



Digitální svět: technologie, potenciál, rizika

## **Dlouhodobé uchování dat**

**Miroslav Bartošek**, [bartosek@ics.muni.cz](mailto:bartosek@ics.muni.cz)

**David Antoš**, [antos@ics.muni.cz](mailto:antos@ics.muni.cz)

# Přednášející



## Miroslav Bartošek

- ÚVT MU, Knihovnicko-informační centrum MU
- Automatizace knihoven
- Digitální knihovny
- Open Science



## David Antoš

- ÚVT MU, CESNET, z.s.p.o.
- Oddělení datových úložišť
- Digital Preservation

# Obsah přednášky

## 1. Uchování informací v historii lidstva

Paměťové instituce • Nosiče informací a jejich životnost • Co se nám z minulosti dochovalo/nedochovalo a proč

## 2. Uchování informací v digitálním světě

Specifika a hrozby digitálních informací • Křehkost médií • Technologické zastarávání • Ochranné strategie • Aspekty digitální ochrany

## 3. Digitální archiv a osobní archivace

OAIS model – nahlédnutí za oponu • Důvěryhodnost • Tipy pro osobní archivaci • Chmurné perspektivy?

## 4. The Long Now

Podpora dlouhodobého uvažování • 10.000 Year Clock • Rosetta Disc

# 1. Uchování informací v historii lidstva

# 1.1 Uchování informací

– Uchovávání informací = důležitý úkol v historii lidstva

– Klasické „paměťové“ instituce

- **muzea** (fyzické artefakty)
- **archivy** (nepublikovaný materiál)
- **knihovny** (publikovaný materiál)

Alexandrijská knihovna  
(295 př.n.l. – 642 ??)

– Základní přístupy k uchování materiálu

- **konzervace**

uchovávání původního artefaktu  
(metoda: obnovování – refreshing)

- **uchování**

uchování informačního obsahu původního artefaktu, i při zániku originálu  
(metoda: migrace)

# 1.1 Uchování informací



**Národní archiv**  
Praha, Chodov

<https://www.nacr.cz/o-nas/historie>

**Moravský zemský archiv**  
Brno, Bohunice



<https://www.asb-portal.cz/wp-content/uploads/images/fotogaleria/>

# 1.1 Uchování informací



**Národní knihovna ČR**  
Praha, Klementinum  
(Barokní sál)

**Moravská zemská knihovna**  
Brno, Kounicova ul.



Foto : varadikamen.cz

# 1.2 Nosiče informací a jejich životnost

## Rozdílné zkušenosti

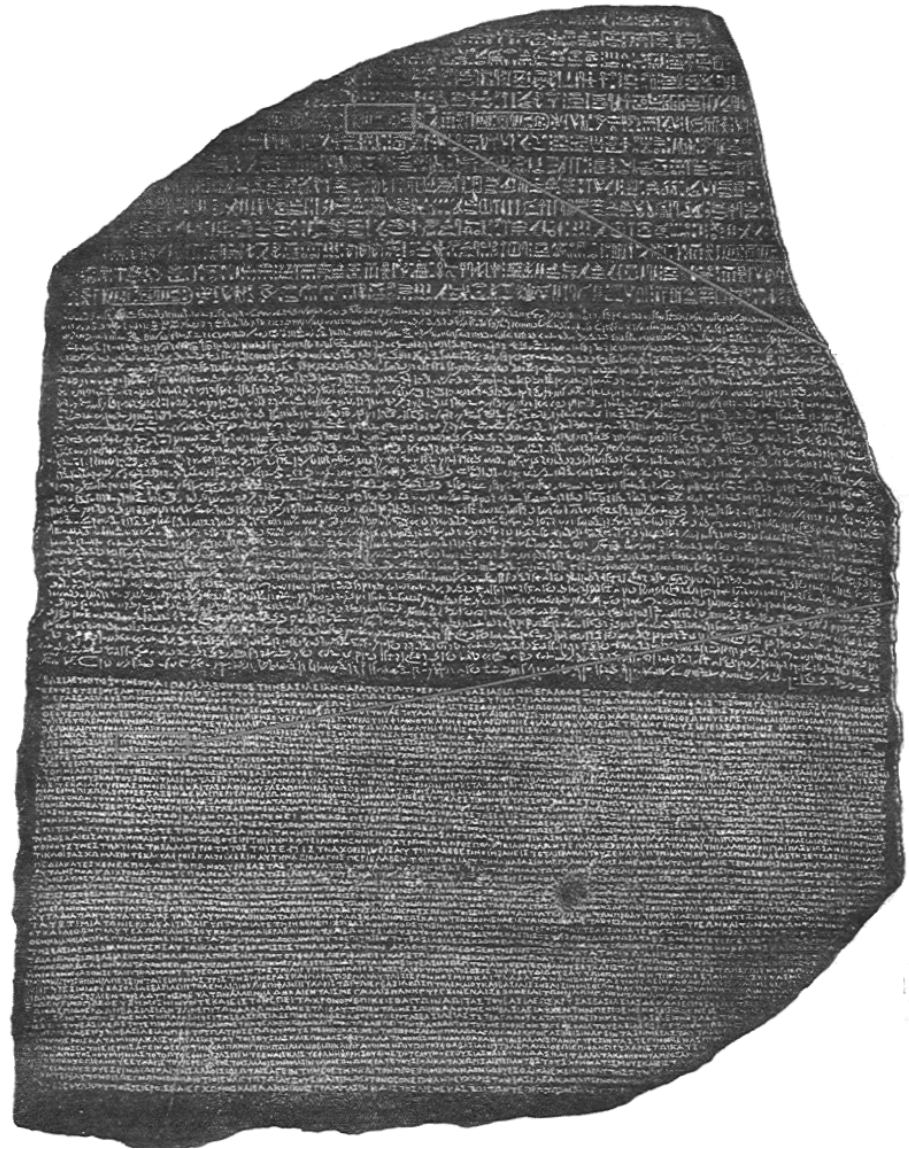
- **starověké záznamy** ~ 4000 let  
(kosti, kámen, hliněné tabulky, papyrus, pergamen, papír)
- **fotografické dokumenty** ~ 200 let (od 1839)  
(fotografické desky, film, fotopapír)
- **audiovizuální záznamy** ~ 100 let  
(voskové a celuloidové válečky, šelakové desky, LP-desky, magnetické pásky)
- **elektronické dokumenty** ~ desítky let  
(magnetický záznam, optický záznam, SSD)

