka, Danska, Hrvaška, Italija, Latvija, Nemčija, Norveška, Poljska, Španija, Švedska, Portugalska, Luksemburg) ne uporablja pretvornikov in validatorjev za e-pošto (graf 21).



Graf 21: Uporaba pretvornikov in validatorjev za e-pošto

Elektronska pošta: PRETVORNIKI

Da uporabljajo pretvornike in validatorje za e-pošto, je sicer odgovorilo pet (5) državnih arhivov, vendar pa je v nadaljevanju le slovenski državni arhiv navedel, katere, in sicer pretvornika **Acrobat PDFMaker** (Adobe) ter **PDF Compressor** (Foxit), oba pa zaenkrat samo v fazi analize (graf 22).



Graf 22: Pretvorniki za e-pošto

Ostali državni arhivi ne poročajo o programski opremi, ki bi jo uporabljali za namene pretvorbe e-pošte. V državnem arhivu Islandije se pretvorba izvede pred prevzemom v arhiv na strani ustvarjalcev.

Elektronska pošta: VALIDATORJI

Za validacijo e-pošte državni arhiv Islandije uporablja programsko orodje **ADA**, vendar le v testne namene, slovenski državni arhiv pa **PDF Compressor** podjetja Foxit (graf 23).



Graf 23: Validatorji za e-pošto

Elektronska pošta: FORMATI

Najpogosteje hranijo državni arhivi e-pošto v formatu **EML** (27,27-odstotni delež – Islandija, Estonija, Belgija, slednja za hrambo prilog v e-poštnih sporočilih uporablja specifično procesiranje).

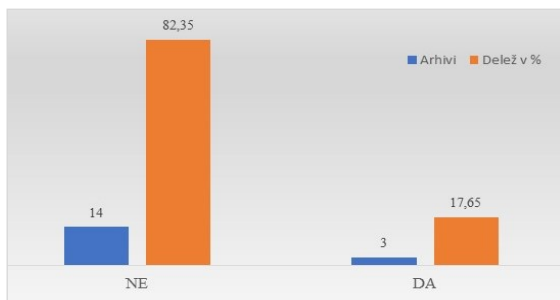


Graf 24: Formati za hrambo e-pošte

Format **TIFF** (18,18 %) uporabljata državna arhiva Danske in Islandije. Poleg formata **EML** uporabljajo v Estoniji za shranjevanje e-pošte še formata **MSG** in **MBOX** (delež 9,09 %). Format **JPEG-2000** (9,09 %) uporabljajo v državnem arhivu Islandije (graf 24).

G SPLETNE STRANI

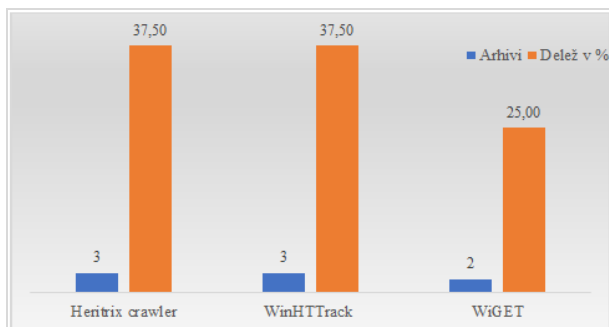
Za pretvorbo ali validacijo spletnih strani se programska orodja uporabljajo v treh (3) državnih arhivih (17,65-odstotni delež – Belgija, Norveška, Slovenija). Preostali državni arhivi (82,35 % – Estonija, Češka, Islandija, Litva, Danska, Hrvaška, Italija, Latvija, Poljska, Španija, Švedska, Portugalska, Luksemburg) programskih orodij za spletne strani ne uporabljajo (graf 25).



Graf 25: Uporaba pretvornikov in validatorjev za spletne strani

Spletne strani: SPLETNI PAJKI

Trije (3) anketiranci so odgovorili, da uporabljajo programsko opremo za žetje spletnih strani, in sicer v 37,5-odstotnem deležu **Heritrix crawler** (Norveška, Belgija, Slovenija) in **WinHTTrack** (Norveška, Švedska, Slovenija), medtem ko uporaba programskega orodja **WiGET** predstavlja 25-odstotni delež (Norveška, Švedska). To je prikazano v grafu 26.



