

Digitální knihovny



Miroslav Bartošek

Ústav výpočetní techniky MU

Knihovnicko-informační centrum MU

Obrazový kvíz

Znáte tato místa/zařízení?

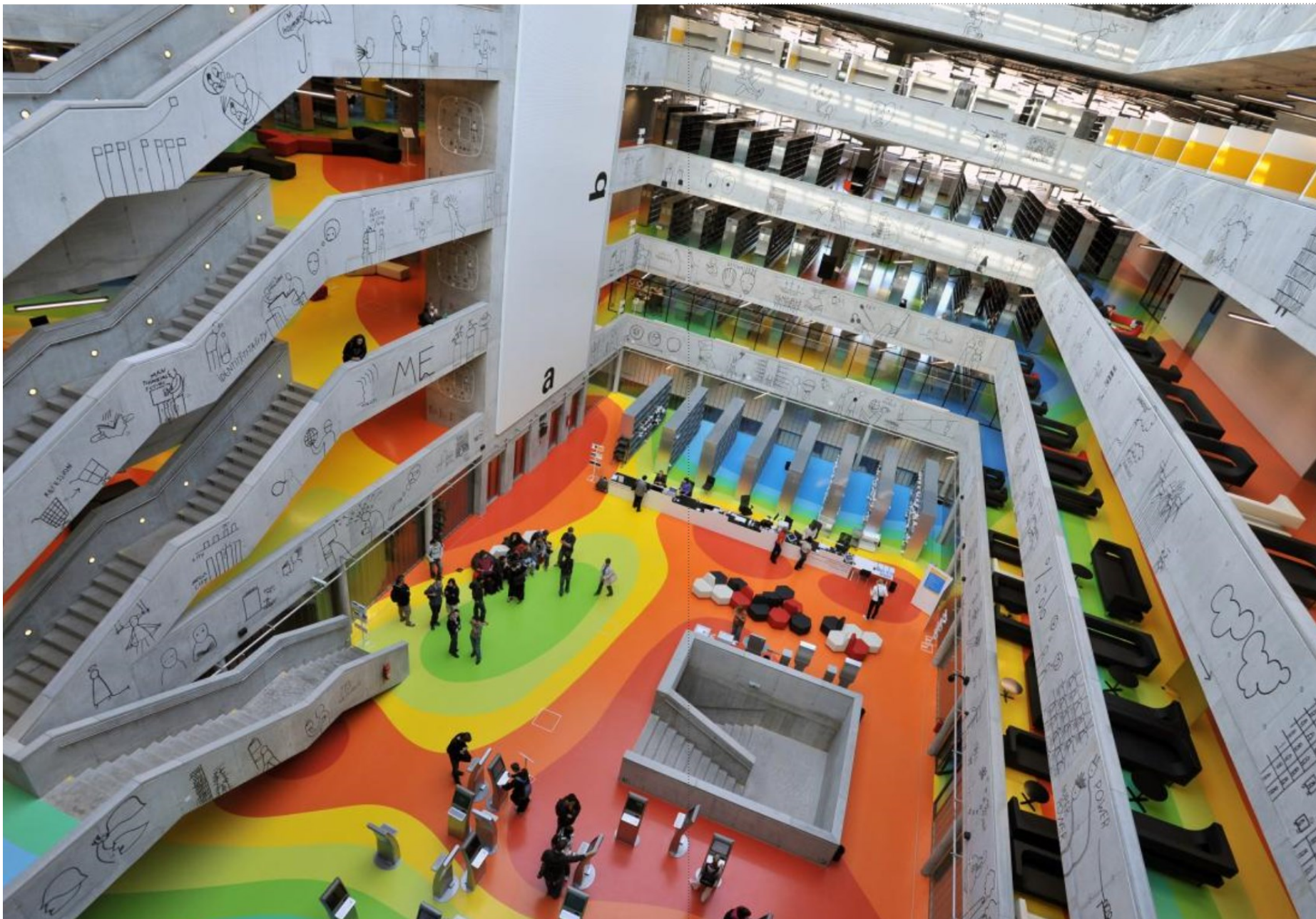