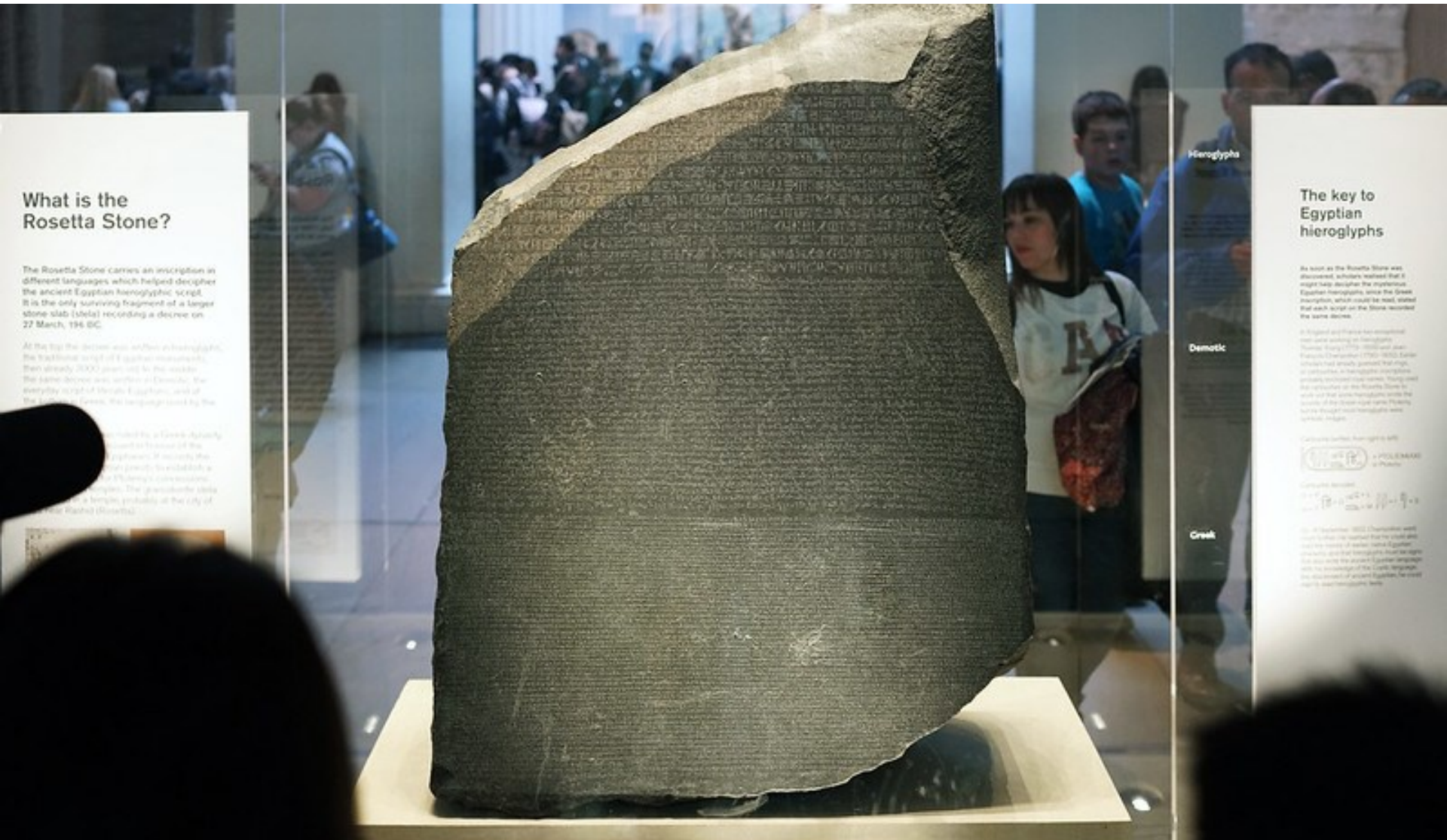
**! Novější nosiče: větší kapacita, ale obvykle kratší životnost!**



