

## Ochrana digitálních dat a otevřené archivy Trvalé zpřístupnění volně dostupných digitálních zdrojů.

Otevřené digitální repozitáře podporují šíření výstupů vědecké a badatelské práce. Badatelé a učitelé na vysokých školách jejich prostřednictvím volně poskytují digitální dokumenty jako doklady své činnosti, a sdílejí tak výsledky výzkumu s jinými badateli. Volnou dostupnost digitálních dat podporují především univerzity, výzkumná centra, v omezené míře i některé knihovny, muzea, archivy a další kulturní instituce. Podstatná část badatelské a akademické produkce i produkce paměťových institucí dnes již vzniká pouze digitálně a bude uložena v otevřených archivech. Otevřené archivy digitálních dat budou mít stále větší význam pro předávání kulturního a vědeckého dědictví dalším generacím. Badatelské a akademické komunity by měly věnovat pozornost dlouhodobé ochraně digitálních dat ve svých digitálních repozitářích tak, aby tyto objekty zůstaly zachovány a mohly být dále využívány a citovány.

### Kontext a cíle

Podle JISC je dlouhodobá digitální ochrana „řadou činností a zásahů, které mají zajistit trvalou a spolehlivou dostupnost k autentickým digitálním objektům tak dlouho, dokud budou mít tyto objekty význam pro komunity uživatelů.“ V kontextu otevřených archivů jsou důležité tři body:

- Otevřené archivy uchovávají objekty, které většinou vznikly přímo v digitální podobě a neexistují k nim žádné analogové originály.
- Dostupnost a spolehlivost objektů uložených v otevřených archivech je důležitá pro jejich a badatelskou a vědeckou komunitu.
- Trvalé zpřístupnění digitálních objektů v otevřených archivech znamená zajistit jejich trvalou použitelnost a možnost objekty vyhledat, získat, citovat po té, co byly uloženy do archivu. Dostupnost musí být zajištěna tak dlouho, jak je to jen třeba, proto je nutné předcházet technickým výpadkům a dopadům technologických změn na digitální objekty. Proces dlouhodobé ochrany digitálních dat by měl být dobře promyšlený, plánovaný a organizačně zajištěný.

### Modely

Referenční model OAIS definuje informační systém otevřeného archivu jako „archiv složený z nějaké organizace a technického systému, který na sebe bere odpovědnost za dlouhodobou ochranu informací ve prospěch nějaké komunity uživatelů.“ Model OAIS poskytuje kompletní funkční a informační popis repozitáře a definuje „závazky, které musí organizace naplňovat, aby fungovala jako archiv „kompatibilní s OAIS“.

Aplikace principů OAIS a zvláště budování archivů volně dostupných dat podle specifikace funkčního a informačního modelu OAIS jsou základními požadavky na důvěryhodný repozitář a předpokladem dlouhodobé ochrany digitálních objektů.

### Vlastnosti

V otevřeném archivu je dlouhodobá ochrana digitálních dat základem pro budování důvěryhodnosti archivu u autorů dokumentů/objektů i u uživatelů. Důvěryhodný repozitář musí splňovat minimálně následující požadavky:

**Autenticita.** Jistota, že digitální objekt byl vytvořen osobou, která tvrdí, že ho vytvořila. Autenticita znamená jistotu, že tvůrce digitálního objektu nemůže popřít, že objekt vytvořil. Ke garanci autenticity se používají techniky jako digitální podpis a digitální vodoznak.



## Další informace a zdroje:

CrossRef je systém trvalé identifikace badatelského obsahu a mezi-vydavatelský citační systém odkazů k plným textům a souvisejícím zdrojům, využívající DOI.  
[www.crossref.org/02publishers/16openurl.html](http://www.crossref.org/02publishers/16openurl.html)

Citebase Search je semi-autonomní citační index pro volně dostupnou výzkumnou literaturu on-line. Sbírá data z OAI-PMH archivů.  
[www.citebase.org](http://www.citebase.org)

OAI-ORE – Iniciativa otevřených archivů pro výměnu digitálních objektů.  
[www.openarchives.org/ore](http://www.openarchives.org/ore)

OAIS referenční model.  
<http://nssdc.gsfc.nasa.gov/nost/isoas/>

Digital Preservation Europe (DPE) – Sbírka on-line zdrojů důležitých pro archivy a digitální repozitáře.  
[www.digitalpreservationeurope.eu/resources](http://www.digitalpreservationeurope.eu/resources)