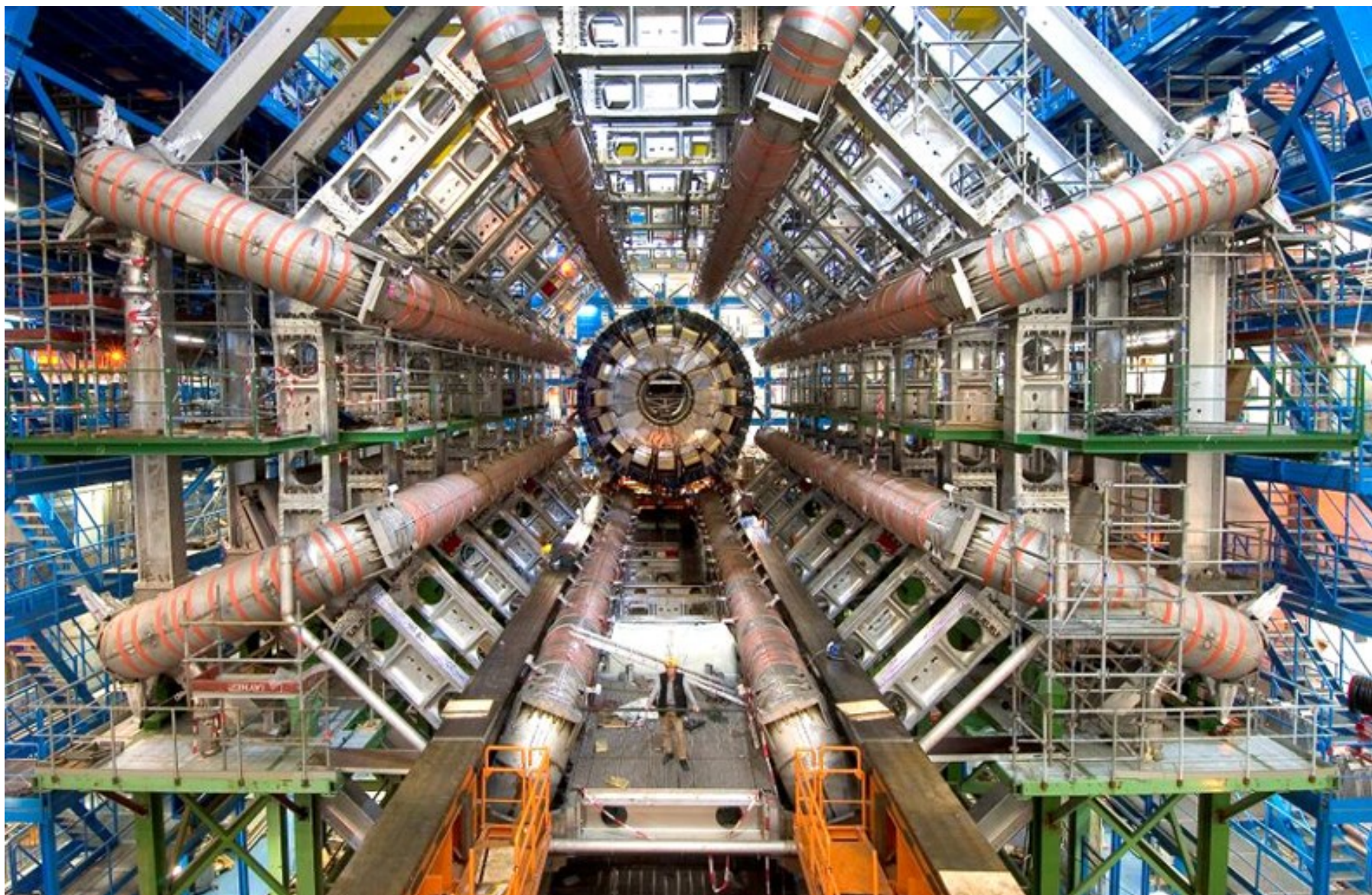


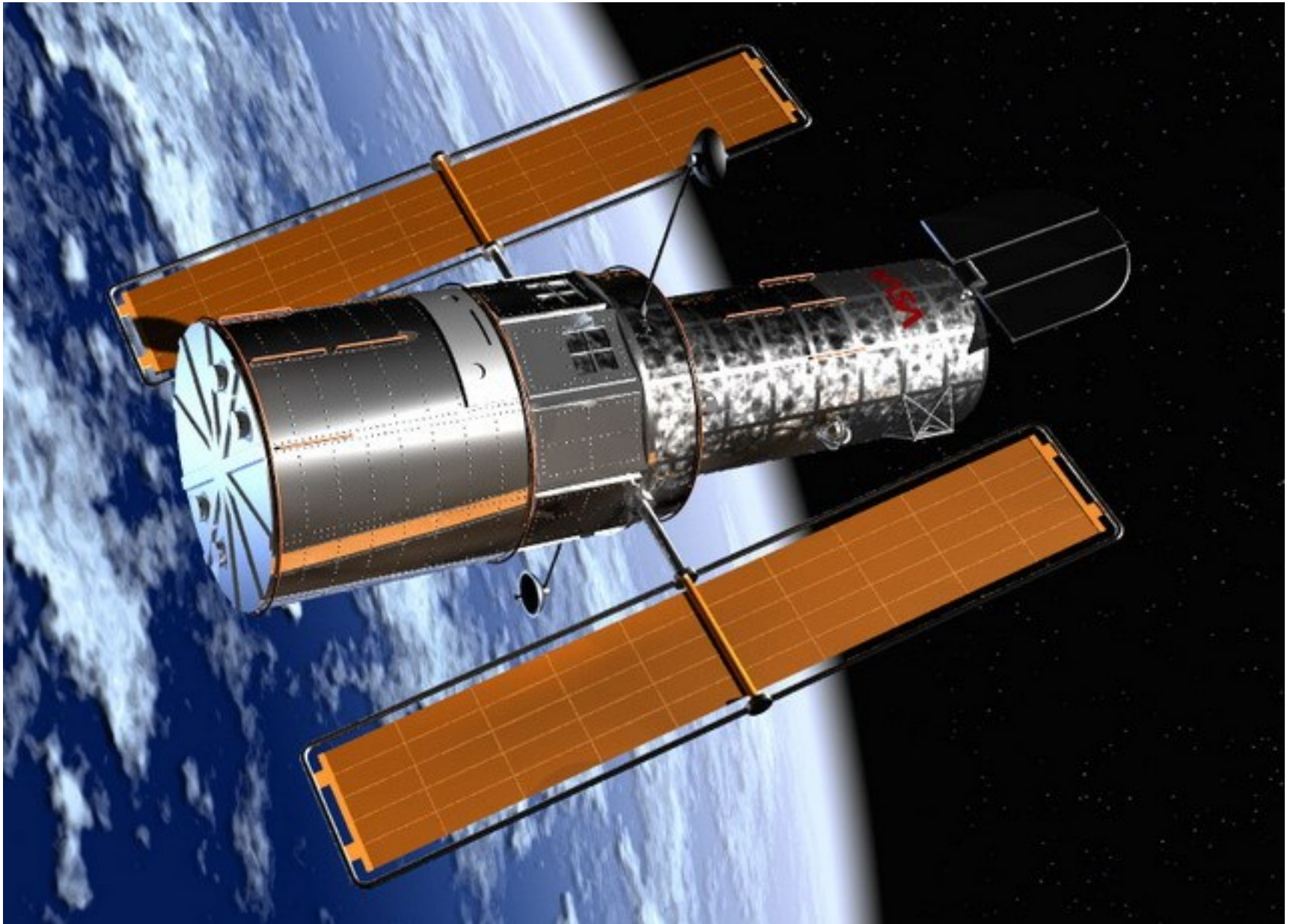












Co mají tato „zařízení“ společného?

Rozprava na úvod

- Byli jste / chodíte do nějaké knihovny?
- Digitální knihovna (DL) – co to je?
- Znáte nějaké digitální knihovny?
- Vztah
 - DL x klasické knihovny
 - DL x informační systémy