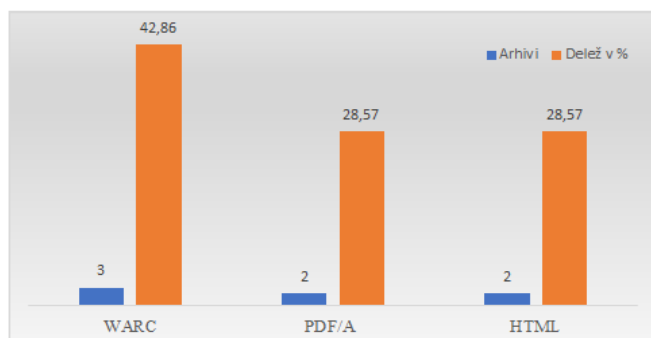
Graf 26: Spletni pajki

Spletne strani: VALIDATORJI

Samo en (1) državni arhiv (Belgija) za validacijo spletnih strani uporablja programsko opremo, in sicer je to **Warctools 4.10.0**, ki so jo razvili Hanzo Archives ter Internet Archive Web ekipa s podporo International Internet Preservation Consortium (IIPC).

Spletne strani: FORMATI ZA HRAMBO

Za hrambo spletnih strani trije (3) državni arhivi (Belgija, Norveška, Slovenija) uporabljajo format **WARC** (delež 42,86 %). Format **PDF/A** (28,57 %) uporabljata državni arhiv na Norveškem in v Sloveniji, format **HTML** pa švedski in slovenski državni arhiv (28,57 %). Deleže prikazuje graf 27.

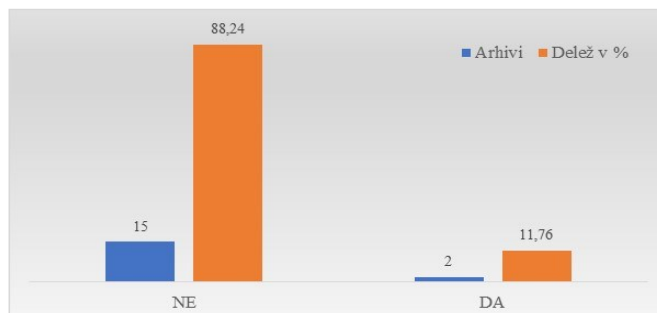


Graf 27: Formati za hrambo spletnih strani

H DRUŽBENA OMREŽJA

Vprašanje: **Ali uporabljate pretvornike in validatorje za zapise na družbenih omrežjih?**

Odgovorilo je 17 anketirancev. Za pretvorbo ali validacijo družbenih omrežij uporabljata programska orodja dva (2) državna arhiva (11,76 % – Norveška, Švedska), ostali (88,24 %) pa teh orodij ne uporabljajo (graf 28).



Graf 28: Uporaba pretvornikov in validatorjev za družbena omrežja

Družbena omrežja: PRETVORNIKI IN VALIDATORJI

V anketnem vprašalniku ni bila ponujena nobena možna izbira, temveč le dana možnost vpisa morebitnih pretvornikov in validatorjev, ki so v uporabi.

Samo norveški državni arhiv je navedel, da je razvil svojo informacijsko rešitev za hrambo objav na družbenih omrežjih (več na <http://beta.some.arkivverket.no/>).

Družbena omrežja: FORMATI ZA HRAMBO

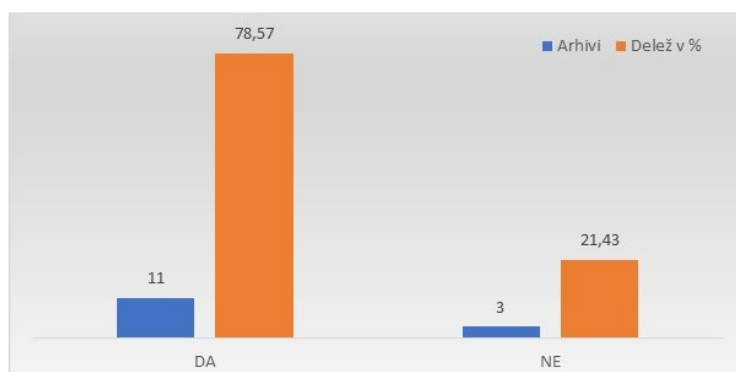
Vprašanje: **V katerih formatih shranjujete zapise z družbenih omrežij?**

Vsebinsko z družbenih omrežij zajemajo v Norveškem državnem arhivu s pomočjo funkcionalnosti *Screengrab*, ki zajame celotno spletno stran in jo shrani v slikovni format. Metapodatki in zajete datoteke se hranijo v podatkovni bazi. Švedski državni arhiv hrani podatke iz družbenih omrežij v formatih **HTML** in **Jacon**.