# 1.3 Rosettská deska

- Objevena 1799 během Napoleonova tažení do Egypta
- Žulová stéla 114 x 71 cm
- **Text z r. 196 př. n. l.** ve třech různých zápisech:
  - egyptské hieroglyfy
  - egyptské démotické písmo
  - starořečtina
- Champollion: rozluštění hieroglyfů
- Londýn, British Museum





## What is the Rosetta Stone?

The Rosetta Stone carries an inscription in different languages which helped decipher the ancient Egyptian hieroglyphic script. It is the only surviving fragment of a larger stone slab (stelae) recording a decree on 27 March, 196 BC.

At the top the decree was written in hieroglyphs, the traditional script of Egyptian monuments, then already 3000 years old by the middle. The same decree was written in Demotic, the everyday script of Egypt's Egyptians, and of the Greek language, the language used by the

...ruled by a Greek dynasty...  
...of the...  
...to establish a...  
...the...  
...of the city of... (Rosetta).

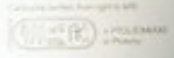
### Hieroglyphs

## The key to Egyptian hieroglyphs

As soon as the Rosetta Stone was discovered, scholars realised that it might help decipher the mysterious Egyptian hieroglyphs, since the Greek hieroglyphs, which could be read, stated that each script on the Stone recorded the same decree.

### Demotic

In England and France two independent men were working on hieroglyphs. Thomas Young (1773-1829) and Jean-François Champollion (1792-1835). Earlier scholars had already guessed that hieroglyphs were a form of writing, but they did not know what the symbols on the Rosetta Stone meant. It was not until 1822 that Champollion realised that some hieroglyphs were the same as the Greek letters. He was the first to decipher the hieroglyphs.

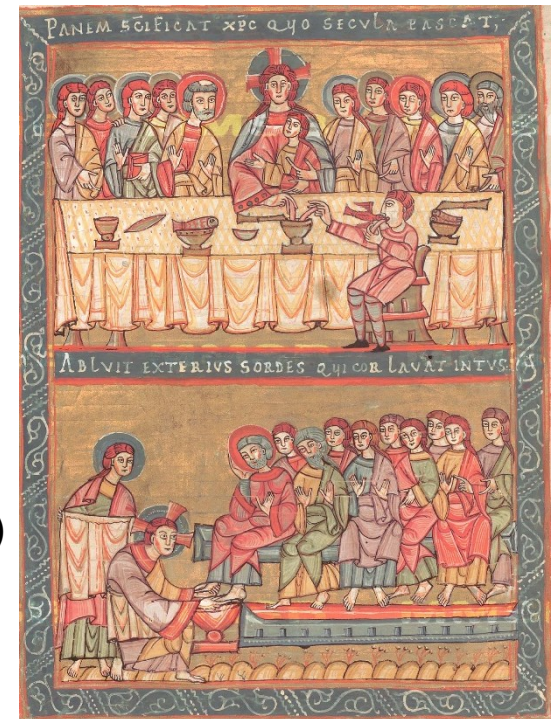


### Greek

On 27 September 1822 Champollion wrote to his brother Luc, saying that he had just deciphered the hieroglyphs. He had done this by using the Greek language as a key. He was the first to decipher the hieroglyphs.

# 1.4 Hodně se dochovalo...

- Jeskynní kresby (Lascaux, Francie, 16 000 let)
- Babylónské hliněné destičky (4 000 let)
- Svitky od Mrtvého moře (cca 2 000 let)
- Starověké rukopisy a staré tisky (inkunábule)
- Antické písemnosti, ...
  - Ne vždy se dochoval originál (médium vs informace)  
(přepisy řeckých děl v kláštorech – běh proti času)
  - „Archimedes codex“, Reviel Netz & William Noel, 2007



- V ČR:
- **Vyšehradský kodex** (cca 1085, korunovace Vratislava II.)  
(nejstarší psaná památka, součást „korunovačnických klenotů“)

## 1.5 Hodně se (nenávratně) ztratilo...

– Originální rukopisy řeckých učenců, různý starobylý materiál...