PV070

Obsah a požadavky

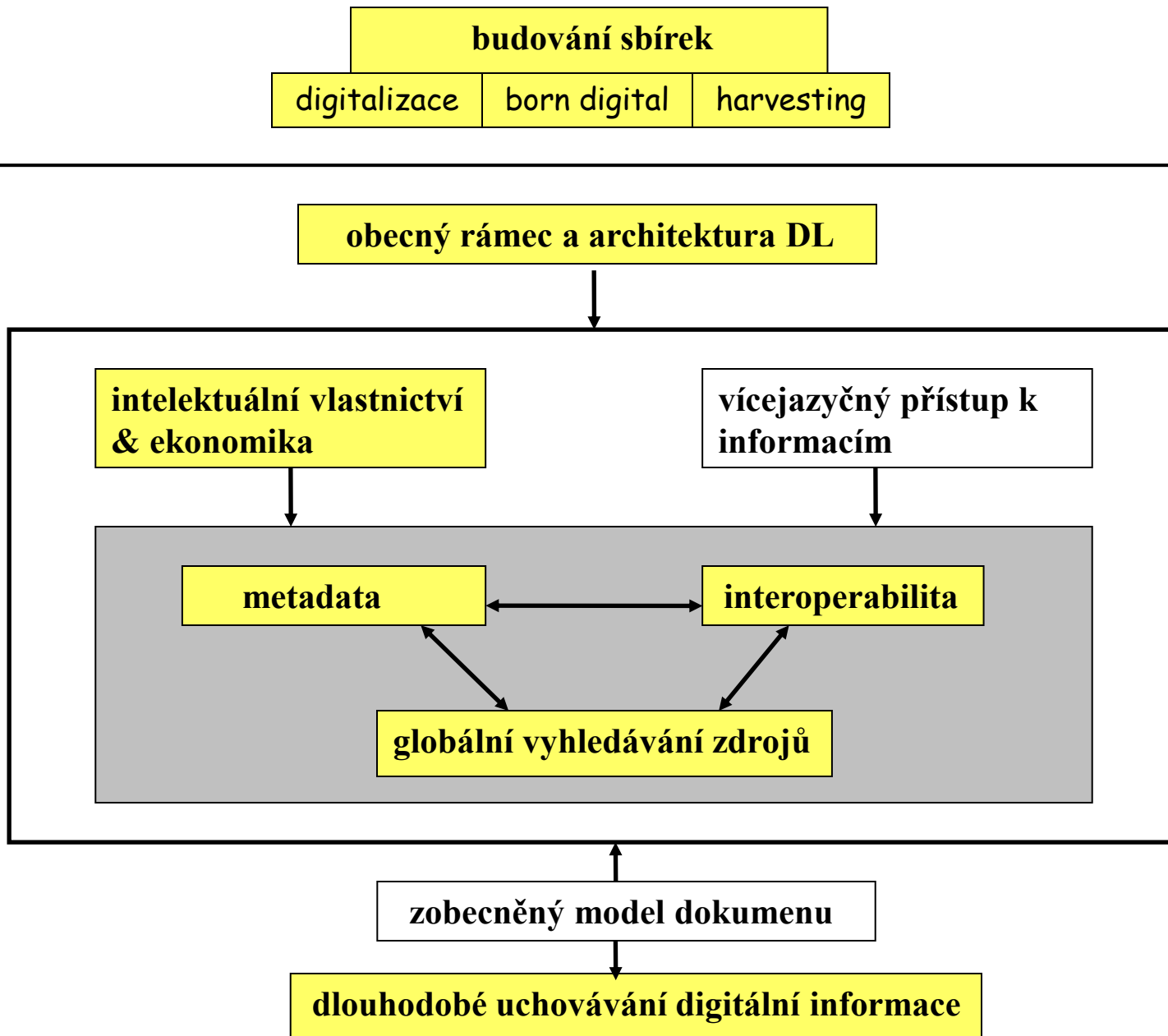


Požadavky k absolvování PV070

1. Znalost problematiky a souvislostí
(**záznamy přednášek**, prezentace z přednášek, doporučená doplňková literatura)
 2. Písemná esej o zvolené DL/projektu/technologii
 3. Ústní zkouška
- Prezentace z přednášek, témata esejí
<http://www.ics.muni.cz/mba/dl-fi2020>

Osnova předmětu PV070

1. Co jsou to digitální knihovny, úvod
2. Digitalizace
3. Obecný rámec a architektura DL
4. Jména a identifikátory
5. Metadata
6. Interoperabilita
7. Globální vyhledávání zdrojů
8. Ekonomika a legislativa (IPR)
9. Otevřený přístup (open access)
10. Problém dlouhodobého uchování (digital preservation)
11. Prezentace esejí (zajímavé projekty a technologie)



