

# Nástroje pro digitální archivaci

## Tools for Digital Preservation

*Mgr. Jan Hutař, Ph.D. /Archives New Zealand/10 Mulgrave Street, Thorndon, Wellington 6011, New Zealand ; Mgr. Marek Melichar / Ústav výpočetní techniky, Univerzita Karlova v Praze (Computer Science Centre, Charles University in Prague) , Ovocný trh 3/5, 116 43 Praha 1, Česká republika*

### Abstrakt:

Tento text podává přehledovou základní informaci o nástrojích a informačních zdrojích používaných v oblasti dlouhodobé archivace digitálního obsahu (digital preservation). Zaměřuje se především na nástroje související s problematikou digitálních formátů, jejich identifikací, validací a extrakcí technických metadat.

**Klíčová slova:** dlouhodobá archivace, digitálních formátů, identifikace formátů, validace formátů, extrakce technických metadat

### Summary:

This text offers a basic introduction about the tools and information sources that are applicable in the field of long-term archiving (digital preservation). It focuses in particular upon the tools dealing with digital formats, their identification, validation and the extraction of technical metadata.

**Key words:** digital preservation, long-term archiving, digital formats, identification of formats, validation of formats, extraction of technical metadata

---

Tento text vznikl v rámci řešení výzkumného projektu Fondu rozvoje CESNET č. 516R1/2014 s názvem „Pilotní projekt pro low-barrier přístup k ochraně digitálního obsahu (LTP-pilot)“.

---

## Úvod

Daný text je úvodem do praktického využití nástrojů pro dlouhodobou archivaci digitálního obsahu. Článek ovšem nepřináší vyčerpávající výčet všech dostupných nástrojů, které se používají v oblasti dlouhodobé archivace (DP – *digital preservation*, LTP – *Long Term Digital Preservation*) nebo digitálního kurátorství (*digital curation*). Na příkladech ukazuje, co všechno dlouhodobá archivace digitálního obsahu zahrnuje, a upozorňuje na nejčastěji používané nástroje. Vedle validace formátů, extrakce technických vlastností z digitálních objektů nebo vytváření a aktualizace metadat musí mít správce digitálního obsahu nástroje pro manipulaci s digitálními daty. Trvalé uchování digitálních dat také do značné míry závisí na globální spolupráci a informační infrastruktuře. Znalostní báze informací o digitálních formátech a rizicích s nimi spojených nebo databáze obsahující jedinečné technické znaky konkrétních formátů nemají sice přímo povahu nástroje, ale řada nástrojů na nich závisí. Proto jsou v tomto textu také zmíněny. Naše chápání „nástroje“ zahrnuje tudíž jak programy nebo moduly, tak nástroje, které podporují intelektuální činnosti v oblasti dlouhodobé archivace, jako je plánování akcí dlouhodobé ochrany, analýza rizik dlouhodobého uchování nebo plánování repozitáře obecně.

Text bude přínosem pro zvědavého čtenáře, který bude sám pokračovat v průzkumech v oblasti dlouhodobé archivace a rozhodne se začít některé z nástrojů používat ve své každodenní praxi při správě digitálních dat. Doufáme však, že text poslouží i těm, kdo se nechtějí stát specialisty na dlouhodobou archivaci, ale potřebují např. vyřešit určitý konkrétní problém. Tento článek je zkrácenou verzí detailnějšího textu, který obsahuje i informace o alternativních zdrojích a nástrojích, vzdělávacích projektech, nástrojích z oblasti archivace webu, nástrojích a projektech z oblasti emulace. Ten je dostupný na internetové stránce projektu LTP Pilot /<http://ltp-portal.cz>.

Pro větší přehlednost textu jsme se rozhodli neuvádět internetové odkazy na příslušné nástroje, projekty, organizace apod. přímo v textu, naleznete je souhrnně v tabulce na konci článku.

## 1 Globální informační zdroje a infrastruktura

Dlouhodobá archivace a digitální kurátorství stojí na kolektivním úsilí mnoha komunit, které poskytují své zdroje volně dalším komunitám. Žádný digitální repozitář neexistuje ve vakuu a žádný není zcela soběstačný, každý potřebuje pro svůj provoz informace a nástroje vytvořené jinými komunitami. Jednoduchá řešení i komplexní systémy pro zajištění dlouhodobé archivace a správu dat se obvykle skládají z celé řady technických a organizačních komponent.

Příklady komponent:

- Hardware (HW) a software (SW), úložná média, technické nástroje pro zpracování dat, databáze atd.
- Lidé a jejich kompetence.
- Standardy a pracovní postupy.
- Finanční zdroje, organizace, mandát.
- Ochraňovaná data.
- Dodavatelé a uživatelé dat a jejich systémy, znalosti a potřeby.
- Strategie a plány mateřské instituce.
- Definice pracovních pozic.
- Procesní dokumentace.
- Bezpečnostní předpisy a dokumentace.

Dlouhodobé uchovávání digitálních dat je *nekonečný sled rozhodnutí* a k nim potřebují správci digitálních dat a manažeři repozitářů odborné informace. Rozhodnutí, která mají vliv na schopnost uchovat digitální materiál, mohou být jak strategické, tak manažerské povahy. Mohou se týkat organizačních hledisek nebo mohou mít čistě odborný a technický charakter, jako např. rozhodování o konkrétních krocích spojených s ochranou konkrétního obrazového souboru. Odborné informace mohou správci získat od relevantní komunity, z volně dostupných zdrojů, ale také právě díky různým nástrojům a jejich výstupům. Každá lokální *digital preservation* komunita sdílí svoje informační zdroje, z nichž některé zpřístupňuje globálně, online, komukoli (např. komunita uživatelů konkrétního systému pro archivaci digitálních dat (LTP systém), projekty apod.).

Příklady procesů rozhodování (více například Rosenthal et al.<sup>1</sup>) :

---

<sup>1</sup> ROSENTHAL, Colin, Asger BLEKINGE-RASMUSSEN a Jan HUTAŘ. 2009. *Průvodce plánem důvěryhodného digitálního repozitáře (PLATTER)*. Praha: Národní knihovna ČR, 2009. 51 s. ISBN 978-80-7050-569-4.