I OHRANJANJE DOKUMENTOV V ORIGINALNI OBLIKI ZAPISA PO PRETVORBI

Na koncu nas je zanimalo tudi, **ali državni arhivi po pretvorbi gradiva v drug format ohranijo gradivo v originalni obliki.**

Od 17 anketiranih državnih arhivov jih je na vprašanje odgovorilo 14. Med temi enajst (11) državnih arhivov (78,57 % – Belgija, Češka, Estonija, Hrvaška, Litva, Norveška, Luksemburg, Nemčija, Slovenija, Švedska, Portugalska) hrani gradivo v originalni obliki tudi po pretvorbi v arhivski format. Trije (3) državni arhivi (21,43 % – Danska, Islandija, Poljska) gradiva ne hranijo v izvornih formatih, v katerih je bilo gradivo pred pretvorbo. Švedski državni arhiv je dodatno pojasnil, da svojim ustvarjalcem pred izročitvijo določijo formate, ti pa pogosto niso v izvorni obliki, vendar vedno ohranijo tudi gradivo v formatu, ki ga dobijo od ustvarjalcev (graf 29).



Graf 29: Ohranjanje gradiva v originalni obliki zapisa (formatu)

2.3 Izbira pretvornikov in validatorjev

V okviru tretje faze naše raziskave smo na podlagi rezultatov prve in druge faze pripravili razširjen seznam pretvornikov in validatorjev, ki smo ga nato zožili in oblikovali **seznam ožjega izbora pretvornikov in validatorjev**. Prvi kriterij za omenjeno zožitev seznama je bil, da programska orodja podpirajo formate s seznama *najpogosteje uporabljenih formatov* za posamezne vrste digitalnega gradiva, ki so primerni za njegovo dolgoročno hrambo oz. jih Arhiv RS priporoča ali označuje kot sprejemljive. Za pretvorbo posameznih vrst digitalnega gradiva smo pripravili tudi **testne primere**. Med testne primere smo vključili dokumente iz različnih obdobj, trenutno uporabljene formate kot tudi do trideset let nazaj. Člani delovne skupine so prevzeli v preizkus pretvornike in validatorje posameznih vrst digitalnega gradiva. Sledila je **namestitev in preizkus orodij na testnih primerih ter izpolnitev kontrolnih seznamov**.

2.4 Sinteza in interpretacija ugotovitev raziskave

V zadnji, četrti fazi raziskave, je sledila sinteza in interpretacija ugotovitev predhodnih faz ter priprava *seznama priporočenih pretvornikov in validatorjev za posamezne vrste digitalnega gradiva*. Omenjeni seznam je del končnega poročila²⁴ delovne skupine KC2.5, ki je objavljeno na portalu državnega računalniškega oblaka (Portal Nacionalnega interoperabilnostnega okvira NIO²⁵), kot je bilo to načrtovano tudi v projektu e-ARH.si: ESS 2016–2020/2021. V tem prispevku smo zgolj na kratko povzeli priporočeno programsko opremo za pretvorbo in validacijo posameznih vrst digitalnega gradiva (tekstovni in mešani dokumenti, slikovno gradivo, spletne strani, avdio-vizualno gradivo in podatkovne baze).

²⁴ Poročilo o izbiri priporočene programske opreme za pretvarjanje v formate za dolgoročno hrambo 'PRETVORNIKI IN VALIDATORJI' je dostopno na naslovu <https://nio.gov.si/nio/asset/porocilo+o+izbiri+priporocene+programske+opreme+za+pretvarjanje+v+formate+za+dolgorocno+hrambo> (14. 8. 2022).

²⁵ Portal NIO je spletišče, ki je namenjeno objavi interoperabilnostnih rešitev in izdelkov javnega sektorja na enem mestu z namenom objave in spodbujanja dobre prakse in ponovne uporabe.