Ale i novodobé dokumenty:

- Značná část novin na kyselém papíru (konec 19. a poč. 20.st.)
- 50 % filmů ze 40. let
- Marvin Minsky (AI, 60. léta) versus Galileo Galilei (16 st.)
- Originální videozáznam z přistání Apollo 11 na Měsíci (19.7.1969)
- 20 % NASA Viking (první průzkum Marsu, 1976)
- První email (1971), obsah první webové stránky (1990)
- ...a mnohé, mnohé další...

**!** Křehkost nosiče, chyby/opomenutí, vandalství, přírodní katastrofy, závislost na technologiích, chybějící systém/infrastruktura

## 2. Uchování informací v digitálním světě

## 2.1 Optimismus – a vystřízlivění

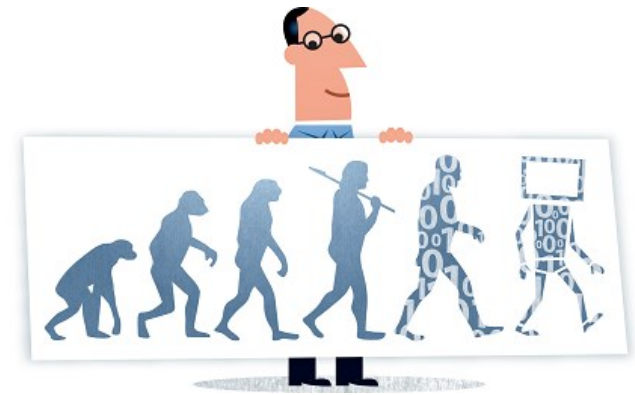
- **Digital information is forever. It doesn't deteriorate and requires little in the way of material media.**

Andy Grove, Intel Corp.

- **Digital information lasts forever – or five years, whichever comes first.**

Jeff Rothenberg, RAND, 1995

## 2.2 Dlouhodobé uchování



DP Handbook

- **Knihovny**: dlouhodobě = staletí
- **Digitální technologie**: inovační cyklus cca 5 let  
(v průměru po 5 letech je daná technologie zastaralá ☹)
- *The trouble with technological progress is that it seems to come at the expense of preservation.* (Ian H.Witten, How to Build a DL)
- *Computer technology is made for information processing, not for long term storage.* (M.Wettengel, NSR)

**!** Digitální technologie dlouhodobé uchování neusnadnily, naopak spíše je komplikují!

## 2.3 V čem je problém s digitální informací?

- Přístup a zobrazení digitální informace jsou závislé na technologiích (nestačí k tomu lidské smysly)
- **Hrozby pro digitální informaci**
  - 1. Křehkost záznamového média**  
(krátká životnost, nízká odolnost vůči změnám)
  - 2. Technologické zastarávání**  
(platformní závislost – nosič, formát, software, hardware)
- Další
  - Velký (trvale rostoucí) objem digitálních dat
  - Finanční nákladnost
  - Nezbytné expertní znalosti



## 2.4 Křehkost digitálního média



### – Krátká životnost

- Pevné disky – roky
- Zapisovatelná optická CD/DVD – roky
- Archivní magnetické pásky – desítky let
  - *Čím vyšší hustota zápisu, tím hůře*

### – Nízká odolnost proti změnám obsahu

- Snadnost provádění (ne)úmyslných změn
- Samovolné změny (bit rot)

Narušení integrity  
a autenticity

### – Rychlá obměna technologií

- Typy médií
- Čtecí zařízení
- Komunikační rozhraní, ovladače

**!** Běžná digitální média nejsou vhodná pro dlouhodobé uchování



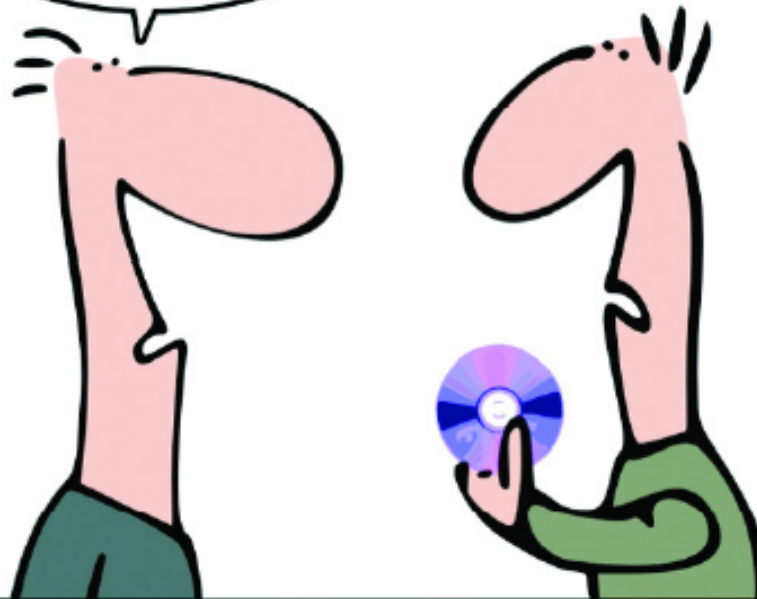
Color photo by Jeff Rothenberg

# HOW TO SAVE YOUR DIGITAL WORK FOR THE POSTERITY?

December 12th 2307

DON'T KNOW.  
LOOKS LIKE A  
FUNNY BEER  
COASTER

WHAT'S  
THAT?