- Definice sbírek a dat, které mají být předmětem ochrany.
- Volba hardwarových (HW) a softwarových (SW) komponent (komerční vs. open source, lokální vs. SaaS (Software as a Service), pásky vs. disky apod.).
- Volba způsobu ukládání dat a jejich záloh (grid, cloud, lokální souborové systémy).
- Návrh metod zajištění integrity dat.
- Volba a specifikace metadat, přijímaných datových formátů, struktur balíčků SIP/DIP/AIP (terminologie OAIS ISO 14721:2012/ČSN ISO 14721<sup>2</sup>).
- Stanovení kvalifikačních požadavků a organizačního uspořádání.
- Definice významných vlastností ukládaného materiálu, které je potřeba dlouhodobě ochránit.
- Technická rozhodnutí o tom, co je a co není validní soubor vhodný pro archivaci.
- Strategie/pravidla pro řešení problémů se soubory.
- Volba strategie ochrany (emulace, migrace, normalizace).

## 1.1 Příklady organizací a projektů

- **Open Planets Foundation** (OPF) - spravuje dědictví projektu Planets (nástroje, vzniklou komunitu, informační zdroje) a podporuje další výzkum v oblasti LTP,
- **DPC** (Digital Preservation Coalition), **DCC** (Digital Curation Centre) – britská sdružení podporující vzdělávání a výzkum v oblasti dlouhodobé archivace,
- **Research Data Alliance** (RDA) – organizace zaměřená na propagaci správy a sdílení vědeckých dat,
- **POWRR** (Preserving digital Objects with Restricted Resources) – americký projekt financovaný nadací National Endowment for the Humanities, zaměřující se na otázky dlouhodobé ochrany při omezeném rozpočtu,
- **FADGI** (Federal Agencies Guidelines Initiative) – americká organizace pracující na vytváření a propagaci standardů digitalizace a dlouhodobé ochrany dat,
- **APARSEN** (Alliance for Permanent Access) – evropský projekt s množstvím užitečných výstupů (nástrojů a publikací),
- **Jisc** – britská organizace zaměřená na využití, správu a ochranu akademických a vědeckých digitálních dat,
- národní iniciativy jako:
  - **Nestor** – německé sdružení pro dlouhodobou archivaci, zabývá se standardizací, podporuje výzkum, vzdělávání, audit a certifikaci,
  - **National Agenda for Digital Stewardship** (NDSA) – komunita institucí zabývajících se dlouhodobou archivací v USA v projektu NDIIPP (National Digital Information Infrastructure and Preservation Program) a dalších projektech,
  - **Netherlands Coalition for Digital Preservation** (NCDD) – nizozemská komunita věnující se problematice dlouhodobé ochrany digitálních dat,
  - **Digitalbevaring.dk** – obdobná dánská komunita.

Většina amerických univerzit má studijní programy týkající se problematiky dlouhodobé digitální archivace, např. University of Michigan.

<sup>2</sup> SIP – Submission Information Package, AIP – Archival Information Package a DIP – Dissemination Information Package.

## 1.2 Globální informační zdroje

Formátové registry: **PRONOM**, **GDFR**, **DPTR** a další.

Během posledních 15 let byly klíčovými informačními zdroji pro dlouhodobou archivaci především tzv. formátové knihovny/formátové registry, které poskytují informace o jednotlivých formátech souborů. Formátová knihovna **PRONOM** se stala hlavní součástí globální infrastruktury a závisí na ní řada repositářů. Obsahuje nejen technické informace o formátech, ale také především distribuuje tzv. *signatures files*, tedy soubory s definicemi, které slouží nástrojům jako je **DROID**, **FIDO** nebo **Siegfried** k automatické identifikaci formátů. Podobně i další nástroje pro identifikaci formátů, jako je například **Unix File** nebo **TRID**, mají svoje vlastní mechanismy, jak tvoří a aktualizují svoje *signatures*.

Registry by měly odpovědět mj. na následující otázky (příklady odpovídají registru **PRONOM**):

- Mám digitální objekt, v jakém je formátu?
- Digitální objekt uvádí, že jde o formát X, je to opravdu formát X?
- Mám objekt ve formátu X a chci jej převést na formát Y, jak?
- Mám digitální objekt ve formátu X, jaké má vlastnosti?
- Mám digitální objekt ve formátu X, jaká dokumentace existuje k tomuto formátu?
- Mám digitální objekt ve formátu X, jaké je s ním spojeno riziko?
- Mám digitální objekt ve formátu X, jak a čím jej mohu zobrazit?

**PRONOM** je projekt Národního archivu Velké Británie (TNA). Registr vznikl jako jejich interní znalostní báze spojená s nástrojem **DROID** (Digital Record Object Identification), až později byla zveřejněna. **PRONOM** je stále ovlivněn tím, že byl původně interní bází a stále jej formují spíše vlastní zájmy a potřeby TNA. Ostatní instituce nemají moc možností ovlivnit jeho vývoj ani fungování, jsou na něm přitom závislé. Z tohoto důvodu neustále vznikají pokusy o vytvoření univerzálního decentralizovaného registru formátů. Např. projekty **GDFR** (Global Digital Format Registry) a pozdější **UDFR** (Unified Digital Format Registry) vyprodukovaly sice nové registry, ty se ale pro své nedostatky rozšíření nedočkaly. Na druhé straně zeměkoule byly v roce 2013 pod záštitou NSLA (National and State Libraries Australasia) zahájeny práce na budování registru Digital preservation technical registry, který má ambice být více než jen formátovou knihovnou. Vedle formátů by měl mít údaje o softwaru, hardwaru apod. Díky tomu bude možné vidět konkrétní formát v širším kontextu.

Proč je **PRONOM** důležitý?

- distribuuje soubory se *signatures*<sup>3</sup> nezbytné k identifikaci formátových souborů, které pak používá nástroj **DROID**,
- udržuje a přiděluje identifikátory formátům digitálních dat (PUID, *Persistent Unique Identifier*), které se dnes používají v řadě komerčních a open source LTP systémech,
- identifikátory **PRONOM**, tzv. PUIDy, jsou již *linguou franca* v komunitě analytiků zabývajících se formáty dat a dlouhodobou archivací obecně<sup>4</sup>,
- umožňuje komukoli přidat nový formát (tedy vlastně přidat *signature*<sup>5</sup>).

<sup>3</sup> DROID signature files (archive). THE NATIONAL ARCHIVES. The National Archives [online]. 2012, 2015 [cit. 2015-10-03]. Dostupné z: <http://www.nationalarchives.gov.uk/aboutapps/pronom/droid-signature-files.htm>

<sup>4</sup> Vědí, že pokud se mluví např. o *fmt/353* (<http://apps.nationalarchives.gov.uk/pronom/fmt/353>), jde o konkrétní formát a jeho verzi, která má určité vlastnosti.