SEZNAM PRIPOROČENIH PRETVORNIKOV IN VALIDATORJEV ZA POSAMEZNE VRSTE DIGITALNIH ZAPISOV

Naziv pretvornika/ validatorja	Naziv lastnika pretvornika/ validatorja	Vrsta licence
1. TEKSTOVNI IN MEŠANI DOKUMENTI		
PRETVORNIKI		
PDF/A Manager	PDFTron	Lastniška programska oprema (plačljiva)
PDF Compressor	Foxit Software Inc	Lastniška programska oprema (plačljiva)
Nitro Pro	Nitro Software, Inc	Lastniška programska oprema (plačljiva)
Adobe DC PRO	Adobe Inc	Lastniška programska oprema (plačljiva)
FileMerlinTM	Advanced Computer Innovations, Inc	Lastniška programska oprema (plačljiva)
VALIDATORJI		
PDF/A Manager	PDFTron	Lastniška programska oprema (plačljiva)
CSV Validator	The National Archives	Odprta programska oprema
Adobe DC PRO	Acobe Inc	Lastniška programska oprema (plačljiva)
veraPDF	Open Preservation Foundation	Odprta programska oprema
Epubcheck	W3C (konzorcij DAISY)	Odprta programska oprema
KOST-Val	KOST-CECO	Javna programska oprema
2. SLIKOVNO GRADIVO		
PRETVORNIKI		
Converseen	Converseen	Odprta programska oprema
XnConvert	XnSoft	Prosto dostopna programska oprema za <i>neprofitne organizacije, šolske ustanove in privatno uporabo</i> Lastniška programska oprema (plačljiva) za <i>podjetja</i>
VALIDATORJI		
JHOVE	Open Preservation Foundation	Odprta programska oprema
KOST-Val	KOST-CECO	Javna programska oprema
DPF Manager	Easy Innova, the Digital Humanities Lab of the University of Basel, Agents Research Lab of the University	Odprta programska oprema

3. SPLETNE STRANI			
ORODJA ZA ZAJEM OZIROMA ŽETEV: SPLETNI PAJKI			
	Heritrix	Internet Archive	Lastniška programska oprema (plačljiva)
	GNU Wget	Free Software Foundation	Odprta programska oprema
VALIDATORJI			
	JHove	Open Preservation Foundation	Odprta programska oprema
4. AVDIO-VIZUALNO GRADIVO			
PRETVORNIKI			
	FFmpegYag, FFmpeg "Rao"	razvojna skupina FFmpegYag	Odprta programska oprema
	Handbrake	The Handbrake team	Odprta programska oprema
5. PODATKOVNE BAZE			
PRETVORNIKI			
	SIARD Suite	Swiss Federal Archives	Odprta programska oprema
	DBPTK	Keep Solutions, Ida	Javna programska oprema
VALIDATORJI			
	DBPTK	Keep Solutions, Ida	Javna programska oprema

Tabela 3: Seznam priporočenih pretvornikov in validatorjev za posamezne vrste digitalnih zapisov

ELEKTRONSKA POŠTA

Kot za druge vrste digitalnega gradiva velja tudi za e-pošto, da je kvalitetna predpriprava gradiva pred prevzemom v arhiv izjemnega pomena. Področje dolgoročnega ohranjanja e-pošte po vseh načelih varne e-hrambe, predvsem pa prevzemanja v arhive, je v svetu precej neraziskano. Zato o tem obstaja zelo malo primerov v praksi.

Pri e-pošti smo se osredotočili na iskanje programskih orodij, ki omogočajo tudi **izvoz**, nato pa še pretvorbo in validacijo datotek.

Pri iskanju programskih orodij za izvoz in pretvorbo smo se osredotočili na najpogostejša poštna odjemalca e-pošte pri naših ustvarjalcih, IBM Lotus Notes in Microsoft Outlook. Osredotočili smo se na pretvorbe iz formatov e-pošte v arhivski format PDF/A ter na njegovo validacijo.

	Naziv programskega orodja	Naziv lastnika	Vrsta licence
E-POŠTA			
PROGRAMSKA ORODJA ZA IZVOZ IN PRETVORBO²⁶			
	MessageSave	TechHit LLC	Lastniška programska oprema (plačljiva)
	Aid4Mail	Fookes Software Ltd	Lastniška programska oprema (plačljiva)
	PDF Compressor	Foxit Software Inc	Lastniška programska oprema (plačljiva)
VALIDATORJI			
	PDF/A Manager	PDFTron	Lastniška programska oprema (plačljiva)
	Adobe DC PRO	Adobe Inc	Lastniška programska oprema (plačljiva)
	veraPDF	Open Preservation Foundation	Odporna programska oprema
	KOST-Val	KOST-CECO	Javna programska oprema

Tabela 4: Seznam priporočenih pretvornikov in validatorjev za e-pošto

PROSTORSKI PODATKI

Ustvarjalci prostorskih podatkov že pri svojem sprotnem delovanju uporabljajo programska orodja, ki omogočajo hrambo aktualnih in priporočenih arhivskih formatov (glej Seznam oblik zapisa, primernih za dolgoročno hrambo gradiva v digitalni obliki). To velja za prostorske podatke v rastrski in vektorski obliki.