1. přednáška

Úvod do digitálních knihoven



Obsah 1. přednášky

1. „definice“ DL
2. DL versus WEB
3. DL versus knihovny
4. Krátce z historie DL
5. Literatura – zdroje informací o DL
6. DL a vědecké publikování

1.1 „definice“ DL

klasický počítačnický pohled:

- Digitální knihovna je **spravovaná sbírka informací** spolu s odpovídajícími **službami**, přičemž informace jsou uloženy v **digitální podobě** a jsou dostupné prostřednictvím **sítě**. *W.Y.Arms, 2000*
 - udržovaná, spravovaná dig. sbírka (maintained)
 - služby
 - vzdálený přístup

- **Focused collection of digital objects**, including text, video, and audio, along with methods for **access and retrieval**, and for **selection, organization, and maintenance**. *I.W.Witten, 2002*
 - digitální obsah (text, video, audio, 3D, simulace, dynam.vizualizace..)
 - uživatel (access and retrieval)
 - „knihovník“ (selection, organization, and maintenance)

1.1 „definice“ DL

knihovnický pohled:

- Digitální knihovny jsou **organizace**, které poskytují zdroje (včetně specializovaného personálu) umožňující provádět **výběr, strukturování a zpřístupnění** sbírek digitálních prací, tyto práce dále distribuovat, udržovat jejich integritu a dlouhodobě uchovávat – a to vše s ohledem na snadné a ekonomické využití **určitou komunitou** nebo množinou komunit uživatelů.

US Digital Library Federation, 1997

- DL jako „instituce“ (např. knihovna)
- důraz na systematickou organizaci a služby
- cílenost DL (využívání určitou komunitou)

1.1 „definice“ DL

archivační pohled:

- DL = the **infrastructure**, **policies** and **procedures**, and organisational, political and economic mechanisms necessary to **enable access** to and **proservation** of digital content. *Ross, 2003*
 - DL jako komplexní infrastruktura a systém pro zajištění
 - a) přístupu
 - b) dlouhodobého uchování digitální informace
 - DL jako „archív“

1.1 „definice“ DL

- Digitální knihovna je on-line sbírkou digitálních objektů spolehlivé kvality, jež jsou vytvořeny či shromážděny a spravovány podle mezinárodně přijatých zásad pro budování sbírek a jsou zpřístupňovány promyšleným a dlouhodobě udržitelným způsobem, podporovány službami, jež jsou nezbytné k tomu, aby uživatelé mohli tyto zdroje vyhledat a využívat.

IFLA/UNESCO Digital Library Manifesto, 2011

**International Federation
of Library Associations
and Institutions**



1.2 Obecné znaky DL

- klíčová je **organizace** dig.sbírky + služby (ne digitalizace)
- DL **není jedna** uzavřená entita (-> DLs)
- inf.zdroje: **heterogenní, dynamické a multimediální**
- technologie pro **propojení autonomních inf. celků**
- transparentní propojení, standardy
- **koherentní přístup** bez ohledu na
 - formu
 - formát
 - způsob a místo uložení informací
- garance **dlouhodobého uchování** digitálního obsahu

1.3 Příklad DL (1)

AMERICAN MEMORY



LIBRARY of CONGRESS

American Memory

Library of Congress

- Digitalizační „Apollo project“ (pilotní projekt 1990-1994)
- 120+ historických sbírek
- > 10 miliónů digit. položek
- knihy, fotky, rukopisy
audio, video, mapy



<https://www.loc.gov/collections/>

1.3 Příklad DL (2)

JSTOR

- 2600 zdigitalizovaných časopisů (od 1.ročníku – po moving wall)
- >100 miliónů naskenovaných stran
- řada tématických kolekcí
- arts, humanities, soc-sciences
- 8.000+ institucí ze 160 zemí
- udržitelný ekonomický model
 - InSTITUTE – poplatek jednorázový/roční
 - INDIVIDUALS – Register & Read (6 článků/30dnů)
 - EARLY JOURNAL CONTENT (free)

<http://www.jstor.org/>

Journal Storage

FOR LIBRARIANS FOR PUBLISHERS FOR INDIVIDUALS SEARCH

JSTOR

Light up your mind.
Scholarly journals, primary sources, and now books!