<sup>5</sup> Kdokoliv chce přidat do báze **PRONOM** nový formát nebo udělat úpravy stávajícího, musí požádat TNA a poslat vysvětlení – ideálně i s návrhem tzv. *signature* – pro ten konkrétní formát. TNA se poté rozhodne, zda formát zařadí, otestuje *signature* a případně jej upraví. Celý proces trvá několik měsíců.

Vedle globálních a otevřených formátových knihoven existuje několik dalších komunit sdružených kolem konkrétních řešení, které společně budují své vlastní formátové knihovny. Při tom využívají obvykle základní informace z PRONOM, ale mezi členy své komunity pak sdílejí další informace o rizicích nebo pravidlech spojených se zpracováním formátů. Příklady takových komunit jsou uživatelské skupiny systémů jako je **Archivemata**, **Rosetta** nebo **Preservica**.

### 1.3 Znalostní báze o digitálních formátech

Kromě vlastních formátových knihoven, jejichž funkce je do značné míry technická, potřebují správci digitálního obsahu kompletní technické informace o digitálních formátech. Potřebují mít informace pro to, aby mohli rozhodnout, kterým digitálním formátům mohou z dlouhodobého hlediska důvěřovat a proč. Existuje několik přístupů k hodnocení vhodnosti formátu pro dlouhodobou archivaci. Prakticky největší globální dopad má přístup Kongresové knihovny, která rozlišuje sedm faktorů udržitelnosti digitálního formátu (typ formátu – např. proprietární; rozsah používání formátu; transparentnost, možnost vložit metadata přímo do souboru, vnější závislosti, vliv patentů, možnost technické ochrany<sup>6</sup>. Na svém webu pak Kongresová knihovna udržuje rozsáhlou znalostní bázi s informacemi o digitálních formátech s odkazy na dokumentaci k nim a s hodnocením jejich vhodnosti pro dlouhodobé ukládání.<sup>7</sup>

### 1.4 Korpusy digitálních formátů a formátových poškození

Pro testování nástrojů a softwaru repozitářů je k dispozici online několik korpusů testovacích digitálních objektů v různých formátech. Některé z nich také obsahují soubory s typickými chybami nebo jsou záměrně poškozené. Pro techničtější orientované zájemce existuje řada nástrojů, které umožňují provést snadné a kontrolované poškození bitového streamu (sekvence bitů) souborů nebo další podobné manipulace s bitovým streamem. Jako jednoduchý nástroj může posloužit samozřejmě jakýkoli hexadecimální editor nebo textový editor. Část informací v této oblasti pochází z komunity zabývající se forenzním výzkumem. Velmi cenné jsou sbírky reálných poškození (*glitch*, *color shift* apod.), bohužel volně dostupných je jich jen málo.

### 1.5 Globální úsilí v oblasti standardizace a certifikace důvěryhodných digitálních repozitářů

Dlouhodobá archivace vyžaduje především standardizaci (formátů dat, metadat a hlavně procesů). Základním orientačním bodem je standard **ISO 14721**<sup>8</sup> (v češtině jako ČSN ISO 14721<sup>9</sup>), tzv. referenční model **OAIS** (Open Archival Information System),

<sup>6</sup> Sustainability Factors. NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM. *Digital Preservation* [online]. 2005, 20.3.2013 [cit. 2015-10-03]. Dostupné z: <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>.

<sup>7</sup> Format Descriptions. NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM. *Digital Preservation* [online]. 2005, 1.5.2015 [cit. 2015-10-03]. Dostupné z: <http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml>.

<sup>8</sup> ISO 14721:2012. *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*. Geneva: International Organization for Standardization, 2012. 126 s.

<sup>9</sup> ČSN ISO 14721. *Systémy pro přenos dat a informací z kosmického prostoru – Otevřený archivační informační systém – Referenční model*. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014.

popisující informační i funkční model digitálního archivu. Z **OAIS** jsou pak odvozeny další standardy, které například popisují mechanismy interakce mezi tvůrci obsahu a archivem (**PAIMAS**<sup>10</sup> a **PAIS**<sup>11</sup>) nebo stanovují metodiku pro audit a certifikaci repozitářů (**ISO 16363**<sup>12</sup>, v češtině jako **ČSN ISO 16363**<sup>13</sup>, a dále **ISO 16919**<sup>14</sup>). Uvedené normy a nástroje pro audit a certifikaci slouží především k prokázání kvality činnosti repozitáře navenek, např. pro zřizovatele, koncového uživatele apod.

Vedle zmíněných standardů ISO existuje celá řada dalších nástrojů, které mohou pomoci při plánování a provozování důvěryhodných dlouhodobých repozitářů. Jsou to např. **DRAMBORA**, **PLATTER**, **Digital Seal of Approval**, **NESTOR Criteria Catalogue**.

V posledních dvou letech se stává velmi aktuálním audit podle **ISO 16363** (Audit and certification of trustworthy digital repositories). Jde o normu podporující provedení auditu a externí certifikace digitálních repozitářů z pohledu jejich důvěryhodnosti. Norma vznikla na základě metodiky popsané v dokumentu TRAC - Audit a certifikace důvěryhodných repozitářů (**Trustworthy Repositories Audit and Certification**). Splnění kritérií této normy je důležité pro projekty financované z veřejných rozpočtů; certifikovaný archiv/repozitář může být považován za důvěryhodný a data jsou a budou v něm dobře ochráněna. Norma byla vydána v roce 2012, prováděcí norma **ISO 16919** (Requirements for bodies providing audit and certification of candidate trustworthy digital repositories) pak v roce 2014. Existují různé pomůcky pro interní audit podle této normy, zatím ale spíše v podobě tabulek než ucelených nástrojů. Tyto pomůcky mají napomoci provádění auditu.

## 2 Nástroje pro práci s digitálními objekty

Digitální soubory nebo objekty je nutno do detailu poznat, abychom je mohli dlouhodobě uchovávat. K tomu existuje mnoho nástrojů různého zaměření. Postup zpracování digitálního objektu před vstupem do dlouhodobého archivu může být sledem kroků, které mají nejprve rozpoznat, o jaký digitální formát se jedná, a pak pomocí dalších nástrojů ověřit, že digitální objekt je validní reprezentací daného formátu. Nástroje mohou také získat technická metadata z objektu nebo z něj extrahovat důležité technické vlastnosti. Jde většinou o nástroje volně dostupné, fungující jako samostatné aplikace ovládané přes příkazovou řádku, výjimečně mají své vlastní uživatelské rozhraní. Velmi často je pak nalezneme integrované v komplexních systémech pro dlouhodobou ochranu.