Digitalni prostorski podatki kot kompleksnejša zvrst arhivskega gradiva za dolgoročno hrambo zahtevajo skrbno predpripravo, ki bo v prihodnosti omogočala njihovo uporabo na enak ali nov način. Glede na razvojne usmeritve lahko pričakujemo, da bo v prihodnosti vedno več zahtev po strojnem dostopu do prostorskih podatkov. S tem namenom je potrebno tehnično dokumentacijo v čim večji meri pripraviti v standardni strojno berljivi obliki. Prostorske podatke je treba pretvoriti v formate za dolgoročno hrambo, pri tem pa ohraniti tudi tehnično in vsebinsko dokumentacijo, ki je bistvena za njihovo razumevanje.

Glede na prakso, ki je opisana v osnovnih gradnikih za eArhiviranje kot dela skupnega Evropskega okvirja (CEF), in glede na smernice standarda ISO 19165-1:2018 naj arhivski paket vsebuje najmanj eno obliko gradiva (reprezentacijo), ki je v formatu za dolgoročno hrambo²⁷, možno pa je hraniti tudi druge prezentacije, kamor se lahko hranijo izvirne oblike zapisa ali trenutno najbolj razširjene oblike.

²⁶ Za izvoz in pretvorbo so v omejenem obsegu primerna tudi druga programska orodja: Adobe PDFMaker, ReadPST, ReliefJet ESSENTIALS for Microsoft Outlook, EMailArchiver. Za več informacij se obrnite na Arhiv Republike Slovenije.

²⁷ Seznam oblik zapisa, primernih za dolgoročno hrambo gradiva v digitalni obliki, različica: 1.0. URL: <https://www.gov.si/assets/organi-v-sestavi/Arhiv-RS/Zakonodaja-2019/Pravilnik-o-enotnih-tehnoloskih-zahtevah-za-zajem-in-hrambo-gradiva-v-digitalni-obliki/Seznam-oblik-zapisa-za-dolgorocno-hrambo-ver-1.0.pdf> (15. 8. 2022).

Obe reprezentaciji naj vsebujeta enako vsebino, a v različnih oblikah. Namen tega je, da se lahko eno od njih v prihodnosti opusti ali nadomesti, ko ni več aktualna. Pred izbiro posamezne oblike za konkreten namen hrambe je treba poznati in uskladiti lastne zahteve, navodila pristojnega arhiva in dolgoročna tveganja.

PRIMER: Evidenca, ki se vodi v obliki prostorskih podatkov, je v obliki ESRI Shapefile, ki je zelo razširjen format v trenutnih GIS orodjih, ni pa primeren za dolgoročno hrambo. Za enostavnejši dostop do gradiva je smiselno narediti dve reprezentaciji, in sicer eno v formatu ESRI Shapefile in drugo v obliki, bolj primerni za dolgoročno hrambo, GML. Obe reprezentaciji naj vsebujeta enake vsebine v tehnični dokumentaciji.

GIS portal GeoHub-SI

Državne institucije, ki imajo pravico uporabe storitev na Državnem računalniškem oblaku, lahko uporabljajo GIS portal GeoHub-SI, platformo za objavo prostorskih vsebin in koriščenje servisov GIS ter aplikacij: <https://gisportal.gov.si/portal/home/>.

3. Zaključki

Rezultati raziskave so na naše raziskovalno vprašanje odgovorili pritrdilno – vsa programska orodja, ki jih je mogoče najti na svetovnem spletu, še zdaleč niso ustrezna za pretvarjanje digitalnega gradiva v formate, primerne za dolgoročno hrambo, prav tako vsi validatorji ne izpolnijo naših pričakovanj oz. zahtev (npr. ne izdelajo pisnega poročila o validaciji). Res je, da vsi pretvorniki opredeljujejo podprte vhodne formate, ki jih znajo pretvoriti v druge, izhodne formate. Vendar pa se večina končnih uporabnikov s tem tudi zadovolji, saj ne vedo, da oznaka formata, npr. .pdf, še ne pomeni, da so digitalno gradivo dejansko pretvorili v pravi format (npr. v zeleni arhivski format PDF/A-1, PDF/A-2 ali PDF-A3). Za to namreč potrebujejo dodatna znanja, npr. poznavanje formatov in njihove značilnosti. V nasprotnem primeru se lahko kaj hitro zgodi, da bodo digitalno gradivo pretvorili v formate, za katere so bili prepričani, da so ustrezni, torej primerni za dolgoročno hrambo, v resnici pa niso.