Source: <http://geekandpoke.typepad.com/>

geek and poke

## 2.5 Technologické zastarávání



### – Formát dat

- způsob zakódování informace do binárního zápisu (doc, docx, xls, jpg, tif, mp3, ...)
- tisíce formátů, nově vznikající, otevřené/proprietární formáty, obsoleté
- software nezbytný pro dekodování/zobrazení informace

### – Software

- Nástroje (textový procesor, tabulkový kalkulátor, videosoftware, DB-systém, poštovní klient)
- Aplikační software (IS MU, ekonomický systém, ... – interpretace dat v souborech)
- Operační systém (MS Windows, Apple OS, Linux, ...)

### – Hardware

- Počítač (IBM PC, Apple, superpočítač Cray, herní konzola, mobil)
- Periferní zařízení (vstupní, výstupní, ...)
- Síť

**!** Závislost dat na „správných“ technologiích je velký problém

## 2.6 Co je „Digital Preservation“



DP Handbook

### – Digital Preservation

The goal of digital preservation is the accurate rendering of authenticated content over time.

Digitální uchování kombinuje postupy, strategie a akce zajišťující přesnou reprodukci ověřeného obsahu v průběhu času, a to s ohledem na případná selhání záznamových médií a na probíhající technologické změny.

### – Dvě úrovně

- **Bit-level preservation** (dostupnost digitálních dat)
- **Logical preservation** (technologické změny + porozumění obsahu)

Terminologie: Digital Preservation (DP), LongTerm Preservation (LTP)

## 2.7 Ochranné strategie

a) **Nosič** (uchování digitálních dat – bitová ochrana)

– oživování a replikace

– nový formát

b) **Informace** (digitální obsah a jeho význam – logická ochrana)

– uchování technologického prostředí

- technologické muzeum

- emulace

– překonání technologické zastaralosti

- migrace

- encapsulation

! V praxi: kombinace přístupů (+ digitální archeologie, když vše selže)

## 2.8 Aspekty digitální ochrany

- Digitální ochrana zdaleka není jen problém technologický
- Ne izolovaný proces, součást širšího digitálního inf. prostředí
- Integrální součást celého životního cyklu digitální informace (nutná podpora od CMS – content management systémů)
- Součást běžného každodenního fungování organizace (ne kampaňovitost jako u analogových inf. zdrojů)
- Nejefektivnější když je pre-emptivní (problematická ex-post záchrana až při vzniku potíží)
- Nutná spolupráce různých subjektů (nakladatelé, paměťové instituce, komerční specializované firmy, instituce, jednotlivci)
- Je dražší než u analogových zdrojů (ale zatím neumíme cenu dlouhodobé archivace vyčíslit)