Následující část se věnuje jak specializovaným nástrojům vytvořeným nebo používaným při dlouhodobém uchovávaní, tak i nástrojům, které jsou užitečné při správě a manipulaci se soubory apod.

<sup>10</sup> ISO 20652:2006. *Space data and information transfer systems -- Producer-archive interface -- Methodology abstract standard*. Geneva: International Organization for Standardization, 2006. 79 s.

<sup>11</sup> ISO 20104:2015. *Space data and information transfer systems -- Producer-Archive Interface Specification (PAIS)*. Geneva: International Organization for Standardization, 2015. 94 s.

<sup>12</sup> ISO 16363:2012. *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*. Geneva: International Organization for Standardization, 2012. 70 s.

<sup>13</sup> ČSN ISO 16363. *Systémy pro přenos dat a informací z kosmického prostoru - Audit a certifikace důvěryhodných digitálních úložišť*. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014.

<sup>14</sup> ISO 16919:2014. *Space data and information transfer systems -- Requirements for bodies providing audit and certification of candidate trustworthy digital repositories*. Geneva: International Organization for Standardization, 2014. 22 s.

## 2.1 Identifikace formátů dat

Identifikace formátu je obvykle prvním krokem při zpracování digitálního objektu v repozitáři, následuje po kontrolách fixity (neporušenosti), antivirové kontrole a kontrolách úplnosti balíčku dat přicházejících od producenta. Cílem identifikace je jednoznačné určení formátu a jeho verze, a také získání jednoznačného identifikátoru formátu. V internetovém prostředí se často používá jako identifikátor typu formátu tzv. MIME-type, který umí zjistit obvyklé nástroje operačních systémů a souborových systémů; ovšem pro dlouhodobou archivaci je to informace nedostatečná. Vedle MIME-typu pracují systémy repozitářů nejčastěji s unikátním identifikátorem databáze **PRONOM**, tzv. PUID. K jednoznačnému určení formátu používají nástroje v oblasti dlouhodobé archivace již zmíněné *signatures* nebo *magic numbers*. Jsou to v podstatě konkrétní bitové sekvence, které se musí nacházet v těle souboru na místech typických pro daný formát. Při identifikaci také mohou pomáhat externí znaky, např. koncovka souboru.

Nástroje pro identifikaci, resp. *postupy*, které tyto nástroje používají, musí řešit situace, kdy:

- koncovka souboru neexistuje;
- koncovka souboru neodpovídá formátu identifikovanému pomocí *signature*;
- více metod identifikace nevede k stejnému výsledku (soubor s koncovkou pdf je identifikovaný jako doc);
- není k dispozici *signature*, která by uměla identifikovat souborový formát (nové nebo exotické formáty);
- je identifikováno více typů/verzí formátů, mezi nimiž není nástroj schopen dále rozlišit (typicky TIFF, DOC, XML nebo TXT soubory);
- soubor je hybridní;
- obsahuje *signatures* více formátů (PDF a HTML/javascript v jednom, OCR TXT obsahují zdrojový kód naskenovaný k článku apod.).

Při praktickém využití nástrojů pro identifikaci velkého množství souborů musí vývojáři volit vždy mezi rychlostí a spolehlivostí. Rychlejší nástroje obvykle používají jednodušší metody identifikace (hledají sekvenci bitů v kratší části souboru, kontrolují jen hlavičky apod.), větší spolehlivost naproti tomu znamená často větší nároky na čas a výpočetní výkon. Např. DROID se vyvinul v nástroj schopný vytvářet, ukládat a ověřovat profily celých sbírek, není jen knihovnou nebo aplikací ovládanou z příkazové řádky, ale má i své uživatelské prostředí. Nicméně většina nástrojů tohoto typu se běžně nepoužívá jako samostatné aplikace, častěji jsou zapojeny do jiné komplexní aplikace.

Nejobvyklejší nástroje pro identifikaci souborových formátů:

- **DROID** – Digital Record Object Identification (Java, používá **PRONOM signatures**) – je nástroj vyvinutý Národním archivem Velké Británie. Provádí automatickou identifikaci formátů jak u jednotlivých objektů, tak hromadně. Výstupem je informace o konkrétní verzi formátu digitálního objektu.
- **FIDO** – Format Identification for Digital Object (Python, používá **PRONOM signatures**) – velmi rychlý nástroj na identifikaci formátů digitálních objektů vyvinutý díky komunitě Open Planets Foundation. S podpisy v databáze **PRONOM** zachází jinak než např. **DROID**. **FIDO** nemusí být zcela přesné.
- Další podobné nástroje: **Nanite**, **Siegfried**, **ffident**, **Unix File**, **Apache Tika**, **TrID** atd.

## 2.2 Validace formátů a extrakce technických metadat

Validace a extrakce technických metadat je souborem kroků, které následují po potvrzení identity souboru. Validací se zjišťuje, nakolik je daný digitální objekt v souladu s konkrétním předpisem nebo standardem pro daný typ formátu. Zjišťuje se, zda jsou naplněny syntaktické a sémantické požadavky. Extrakce technických metadat je pak dalším krokem, jehož cílem je získat ke každému objektu informace o pokud možno všech jeho vlastnostech (včetně těch, které je klíčové zachovat – *significant properties*).<sup>15</sup> Tyto informace jsou pak uloženy v repozitáři jako metadata. Každý takový nástroj pochopitelně produkuje větší množství metadat a repozitář je musí nějak konvertovat nebo zasadit do datového modelu svého balíčku AIP.

Protože zde jde už o podrobnější zpracování formátů, nelze očekávat, že jeden nástroj bude umět bezvadně pracovat s jakýmkoli formátem. Podobně jako v oblasti identifikace formátů jsou i tyto nástroje vyvíjené jako open source a každý projekt má své limity. Nejobvyklejší nástroje, jako je **JHOVE**, podporují jen několik základních skupin formátů a validaci/extrakci jiných neumožňují. Komunita proto našla dvě řešení – jednou cestou je vývoj formátově specifických nástrojů (jako je Jpylyzer nebo mnoho validátorů PDF/A), druhou cestou je balení více nástrojů do jednoho nástroje (jako je např. **FITS a Flint**).