Raziskava je bila izvedena na podlagi več raziskovalnih metod: pregleda literature na tem področju, predhodnih raziskav in dobrih praks ter izvedbe ankete na reprezentativnem vzorcu državnih arhivov. Raziskava torej v večjem delu temelji na najbolj razširjeni uporabi pretvornikov in validatorjev v državnih arhivih EU. Na podlagi rezultatov teh metod smo pripravili seznam kriterijev za izbiro pretvornikov in validatorjev. Ta je v izhodišču temeljil na predhodni raziskavi in znanju ter izmenjavi izkušenj na tem področju v okviru članov delovne skupine KC2.5. Tako oblikovan seznam kriterijev po eni strani lahko služi za ponovljivost raziskave, po drugi strani pa še vedno ostaja možnost, da se ga da prilagoditi svojim potrebam, morda tudi še bolj precizirati in na ta način priti do drugačnih rezultatov. Kot pomemben kriterij smo npr. izpostavili, da morajo programska orodja podpirati formate, ki jih je Arhiv RS opredelil kot priporočene oz. sprejemljive za dolgoročno hrambo digitalnega gradiva. S tem smo želeli s praktičnega vidika podpreti določila slovenskih predpisov (ZVDAGA, UVDAG, PETZ) in tako priskočiti na pomoč predvsem ustvarjalcem arhivskega gradiva. V tem delu gotovo obstaja še prostor za nadaljnje raziskave, saj se zavedamo, da omenjeni seznam formatov še zdaleč ni popolnjen. Ne zajema vseh vrst digitalnih zapisov, prav tako izpostavlja samo nekaj formatov za posamezne vrste digitalnega gradiva. Razširjen seznam formatov pa bi gotovo vplival tudi na ožji izbor priporočenih pretvornikov in validatorjev.

Če povzamemo – ne glede na določene omejitve raziskave njeni rezultati predstavljajo pomagalo vsem tistim, ki so zadalženi za dolgoročno ohranjanje digitalnega gradiva in v okviru tega za izvajanje postopkov pretvarjanja digitalnega gradiva v ustrezne formate, da bo le-to zanesljivo. Hkrati pa rezultati te raziskave predstavljajo tudi dobro izhodišče za nadaljnje raziskave, saj ostaja dejstvo, da poleg izbire ustreznega formata tudi izbira pretvornikov in validatorjev predstavlja pomemben del postopka pretvarjanja digitalnega gradiva v formate, primerne za dolgoročno hrambo. Uporaba ustreznih pretvornikov in validatorjev namreč lahko pomembno vpliva na uporabnost, predvsem pa na celovitost in avtentičnost pretvorjenega digitalnega gradiva.

4. Viri in literatura

HAJTNIK, T. (2016). *Celovit pristop k pretvorbi elektronskih dokumentov v obliko za dolgoročno hrambo : doktorska disertacija*. Maribor: [T. Hajtnik]. XIV, 243 str., ilustr. <https://dk.um.si/Dokument.php?id=104330>. [COBISS.SI-ID 19779094].

NIO. (2022). *Državni računalniški oblak DRO*. URL: <http://nio.gov.si/nio/asset/drzavni+racunalniski+oblak+dro> (14. 8. 2022).

PETZ: Pravilnik o enotnih tehnoloških zahtevah za zajem in hrambo gradiva v digitalni obliki (Uradni list RS, št. 118/20).

Poročilo o izbiri priporočene programske opreme za pretvarjanje v formate za dolgoročno hrambo: PRETVORNIKI IN VALIDATORJI. (2021). URL: <https://nio.gov.si/nio/asset/porocilo+o+izbiri+priporocene+programske+opreme+za+pretvarjanje+v+formate+za+dolgorocno+hrambo> (15. 8. 2022).

Projektna dokumentacija projekta e-ARH.si: ESS 2016 – 2020/2021. Interno gradivo.

PRONOM. (2022). URL: <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx> (15. 8. 2022).

Seznam oblik zapisa, primernih za dolgoročno hrambo gradiva v digitalni obliki, različica: 1.0. (2020). URL: <https://www.gov.si/assets/organi-v-sestavi/Arhiv-RS/Zakonodaja-2019/Pravilnik-o-enotnih-tehnoloskih-zahtevah-za-zajem-in-hrambo-gradiva-v-digitalni-obliki/Seznam-oblik-zapisa-za-dolgorocno-hrambo-ver-1.0.pdf> (15. 8. 2022).

UVDAG: Uredba o varstvu dokumentarnega in arhivskega gradiva (Uradni list RS, št. 42/17).

ZVDAGA: Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih (Uradni list RS, št. 30/06 in 51/14).