# Pro zájemce o další informace

**Digital Preservation Handbook**

2nd Edition

<https://www.dpconline.org/handbook>

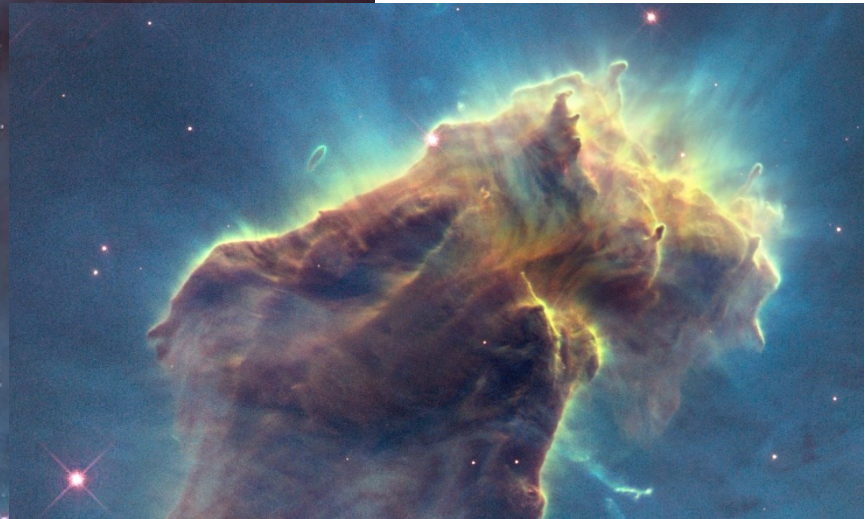
Digital Preservation Coalition © 2015



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark



## 3. Digitální archiv

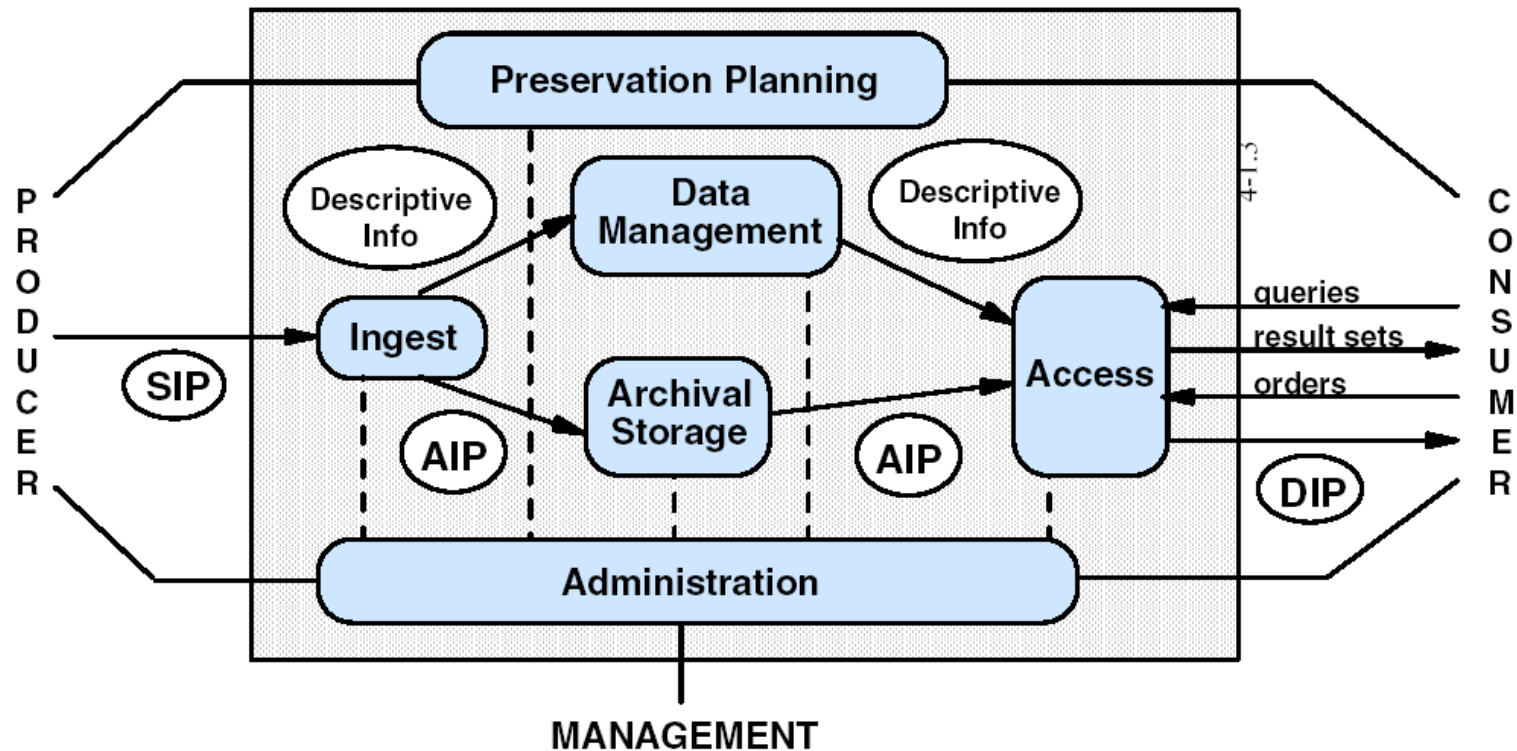


**Sloupy stvoření**  
Hubbleův teleskop, 1995

# 3.1 Systémová ochrana digitálních informací

- Digitální uchování je složité – jak to řešit „systémově“?
- Koncepte standardizovaného „Digitálního archivu“
- Podnět od kosmických agentur (NASA, ESA, ...)  
(obrovské objemy dat, spousta negativních zkušeností)
- 2002: **OAIS** – Open Archival Information System
  - Referenční model pro dlouhodobý Digitální archiv
  - ISO standard (od 2014 i ČSN)
- Pro naše potřeby příliš odborné/obsáhlé
- ...ale pár postřehů i pro osobní inspiraci...