Důležité je si uvědomit, že validace znamená ověření toho, že celé tělo souboru má strukturu a obsah, která je předepsána, nejčastěji specifikací formátu<sup>16</sup>. V praxi to znamená, že k ověření validity musí nástroje zpracovat třeba i celý obsah souborů.

Samotná extrakce technických metadat nebo významných vlastností může být jednodušší. V této oblasti se běžně používají základní nástroje jako je ExifTool, který je populární i mimo oblast dlouhodobé archivace, nebo Metadata Extraction Tool, MedialInfo a další. Ovšem tyto nástroje formáty většinou nevalidují.

Obvyklé nástroje pro validace formátů a extrakci metadat:

- **JHOVE** – JSTOR/Harvard Object Validation Environment – velmi úspěšný a všude využívaný nástroj, vyvinutý na Harvardské univerzitě ve spolupráci s organizací JSTOR. Cílem bylo automatizovat identifikaci, validaci a extrakci metadat digitálních objektů. Nástroj pracuje s několika datovými formáty (AIFF, ASCII, BYTESTREAM, GIF, HTML, JPEG2000, JPEG, PDF, TIFF, UTF8, XML, ZIP a WAV). Je v mnoha ohledech nastavitelný, např. co se týká podoby výstupu (délka, výstupní formát TXT nebo XML, obsah záznamu), způsobu jeho práce s objekty. Dostupné je i GUI.
- **JHOVE2** – pokus o následovníka JHOVE, umí odlišné formáty, nikdy se ale masově neujal.
- **NZME** – New Zealand Metadata Extraction Tool – jde o jeden z prvních nástrojů pro dlouhodobou ochranu vůbec. Vytvořen byl v Národní knihovně Nového Zélandu

<sup>15</sup> Koncept *significant properties* zahrnuje jak technické vlastnosti digitálního objektu (o které jde zde), tak logické vlastnosti informačního obsahu, které nelze vždy technickými prostředky ze souboru extrahovat. I ty jsou ale předmětem zájmu dlouhodobého repozitáře. Logické *significant properties* konkrétní skupiny objektů musí být popsány v dokumentaci k politikám dlouhodobé archivace a musí být jasné, co je třeba uchovat v budoucnu (příklady – pořadí čtení, stránkování, barva a typ fontu, barva pozadí, kontrast textu na pozadí, uspořádání objektů na stránce, hlavička a patička atd.).

<sup>16</sup> Specifikace může být veřejná, tj. publikovaná, nebo neveřejná – u proprietárních formátů (např. MS Office .doc), nebo někdy nemusí existovat vůbec.



v roce 2003 se záměrem mít nástroj, který dokáže extrahovat ochranná metadata (převážně technická) z formátů používaných balíkem Microsoft Office. Nástroj je dodnes vyvíjen a často v reálném nasazení doplňuje **JHOVE**. Oba nástroje pracují s omezenou množinou formátů, některé z nich se překrývají. **NZME** ale na rozdíl od **JHOVE** zvládá formáty sady MS Office do verze 2007, bohužel ne na XML založené formáty docx, xlsx a pptx).

- Další podobné nástroje – **FITS**, **Flint**, **jpylyzer**, **ExifTool**, **ffprobe**, **MediaInfo**, **BWF MetaEdit** a mnoho dalších.

### 2.3 Formátová migrace (a normalizace)

Migrace digitálního objektu z jednoho formátu do jiného (formátová migrace) je jednou ze základních strategií dlouhodobé archivace.

Formátová migrace není jen jednoduchým převodem objektu z formátu A do formátu B.

- Správce obsahu má obvykle tisíce, milióny souborů v různých formátech. Především musí mít každý z nich identifikovaný formát a přidělené ID formátu a jeho verze.
- Správce musí mít nástroje pro identifikaci rizikové vlastnosti (například kompresního algoritmu, který je proprietární či chybný) nebo musí být schopen identifikovat nevalidní soubory či soubory ve formátech, které nepovažuje za vhodné k archivaci. Správce musí být schopen definovat, jaké vlastnosti informačního obsahu souborů se mají při migraci zachovat. Potřebuje nástroje, aby mohl migrační cesty ověřit, zkontrolovat výsledky a odhadnout časové nároky procesu migrace.
- Výsledek migrace, nový soubor a proces migrace samotný musí být popsány v metadatach objektu. Vazba mezi originálním souborem, který se zachovává, a novou verzí musí být z metadat jasná a vystopovatelná.
- Pokud je repozitář skutečně budován s cílem uchovávat informační obsah dlouhodobě, měly by být v používaném LTP systému k dispozici nástroje a mechanismy, které administrátorovi umožní dozvědět se o nutnosti objekty migrovat do jiného formátu. Musí mít také dostatečně kompetentní zaměstnance, kteří dovedou s pomocí LTP systému a dalších nástrojů migraci provést, vyhodnotit rizika apod.

Z výše uvedeného vyplývá, že technická realizace migrace z formátu A do formátu B je jen jedním z mnoha kroků v řetězu aktivit. Těmto aktivitám se říká „plánování dlouhodobé archivace“, *preservation planning*. K podpoře těchto aktivit vznikla řada metodických i praktických nástrojů (Plato, Planets test-bed a další). Příklad, jak proces plánování formátové migrace může vypadat, podávají McKinney a Gattuso v článku o migraci z formátu WordStar.<sup>17</sup>

Obvyklými nástroji pro formátové migrace (používanými jako samostatné aplikace nebo v podobě pluginů pro komplexní systémy dlouhodobé ochrany) jsou například **Imagemagick**, **Ghostscript**, **MEncoder**, **ffmpeg**, **LibreOffice** a mnoho dalších.

<sup>17</sup> MCKINNEY, Peter a Jay GATTUSO, *Converting WordStar to HTML4*. In: COATES, Serena et al. (eds.). *iPRES 2014 – Proceedings of the 11th International Conference on Digital Preservation, 6-10. 10. 2014, Melbourne*. Melbourne: State Library of Victoria, 2014, s. 149–159. ISBN 978-0-642-27881-4. Dostupné z: <http://www.ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf>.