SUMMARY

CONVERTERS AND VALIDATORS: SOFTWARE TOOLS FOR CONVERTING TO FORMATS FOR LONG-TERM DIGITAL PRESERVATION AND VALIDATION

Tatjana HAJTNIK, Ph. D

Archives of the Republic of Slovenia, Ljubljana, Slovenia

tatjana.hajtnik@gov.si

Klavdija KRIVEC

Arhiv Republike Slovenije, Ljubljana, Slovenija

klavdija.krivec@gov.si

The conversion of digital records into formats suitable for long-term storage is stipulated by law in Slovenia and thus mandatory, especially for public entities as the largest creators of public archival records. At the same time, this conversion must be reliable. This means that the digital records will remain integral after conversion, and the content of the converted records will be as useful as the content of their originals. The authenticity of the converted records will also be preserved, means, that their content will be identical to the original. The entrance to the implementation of the conversion process is not only the selection of appropriate formats, a key role plays the selection of appropriate software tools, converters.

To determine the reliability of the conversion, we also need suitable software tools, validators. However, in order to choose suitable converters and validators, we need specific knowledge, and above all, the awareness that every software tool found on the web is not suitable for use. At least not when we need to ensure that the conversion will be reliable. The process of selecting suitable converters and validators is therefore a demanding task, which, in addition to specific skills, also requires a certain amount of time. In public archives aware of this and included the task of selecting suitable converters and validators in the project of further development of the Slovenian public electronic archive e-ARH.si in the period 2016 - 2021.

With this, we wanted to facilitate the work of the creators of archival records and at the same time ensure that they used appropriate software tools and thereby ensured the reliable conversion of digital records into formats suitable for long-term storage. Only in this way will we all be able to take care of maintaining the accessibility, usability, integrity and authenticity of digital records, which must be preserved in the long term. This is especially important for that part of digital records that must be preserved as a cultural heritage for our descendants and for the legal security of all of us, for archival records.

The task was carefully planned, for its implementation we used several research methods, from a review of literature and good practices to the use of a structured questionnaire (survey), testing of accessible converters and validators on the market, and synthesis and interpretation of research findings. The paper systematically shows the course of the research in all phases; from the starting points we found in the existing literature in this field and examples of good practices, to conducting a survey on a representative sample of state archives of EU member states and three state archives outside the EU. This part of the research was then followed by the preparation of criteria for the selection of converters and validators. Based on this, we prepared of a wider selection of converters and validators for different types of digital records. We installed them in our IT environment and checked (tested) them based on the criteria prepared in

advance. Based on the synthesis and interpretation of the results of all parts of the research, we created a list of recommended converters and validators for different types of digital records. We published the list on the NIO portal managed by the Ministry of Public Administration. This makes it available to everyone, not only to public entities that are obliged by regulations to convert digital records into formats for long-term preservation, but to everyone who wants to ensure the long-term preservation of digital records.

Priloga 1: Anketa

Institution: Please, click here to enter text

Name and email of the person completing the survey: Please, click here to enter text

AREAS

III. CONVERSION OF ARCHIVAL RECORDS INTO FORMATS SUITABLE FOR LONG-TERM PRESERVATION AND VALIDATION OF SUCH CONVERSIONS

Do you use any software for batch conversion of archival records into formats suitable for long-term preservation and/or validation of such conversions?

- YES
- NO

For which types of digital archival records do you use format converters and validators:

1. TEXTS (e.g. .docx, .txt, .xlsx)

- YES
- NO: skip

Select one or more options or add:

a) CONVERTERS

- PDF compressor, Foxit
- Adobe Acrobat DC Pro, Adobe
- pdfaPilot, callas software
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text. Please, click here to enter text.

b) VALIDATORS (PDF/A-1, PDF/A-2, ODF, W3C XML)

- veraPDF, consortium veraPDF
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text. Please, click here to enter text.

c) Which formats do you use to preserve digital text records?

- PDF/A-1
- PDF/A-2
- ODF
- W3C XML
- Other (please state other formats):
Please, click here to enter text. Please, click here to enter text.

2. IMAGES (e.g. .jpg, .tiff)

- YES
- NO: skip

Select one or more options or add:

a) CONVERTERS

- IrfanView, Irfan Škiljan
- ImageMagic, ImageMagick Studio LLC
- XnConvert, XnSoft
- DJVU Converter
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text.Please, click here to enter text.

b) VALIDATORS (PDF/A-1, TIFF, JPEG, JPEG2000, PNG)

- Python Library JPYZER (JPEG2000)
- veraPDF, veraPDF consortium
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text.Please, click here to enter text.

c) Which formats do you use to preserve digital images?