## 3.2 OAIS



- **SIP** – Submission Information Package
- **AIP** – Archival Information Package
- **DIP** – Dissemination Information Package

# Papírový archiv



<https://www.ceskatelevize.cz/porady/1142743803-reporteri-ct/212452801240049/0/44211-archivalie-zidovskych-obcanu-na-ceskem-sternberku/>

# Digitální „archiv“ (superpočítač Frontera)



<https://news.utexas.edu/2019/06/19/ut-austins-frontera-named-5th-fastest-computer-in-the-world/>

## 3.3 Důvěryhodnost Digitálního archivu

**Důvěryhodnost** – prokázaná schopnost úložiště zachovat digitální dokumenty v dlouhodobém horizontu přístupné a použitelné

Různé stupně jak prokázat důvěryhodnost (certifikace)

- **Základní certifikace**  
(Data Seal of Approval, 16 guidelines, self-assessment)
- **Rozšířená certifikace**  
(self-audit podle ISO 16363)
- **Formální certifikace**  
(externí nezávislý audit dle ISO 16363 certifikovanou autoritou)
- ISO 16363:2012 – **Audit and certification of trustworthy digital repositories**



## 3. ...a osobní archivace



## 3.4 Osobní digitální archivace

- Má smysl starat se „rozumně“ o svá osobní digitální data
- Library of Congress: <https://digitalpreservation.gov/personalarchiving/>
- **Pár tipů:**
  - Zálohovat: alespoň 2 zálohy na separátních médiích
  - Zálohy uchovávat na vzdálených lokalitách
  - Média označit a držet v bezpečných místech (jako důležité dokumenty)
  - Namátkově ověřit čitelnost médií
  - Alespoň každých 5 let vytvořit nová záložní média
  - Systematicky roztrdit své osobní sbírky (fotografie, audio, video, e-mail, osobní)
  - Vybrat nejdůležitější materiály, rozumně popsat (jména souborů, metadata)
  - Používat rozšířené (otevřené) formáty, důležitá data migrovat na nové

**!** Další? Osobní zkušenosti?

## 3.5 Chmurné perspektivy?

- Velké množství digitálních informací již dnes nenávratně ztraceno
- Trvale roste množství informací existující pouze v digitální podobě
- Stále se rozšiřuje množství formátů dokumentů a médií
- Informační technologie zastarávají velmi rychle
- Rychlé zastarávání fyzických paměťových nosičů
- Při vytváření digitálních zdrojů není počítáno s náklady na archivaci
- Snižování finančních zdrojů pro knihovny a archivy
- Nesmyslné ochrannářské trendy omezující dostupnost info (právo být zapomenut)

„There is, at present, no way to guarantee the preservation of digital information. The first line of defense against loss of valuable digital information rests with the creators, providers, and owner of that information. It's every man for himself.“

U.S. Commission on Preservation and Access.  
Final report of a Task Force on the Archiving of Digital Information. 1996

## 3.5 Chmurné perspektivy?



Analog objects can survive with minimal care for centuries, but no electronic format can hope to persist more than a short while without careful (and perhaps expensive) intervention. There will be no digital equivalent of the Lascaux cave paintings, Mayan stone scripts, Dead Sea scrolls, or other kinds of rediscovered ancient knowledge.

Building Preservation Partnership. The LoC NDIP Program

## 4. The Long Now

## 4.1 The Long Now

"When I was a child, people used to talk about what would happen by the year 2000. For the next thirty years they kept talking about what would happen by the year 2000, and now no one mentions a future date at all. The future has been shrinking by one year per year for my entire life. I think it is time for us to start a long-term project that gets people thinking past the mental barrier of an ever-shortening future.

I would like to propose a large (think Stonehenge) mechanical clock, powered by seasonal temperature changes. It ticks once a year, bongs once a century, and the cuckoo comes out every millennium."