### 3 Plánování dlouhodobé ochrany

Nejznámější volně dostupné nástroje plánování dlouhodobé ochrany jsou výstupy projektu **PLANETS**, konkrétně jde o **PLANETS testbed** a **Plato**. Jsou to aplikace pro testování migrací, vytváření plánů ochrany, jejichž výstupy lze zakomponovat do postupů LTP systému. Obě aplikace mají svůj prapůvod v evropském projektu DELOS (2002–2006), v jehož rámci existovala i sekce pro dlouhodobou ochranu digitálních dat. V této sekci vznikly první návrhy a prototypy obou nástrojů. Jejich vývoj poté pokračoval v projektu **PLANETS**.

- **Plato** – nástroj na plánování dlouhodobé ochrany, který umožní organizaci vytvářet tzv. plány ochrany, které lze automaticky provádět ve spojení s LTP systémy nebo se systémy uložení dat. Jde o online i offline prostředí, kde lze definovat požadavky na ochranu – způsoby, metody, zachování nebo pominutí signifikantních vlastností různých typů dat.
- kompletní LTP systémy jako **Rosetta**, **Preservica**, **RODA** (používá **Plato**) mají moduly Plánování dlouhodobé ochrany
- **PLANETS Testbed** – šlo o webovou aplikaci, která poskytovala prostředí pro vědecké experimenty v oblasti dlouhodobé ochrany digitálních dat. Přes webový prohlížeč nabízel a kombinoval data, softwarové a hardwarové nástroje k testování různých metod dlouhodobé ochrany pro různé typy digitálních objektů. Výsledky bylo možné podrobně manuálně i automaticky porovnávat. V současné době není nástroj dostupný.

### 4 Nástroje pro práci s metadaty

Dlouhodobé uchování digitálních dat je nedílně spojeno s metadaty. Ta jsou potřebná pro procesy ochrany, jako jsou hodnocení rizik, plánování ochranných akcí apod. V této části se budeme věnovat pouze nástrojům, které pomáhají vytvářet nebo pracovat se standardy (schémata) **METS** a **PREMIS**. Tyto standardy se většinou vyskytují společně, konkrétně **PREMIS** je vnořen do záznamu **METS**; oba standardy vznikly ve snaze naplnit požadavky systémů budovaných podle modelu **OAIS**. **METS** (Metadata Encoding and Transmission System) je metadatový standard navržený pro zápis, výměnu či sdílení různých typů metadat ke konkrétnímu digitálnímu objektu. Umožňuje „zabalit“ do jednoho XML souboru popisná, administrativní, ochranná, strukturální metadata i metadata práv.

**PREMIS** (Preservation Metadata: Implementation Strategies) je schéma specializované na tzv. ochranná metadata, tedy metadata o digitálním objektu, která podporují procesy dlouhodobého uchování toho konkrétního objektu. **PREMIS** je jedním z doporučených standardů pro část administrativních metadat schématu **METS** (amdSec). Existuje konkrétní doporučení, jakým způsobem **PREMIS** do **METS** „zabalit“<sup>18</sup>. K dispozici je celá řada nástrojů pro vytváření a správu metadat v těchto standardech, stejně jako nástrojů, které uvedené standardy využívají. Příklady mohou být **Curator's Workbench**, **PIMTOOLS** – PREMIS in METS Toolbox, **Sobek CM METS Editor**, **Bagit** a další. Mnohé systémy pro správu digitálních dat (DAM – *digital asset management system*) používají **PREMIS** a **METS** nebo jsou schopny je zpracovávat.

<sup>18</sup> GUENTHER, Rebecca et al. *Guidelines for using PREMIS with METS for exchange* [online]. PREMIS Committee 2008 [cit. 2015-10-03]. Dostupné z: <http://www.loc.gov/standards/premis/guidelines-premismets.pdf>.

## Namísto závěru

Kromě výše uvedených specifických nástrojů spoléhají správci digitálního obsahu na běžné systémové nástroje unixových operačních systémů (**Rsync** atd.) nebo nástroje jako **WinSCP**, **putty**. Při analýze digitálních objektů a práci s metadaty jsou užitečné zase **PSPad**, **Notepad++**, **LibreOffice**, **ExifTool**, knihovny pro kontroly a generování kontrolních součtů a další.

Příkladem životaschopných open source projektů vývoje kompletních systémů určených přímo pro dlouhodobou ochranu digitálních dat jsou **Archivematica**, **RODA**, případně komerční řešení jako **Preservica** nebo **Rosetta**. V praxi pak řada institucí využívá řešení vytvořená jako kombinaci volně dostupných a na míru dodělaných komponent. Mezi nimi mohou být některé z výše uvedených nástrojů a repositářové systémy jako **Fedora commons**, **Dspace** a další.

Do globální informační infrastruktury, bez které si dnes nelze dlouhodobou archivaci představit, je třeba započítat také blogy a seznamy dostupných nástrojů. Užitečné jsou především informace publikované na webu OPF (Open Preservation Foundation), nástupce zmiňovaného projektu **Planets** (např. **OPF Knowledge Base**, **OPF Tools registry**), nebo databáze vzniklé v dalších projektech jako jsou **Presto Prime Tool**, **COPTR**, **FileFormat.Info Registry**, **POWRR Tool Grid**, **AVPreserve tools**, **DCH-RP project registry**, **NARA open source tools** aj.

## Použité zdroje a literatura:

ČSN ISO 16363. *Systémy pro přenos dat a informací z kosmického prostoru – Audit a certifikace důvěryhodných digitálních úložišť*. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014.

ČSN ISO 14721. *Systémy pro přenos dat a informací z kosmického prostoru – Otevřený archivační informační systém – Referenční model*. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014.

DROID signature files (archive). THE NATIONALARCHIVES. The National Archives [online]. 2012, 2015 [cit. 2015-10-03].

Dostupné z: <http://www.nationalarchives.gov.uk/aboutapps/pronom/droid-signature-files.htm>.

Format Descriptions. NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM. *Digital Preservation* [online]. 2005, 1. 5. 2015 [cit. 2015-10-03]. Dostupné z: <http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml>.

GUENTHER, Rebecca et al. *Guidelines for using PREMIS with METS for exchange* [online]. PREMIS Committee 2008 [cit. 2015-10-03]. Dostupné z: <http://www.loc.gov/standards/premis/guidelines-premismets.pdf>.

ISO 16919:2014. *Space data and information transfer systems -- Requirements for bodies providing audit and certification of candidate trustworthy digital repositories*. Geneva: International Organization for Standardization, 2014. 22 s.

ISO 16363:2012. *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*. Geneva: International Organization for Standardization, 2012. 70 s.

ISO 20104:2015. *Space data and information transfer systems -- Producer-Archive Interface Specification (PAIS)*. Geneva: International Organization for Standardization, 2015. 94 s.