The seeds of something new

Search, view, comment, and tag 2+ million plant type specimens and other primary sources from hundreds of herbaria worldwide.

For more information go to >>

THE BUZZ

- RT @rschon: Antisocial in Social World – Scholarly Kitchen on publisher / society implications of the @IthakaSR US Faculty Survey <http://t...> @JSTOR
- "An Account of the Blue Shark, Together with a Drawing of the Same." <http://t.co/UnszDqgNCF> 1778 article #sharkweek @JSTOR

JSTOR EVIDENCE IN UNITED STATES VS. SWARTZ

Read more

10 THINGS YOU SHOULD KNOW ABOUT JSTOR

1. JSTOR is a not-for-profit, founded to help academic libraries and publishers.
2. JSTOR is not a publisher, and we do not ask for exclusive rights to the content on the platform.

JSTOR IN NUMBERS

151M

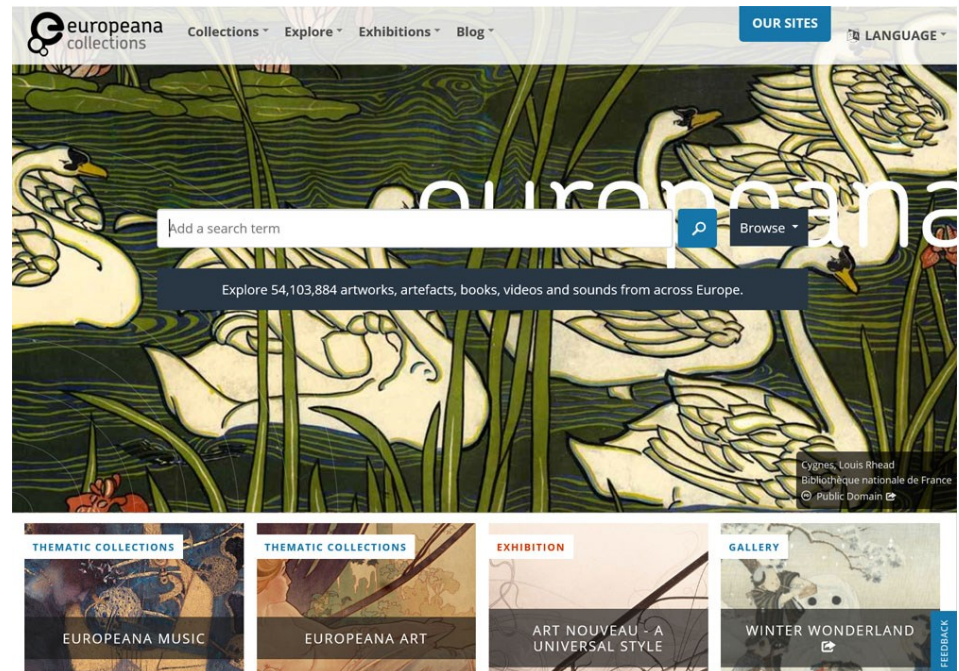
Searches Performed in 2012

1.3 Příklad DL (3)



Europeana

- **Portál pro evropské kulturní dědictví**
- EC = iniciátor a financování
- 2010 první verze
- nyní:
- Data od 4.000 institucí ze 48 zemí
 - Knihovny muzea, archivy, galerie...
- 60 miliónů objektů
 - obrázky, texty, zvuk, video
- Metadata, náhled + link na zdroj v domovské DL



<http://europeana.eu/>

1.3 Příklad DL (4)

Klíčové komerční digitální knihovny pro Computer Science

- **Springer – Lecture Notes in CompSci (LNCS)**
- **Digitální knihovna ACM**
- **IEEEXplore**
- **Knihovny e-časopisů Elsevier, Springer, Wiley**
 - **ScienceDirect (Elsevier), SpringerLink, ...**

<https://ezdroje.muni.cz> (Portál elektronických informačních zdrojů MU)

DL x Web



2.1 DL x Web

- ***Proč DL?***
Máme přece Web a tam je „vše“ !
- Je Web digitální knihovnou ?