- PDF / A-1
- TIFF
- JPEG
- JPEG2000
- PNG
- Other (please state other formats):
Please, click here to enter text.Please, click here to enter text.

3. SPATIAL DATA (ESRI ArcGIS formats, other spatial formats)

- YES
- NO: skip

Select one or more options or add:

a) CONVERTERS

- Extension ArcGIS data interoperability, ESRI
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text.Please, click here to enter text.

b) VALIDATORS (.gml, .tiff, .tfw, .pr)

- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text.Please, click here to enter text.

c) Which formats do you use to preserve spatial data?

- GML
- TIFF
- TFW
- PRJ
- Other (please state other formats):
Please, click here to enter text.Please, click here to enter text.

4. AUDIOVISUAL MATERIAL

- YES
- NO: skip

Select one or more options or add:

a) **CONVERTERS**

- Cubeworkflow, Cube-Tec
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text. Please, click here to enter text.

b) **VALIDATORS** (ANSI / SMPTE 268M (DPX), FFV1 Video Codec, Motion JPEG2000, FLAC, DPX, MPEG-2 Audio Layer III, MPEG-2 Audio AAC, MPEG-4 Audio AAC, H.264 / MPEG-4 AVC , H.265)

- Cubeworkflow, Cube-Tec
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text. Please, click here to enter text.

c) **Which formats** do you use to preserve digital audiovisual material?

- ANSI / SMPTE 268M (DPX)
- FFV1 Video Codec
- Motion JPEG2000
- FLAC
- DPX
- MPEG-2 Audio Layer III
- MPEG-2 Audio AAC
- MPEG-4 Audio AAC
- H.264 / MPEG-4 AVC
- H.265
- Other (please state other formats):
Please, click here to enter text. Please, click here to enter text.

5. DATABES (dBase (.dbf), Oracle, Microsoft SQL Server, MySQL, DB / 2, Microsoft Access)

- YES
- NO: skip

Select one or more options or add:

a) **CONVERTERS** (dBase (.dbf), Oracle, Microsoft SQL Server, MySQL, DB / 2, Microsoft Access)

- SIARD Suite, Enter AG
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text. Please, click here to enter text.

b) **VALIDATORS** (SIARD, .csv)

- DBPTK (Database Conservation Toolkit), KEEP Solutions
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text. Please, click here to enter text.

c) **Which formats** do you use to preserve databases?

- SIARD
- CSV
- Other (please state other formats):
Please, click here to enter text. Please, click here to enter text.

6. E-MAIL (Microsoft Outlook, Lotus Notes, Mozilla Thunderbird, Online Mailboxes)

- YES
- NO: skip

Select one or more options or add:

a) **CONVERTERS**

- Adobe Acrobat PDFMaker, Adobe
- PDF Compressor, Foxit
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text. Please, click here to enter text.

b) **VALIDATORS** (PDF / A-1, PDF / A-2)

- veraPDF, veraPDF consortium
- Other (please enter the name of the converter and manufacturer):
Please, click here to enter text. Please, click here to enter text.

c) Which formats do you use to preserve e-mail?

- PDF / A-1
- PDF / A-2
- Other (please state other formats):
Please, click here to enter text. Please, click here to enter text.

7. WEBSITES

- YES
- NO: skip

Select one or more options or add:

a) **WEBSITE CRAWLERS**

- Heritrix crawler
- Other (write the name of the converter and manufacturer):
Please, click here to enter text. Please, click here to enter text.

b) **VALIDATORS** (html 5, .warc)

- Warctools 4.10.0, IIPC, Hanzo Archives
- Other (please enter the name of the converter and manufacturer):

Please, click here to enter text. Please, click here to enter text.

c) **Which formats** do you use for preservation of websites?

HTML 5

WARC

Other (please state other formats):

Please, click here to enter text. Please, click here to enter text.

8. SOCIAL MEDIA (e.g. Facebook, Instagram, Twitter)

YES

NO: skip

Select one or more options or add:

a) **CONVERTERS** (please enter the name of the converter and manufacturer):

Other (please state other formats):

Please, click here to enter text. Please, click here to enter text.

b) **VALIDATORS** (please enter the name of the converter and manufacturer):

Other (please state other formats):

Please, click here to enter text. Please, click here to enter text.

c) **Which formats** do you use to preserve social media?

Other (please state other formats):

Please, click here to enter text. Please, click here to enter text.

Do you keep the documents in their original format after you have converted them to archival format?

YES

NO