ISO 14721:2012. *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*. Geneva: International Organization for Standardization, 2012. 126 s.

MCKINNEY, Peter a Jay, GATTUSO, *Converting WordStar to HTML4*. In: COATES, Serena et al. (eds.). *iPRES 2014 – Proceedings of the 11th International Conference on Digital Preservation, 6-10. 10. 2014, Melbourne*. Melbourne: State Library of Victoria, 2014, s. 149–159. ISBN 978-0-642-27881-4. Dostupné z:

<http://www.ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf>.

ROSENTHAL, Colin, Asger BLEKINGE-RASMUSSEN a Jan HUTAŘ. Průvodce plánem důvěryhodného digitálního repozitáře (PLATTER). 1. vyd. Praha: Národní knihovna České republiky, 2009. 51 s. ISBN 978-80-7050-569-4. Dostupné také online z: <http://www.ndk.cz/platter-cz>.

Sustainability Factors. NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM. *Digital Preservation* [online]. 2005, 20. 3. 2013 [cit. 2015-10-03]. Dostupné z: <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>.

Tab.1 Odkazy na nástroje, projekty, instituce a normy uvedené v textu

Název	URL
Apache Tika	<a href="http://tika.apache.org/">http://tika.apache.org/</a>
APARSEN (Alliance for Permanent Access)	<a href="http://www.alliancepermanentaccess.org/index.php/aparsen/">http://www.alliancepermanentaccess.org/index.php/aparsen/</a>
Archivematica	<a href="http://archivematica.org">http://archivematica.org</a>
AVPreserve tools	<a href="http://www.avpreserve.com/avpsresources/tools/">http://www.avpreserve.com/avpsresources/tools/</a>
Bagit	<a href="http://www.digitalpreservation.gov/documents/bagitspec.pdf">http://www.digitalpreservation.gov/documents/bagitspec.pdf</a>
BWF MetaEdit	<a href="http://bwfmetaedit.sourceforge.net/">http://bwfmetaedit.sourceforge.net/</a>
COPTR (Community Owned digital Preservation Tool Registry)	<a href="http://coptr.digipres.org/Main_Page">http://coptr.digipres.org/Main_Page</a>
Curator's Workbench	<a href="http://blogs.lib.unc.edu/cdr/index.php/about-the-curators-workbench/">http://blogs.lib.unc.edu/cdr/index.php/about-the-curators-workbench/</a>
ČSN ISO 14721 - Systémy pro přenos dat a informací z kosmického prostoru - Otevřený archivační informační systém - Referenční model	<a href="http://seznamcsn.unmz.cz/Detailnormy.aspx?k=95767">http://seznamcsn.unmz.cz/Detailnormy.aspx?k=95767</a>
ČSN ISO 16363 - Systémy pro přenos dat a informací z kosmického prostoru - Audit a certifikace důvěryhodných digitálních úložišť	<a href="http://seznamcsn.unmz.cz/Detailnormy.aspx?k=96148">http://seznamcsn.unmz.cz/Detailnormy.aspx?k=96148</a>
DCC (Digital Curation Centre)	<a href="http://www.dcc.ac.uk/">http://www.dcc.ac.uk/</a>
DCH-RP project registry (Digital Cultural Heritage Roadmap for Preservation)	<a href="http://www.digitalmeetsculture.net/heritage-showcases/dch-rp/registry-of-services-and-tools/">http://www.digitalmeetsculture.net/heritage-showcases/dch-rp/registry-of-services-and-tools/</a>
Digital Seal of Approval	<a href="http://datasealofapproval.org/en/">http://datasealofapproval.org/en/</a>
Digitalbevaring.dk	<a href="http://digitalbevaring.dk/">http://digitalbevaring.dk/</a>
DPC (Digital Preservation Coalition)	<a href="http://www.dpconline.org/">http://www.dpconline.org/</a>
DPTR (Digital Preservation Technical Registry)	<a href="http://ndha-wiki.natlib.govt.nz/current-initiatives/technical-registry/">http://ndha-wiki.natlib.govt.nz/current-initiatives/technical-registry/</a>
DRAMBORA (Digital Repository Audit Method Based on Risk Assessment)	<a href="http://www.repositoryaudit.eu/">http://www.repositoryaudit.eu/</a>
DROID (Digital Record Object Identification)	<a href="http://droid.sourceforge.net/">http://droid.sourceforge.net/</a>