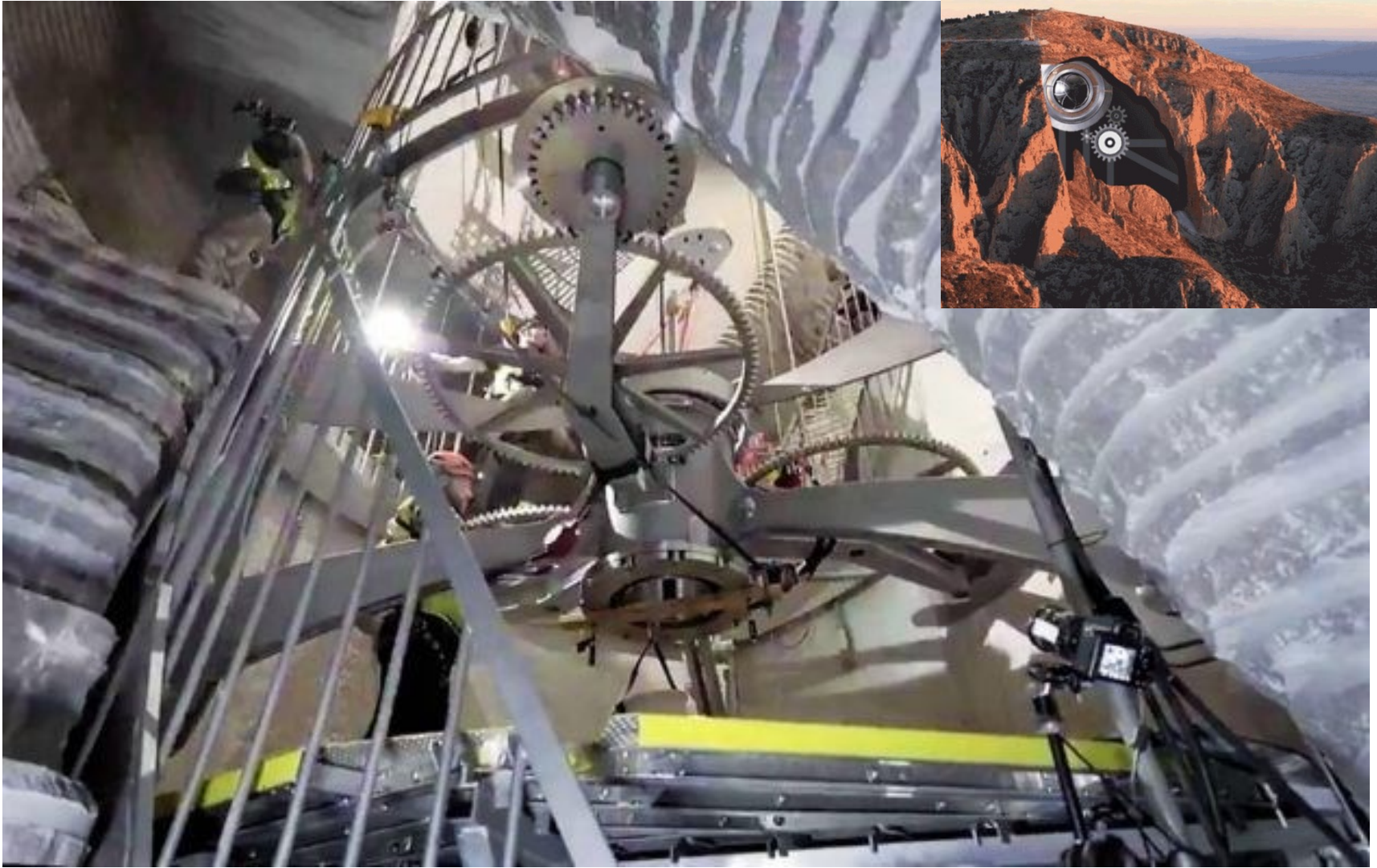
**Daniel Hillis, The Long Now Foundation**  
<http://www.longnow.org>

The Long Now Foundation was established in 1996 *to creatively foster long-term thinking and responsibility in the framework of the next 10,000 years.*

## 4.2 The 10.000 year Clock

- Ikonický projekt pro dlouhodobé uvažování
- Postavit mechanické hodiny (udržovatelné „technologiami doby bronzové“) které přežijí a budou fungovat 10 tisíc let!
- **Výzvy:**
  - Jaké použít materiály, technologie, aby vydržely
  - Čím hodiny pohánět
  - Jak zajistit správné zobrazování času po tak dlouhou dobu
  - Jak prezentovat čas/datum na hodinách budoucím uživatelům
  - Ochrana proti vandalství (aby lidé hodiny sami nezničili), ...
- 01999 prototyp: 28bitový mechanický počítač (Science Museum, Londýn)
- Zakoupen pozemek v poušti v Texasu
- Výstavba hodin ve skále (výška 200 stop)





Workers install the 10,000 Year Clock inside an underground chamber in Texas. (Long Now Foundation)

## 4.3 Rosetta Disc

- Dlouhodobý lingvistický archív a překladatelský nástroj pro obnovu jazyků ztracených v hluboké budoucnosti
- Inspirace Rosettskou deskou (rozluštění egyptských hieroglyfů)
- Paměťové médium: 3“ niklová deska, litograf. nano-technologií vyrytý analogový text a obrázky (až 200.000 stran, životnost 2K-10K let)
  - Archiv (Languages of the World)
  - Cíl: lingvisticky podrobně zdokumentovat 7.000 současných jazyků světa
  - Pro každý jazyk stejná struktura informací:  
podrobný lingvistický popis jazyka, fonologie, gramatika, stejné texty (Genesis, OSN – Deklarace lidských práv), mapy a další
  - Aktuální stav: 70.000 stran o 2.500 jazycích
- **Rosettský disk v1**
  - 13.000 stran informací o 1.500 jazycích
  - čitelné při 650násobném zvětšení
  - dar při 10 000 Membership  
(předávat z pokolení na pokolení)





# Diskuse