Digital Preservation Coalition (DPC), založená v roce 2001, iniciativa reagující na rostoucí potřebu dlouhodobé ochrany digitálních dat ve Velké Británii. Spolupracuje i na mezinárodní úrovni s cílem zajistit ochranu globálního digitálního dědictví a koordinovat znalosti z oblasti digitálních ochrany.  
[www.dpconline.org](http://www.dpconline.org)

JISC Program dlouhodobé ochrany a repozitářů se snaží rozvíjet ochranu digitálních dat, podporuje sdílené vyhledávání, shromažďování dat, hledání řešení pro digitální ochranu a vytváření repozitářů.  
[www.jisc.ac.uk/whatwedo/programmes/programme\\_re\\_p\\_pres.aspx](http://www.jisc.ac.uk/whatwedo/programmes/programme_re_p_pres.aspx)

International Research on Permanent Authentic Records in Electronic Systems (InterPARES) – Dva projekty, které se v letech 1999-2001 a 2002-2006 zabývaly vývojem teoretické a metodologické základny pro dlouhodobou ochranu digitálních dat.  
[www.interpares.org](http://www.interpares.org)

PREMIS (Preservation Metadata: Implementation Strategies) – Společná iniciativa OCLC a RLG, vyvinula základní metadatové schéma pro účely dlouhodobé ochrany digitálních dat a široce použitelný datový slovník. Cílem je identifikovat a zhodnotit jednotlivé strategie kódování, uchovávání a managementu metadat v systémech dlouhodobé ochrany digitálních dat.  
[www.oclc.org/research/projects/pmwg/resources.htm](http://www.oclc.org/research/projects/pmwg/resources.htm)

**Integrita.** Schopnost udržovat data úplná a bez chyb, předcházet nehodám nebo nežádoucím změnám dat (poškození dat). Základní technikou sloužící k detekci jakýchkoli změn digitálních objektů po jejich vstupu do archivu je pořizování a uchovávání *bit/byte checksum* jako je MD5.

**Spolehlivost a dostupnost.** Spolehlivost je schopnost hardwaru a softwaru fungovat bez chyb nebo poruch (v teoretické rovině stoprocentně, v praxi v rovině zajištění vysoké procento spolehlivosti fungování). Dostupnost je procento času, ve kterém je daný systém, aplikace nebo komponent v provozu, v poměru k času, kdy by měl fungovat. Běžnými technikami zajištění vysoké procentuální úrovně spolehlivosti a dostupnosti jsou: vytváření záložních kopií, používání antivirů, firewallů, záplat k operačním systémům, upgradování softwaru, hardwarová redundance a používání hardwaru odolného vůči výpadkům.

**Možnost opakovaného použití.** Možnost získat digitální dokument po celou dobu, kdy jej archiv nebo archivující instituce skladuje. Musí být umožněno badatelské nebo vědecké digitální objekty správně vyhledávat a používat po dlouhou dobu (užití trvalých identifikátorů, zajištění trvalosti skladovacích médií a dostupných formátů).

Dlouhodobá ochrana digitálních dat se nezaměřuje jen na digitální objekty samotné, ale i na všechny ostatní komponenty otevřeného archivu: metadata a identifikátory objektů a metadata pro jejich resolvery.

## Politiky a strategie

- Spolehlivý citační systém musí být založen na trvalých a jedinečných identifikátorech, které:
  - zajistí důvěryhodnost a spolehlivost citovaných digitálních objektů,
  - zajistí dlouhodobou strategii ochrany těchto objektů,
  - zajistí po technické a organizační stránce, aby odkazy na citované objekty spolehlivě fungovaly nad rámec možností URL. (Aby se nezobrazovala stránka „404 objekt nenalezen.“)
- Strategie organizace provozující otevřený archiv musí jasně stanovit, které objekty jsou určeny k dlouhodobé digitální ochraně. Měla by specifikovat organizační a finanční podmínky ochrany dat a přesně delegovat odpovědnost za jednotlivé činnosti v oblasti dlouhodobé ochrany digitálních dat. Role a odpovědnost musí být přesně definovány, je třeba podporovat odborný přístup a zvyšovat kvalifikaci zaměstnanců.
- Ochrana digitálních multidisciplinárních dat musí být založena na úzké spolupráci odborníků z různých oblastí.
- Je třeba zvyšovat povědomí autorů o strategii digitální ochrany tak, aby si uvědomovali význam této činnosti pro zajištění trvalé dostupnosti jejich digitálních objektů.
- Také metadata zajišťují trvalou dostupnost digitálních objektů: Ochrana digitálních dat proto musí obsahovat i strategie a procesy k ochraně metadat.
- Neustálý audit a hodnocení činnosti archivu v oblasti dlouhodobé ochrany musí reagovat na změny v organizaci, strategiích, právních a technických komponentech archivu.