ExifTool	<a href="http://www.sno.phy.queensu.ca/~phil/exiftool/">http://www.sno.phy.queensu.ca/~phil/exiftool/</a>
FADGI (Federal Agencies Guidelines Initiative)	<a href="http://www.digitizationguidelines.gov/">http://www.digitizationguidelines.gov/</a>
ffident	<a href="https://github.com/gmcgath/ffident">https://github.com/gmcgath/ffident</a>
ffmpeg	<a href="http://www.ffmpeg.org/about.html">http://www.ffmpeg.org/about.html</a>
FIDO (Format Identification for Digital Objects)	<a href="https://github.com/openpreserve/fido">https://github.com/openpreserve/fido</a>
FileFormat.Info Registry	<a href="http://www.fileformat.info/index.htm">http://www.fileformat.info/index.htm</a>
FITS (The File Information Tool Set)	<a href="http://projects.iq.harvard.edu/fits">http://projects.iq.harvard.edu/fits</a>
Flint	<a href="https://github.com/openpreserve/flint">https://github.com/openpreserve/flint</a>
GDFR (Global Digital Format Registry)	<a href="http://hul.harvard.edu/gdfr/documents.html">http://hul.harvard.edu/gdfr/documents.html</a>
Ghostscript	<a href="http://www.ghostscript.com/">http://www.ghostscript.com/</a>
Imagemagick	<a href="http://www.imagemagick.org/index.php">http://www.imagemagick.org/index.php</a>
ISO 14721 - Space data and information transfer systems - Open archival information system (OAIS) - Reference model	<a href="http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284">http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284</a>
ISO 16363 - Space data and information transfer systems - Audit and certification of trustworthy digital repositories	<a href="http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=56510">http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=56510</a> a <a href="http://public.ccsds.org/publications/archive/652x0m1.pdf">http://public.ccsds.org/publications/archive/652x0m1.pdf</a>
ISO 16919 - Space data and information transfer systems -- Requirements for bodies providing audit and certification of candidate trustworthy digital repositories	<a href="http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57950">http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57950</a>
JHOVE (JSTOR/Harvard Object Validation Environment)	<a href="http://hul.harvard.edu/jhove/">http://hul.harvard.edu/jhove/</a>
JHOVE2	<a href="https://bitbucket.org/jhove2/main/wiki/Home">https://bitbucket.org/jhove2/main/wiki/Home</a>
Jisc	<a href="https://www.jisc.ac.uk/">https://www.jisc.ac.uk/</a>
jpglyzer	<a href="https://github.com/openpreserve/jpylyzer">https://github.com/openpreserve/jpylyzer</a>
LibreOffice	<a href="http://www.libreoffice.org/">http://www.libreoffice.org/</a>
MediaInfo	<a href="https://mediainfo.net/cs/MediaInfo">https://mediainfo.net/cs/MediaInfo</a>
MEncoder	<a href="http://www.mplayerhq.hu/DOCS/HTML/en/encoding-guide.html">http://www.mplayerhq.hu/DOCS/HTML/en/encoding-guide.html</a>
METS (Metadata Encoding and Transmission Standard)	<a href="http://www.loc.gov/standards/mets/">http://www.loc.gov/standards/mets/</a>
Nanite	<a href="https://github.com/openplanets/nanite/">https://github.com/openplanets/nanite/</a>
NARA (National Archives and Records Administration) open source tools	<a href="http://www.archives.gov/records-mgmt/prmd/open-source-tools-for-records-mgmt-report.pdf">http://www.archives.gov/records-mgmt/prmd/open-source-tools-for-records-mgmt-report.pdf</a>
NDSA (National Agenda for Digital Stewardship)	<a href="http://www.digitalpreservation.gov/ndsas/">http://www.digitalpreservation.gov/ndsas/</a>
Nestor	<a href="http://www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html">http://www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html</a>
Nestor Criteria Catalogue	<a href="http://files.d-nb.de/nestor/materialien/nestor_mat_08_eng.pdf">http://files.d-nb.de/nestor/materialien/nestor_mat_08_eng.pdf</a>
NCDD (Netherlands Coalition for Digital Preservation)	<a href="http://www.ncdd.nl/en/">http://www.ncdd.nl/en/</a>
Notepad++	<a href="http://notepad-plus-plus.org">http://notepad-plus-plus.org</a>
NSLA (National and State Libraries Australasia)	<a href="http://www.nsla.org.au/">http://www.nsla.org.au/</a>
NZME (New Zealand Metadata Extractor)	<a href="http://meta-extractor.sourceforge.net/">http://meta-extractor.sourceforge.net/</a>
OPF (Open Planets Foundation)	<a href="http://openpreservation.org/">http://openpreservation.org/</a>
OPF Knowledge Base	<a href="http://wiki.opf-labs.org/display/KB/Home">http://wiki.opf-labs.org/display/KB/Home</a>
OPF Tools registry	<a href="http://wiki.opf-labs.org/display/TR/Digital+Preservation+Tool+Registry">http://wiki.opf-labs.org/display/TR/Digital+Preservation+Tool+Registry</a>

PAIMAS (Producer-Archive Interface Methodology Abstract Standard)	<a href="http://public.ccsds.org/publications/archive/651x0m1.pdf">http://public.ccsds.org/publications/archive/651x0m1.pdf</a>
PAIS (Producer-Archive Interface Specification)	<a href="http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206511R1/Attachments/651x1r1.pdf">http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206511R1/Attachments/651x1r1.pdf</a>
PIMTOOLS (PREMIS in METS Tool)	<a href="http://pim.fcla.edu/">http://pim.fcla.edu/</a>
PLANETS	<a href="http://www.planets-project.eu/">http://www.planets-project.eu/</a>
Plato	<a href="http://www.ifs.tuwien.ac.at/dp/plato/intro/">http://www.ifs.tuwien.ac.at/dp/plato/intro/</a>
PLATTER (Planning Tool for Trusted Electronic Repositories)	<a href="http://www.ndk.cz/platter-cz/Platter.pdf">http://www.ndk.cz/platter-cz/Platter.pdf</a>
POWRR (Preserving digital Objects with Restricted Resources)	<a href="http://digitalpowrr.niu.edu/">http://digitalpowrr.niu.edu/</a>
POWRR Tool Grid	<a href="http://digitalpowrr.niu.edu/tool-grid/">http://digitalpowrr.niu.edu/tool-grid/</a>
PREMIS (Preservation Metadata: Implementation Strategies)	<a href="http://www.loc.gov/standards/premis/">http://www.loc.gov/standards/premis/</a>
Preservica	<a href="http://preservica.com/">http://preservica.com/</a>
Presto Prime Tools	<a href="https://prestocentre.org/library/tools">https://prestocentre.org/library/tools</a>
PRONOM	<a href="http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx">http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx</a>
PSPad	<a href="http://www.pspad.com/">http://www.pspad.com/</a>
putty	<a href="http://www.chiark.greenend.org.uk/~sgtatham/putty">http://www.chiark.greenend.org.uk/~sgtatham/putty</a>
Research Data Alliance	<a href="https://rd-alliance.org/">https://rd-alliance.org/</a>
RODA	<a href="http://roda.di.uminho.pt/?locale=en#home">http://roda.di.uminho.pt/?locale=en#home</a>
Rosetta	<a href="http://www.exlibrisgroup.com/category/RosettaOverview">http://www.exlibrisgroup.com/category/RosettaOverview</a>
Rsync	<a href="https://rsync.samba.org/">https://rsync.samba.org/</a>
Siegfried	<a href="http://www.itforarchivists.com/siegfried">http://www.itforarchivists.com/siegfried</a>
Sobek CM METS Editor	<a href="http://ufdc.ufl.edu/software/mets">http://ufdc.ufl.edu/software/mets</a>
TrID	<a href="http://mark0.net/soft-trid-e.html">http://mark0.net/soft-trid-e.html</a>
Trustworthy Repositories Audit and Certification (TRAC)	<a href="https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf">https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf</a>
UDFR (Unified Digital Format Registry)	<a href="http://udfr.cdlib.org/">http://udfr.cdlib.org/</a>
Unix File	<a href="http://unixhelp.ed.ac.uk/CGI/man-cgi?file=">http://unixhelp.ed.ac.uk/CGI/man-cgi?file=</a>
WinSCP	<a href="http://winscp.net/eng/docs/lang:cs">http://winscp.net/eng/docs/lang:cs</a>