2.2 Web je skvělý, ale...

- obrovské množství informací, snadný přístup
- společná technologie
- pokroky v oblasti vyhledávání (Google)
- autor sám sobě nakladatelem
- široká dostupnost

Ale:

- vyhledávání
- správa práv
- trvalá dostupnost
- autenticita
- kvalita



2.3 Rozdíl DL x Web

- základním rozdílem je **organizace**
(web jako celek není organizován)
- Inet vyhledávače většinou spoléhají na jednoduché **fulltextové vyhledávání** (podle „klíčových“ slov)
- další důležitý rozdíl – **řízení vstupu dat** do systému
(DL – pevnější řízení, kontrola vstupu, kvalitní metadata)
- Web **není cílen** na konkrétní komunitu uživatelů
- krátký **poločas rozpadu** dokumentů na Webu
- povrchový a skrytý **hluboký Web** (deep web)

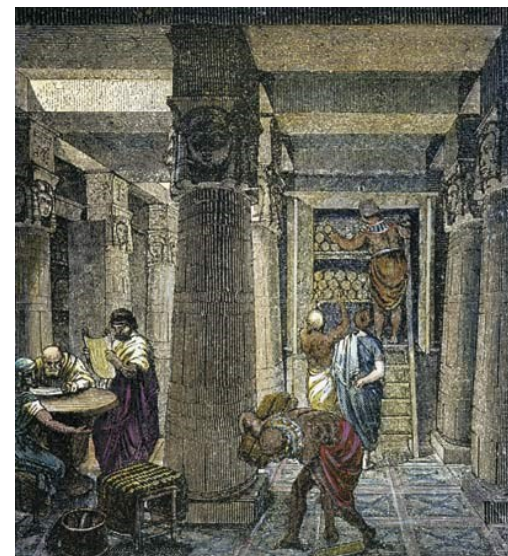
DL x knihovny



3.1 Společné rysy DL a knihoven


- systematicky budovaná sbírka **datových objektů**
- sbírky **metadatových struktur** (katalogy, rejstříky, indexy)
- množina **služeb** (přístup pro různé třídy uživatelů, IPR management, správa, vyhodnocování, SDI)
- **tématické zaměření**
- sledování **kvality** (nezařazuje se cokoliv)
- **dlouhodobé uchovávání** (v knihovnách - **staletí** !)

Alexandrijská knihovna: Egypt - Ptolemaiovci, 295 př.n.l. -- ??
700.000 papyrusových svitků (originály Euripida, Aischyla,
Sofokla, Euklida...)



3.2 DL = evoluce knihoven?

Tři typy knihoven:

- 
- 1. tradiční knihovna** (zaměřená na *fyzické objekty*)
 - 2. hybridní knihovna** (současně fyzické i digitální sbírky)
 - 3. digitální knihovna** (pouze *digitální objekty*)

-
- DL = plně digitalizovaná knihovna + síť + interoperabilita
 - DL nemusí vzniknout z klasické knihovny
 - DL jako informační systémy – počítačníci

3.3 Přednosti knihoven

- staletá **tradice** v organizaci/správě/zpřístupňování informací
(evropský versus americký přístup)
- jednotné **standards** a doporučení (MARC, AACR2, ...)
- propracovaný a fungující **system** knihoven
(hierarchický národní systém, různé typy knih., systém služeb - MVS, katalog.)
- **legislativa**, ustálené vzorce chování
(postavení knihoven ve společnosti, povinný výtisk, fair-use)
- vybalancovaný systém
autoři – nakladatelé – knihovny - čtenáři

3.4 Přednosti DL

a) zlepšení klasických knihovních služeb

- vzdálený, nepřetržitý přístup 24/7
- efektivnější vyhledávání
- lepší využití fondu
- sdílení informací
- lepší ochrana (bity, ne atomy)
- prostorové úspory

b) nové možnosti

- integrace a propojování informací
- nové formy/formáty informací
- permanentní aktualizace informací
- informace od uživatelů
- nové typy služeb a modely šíření informace (viz STI)

3.5 Přeměna knihoven v DL?

- „Paper libraries would disappear by 1984.“

Arthur Samuel (1964, The Banishment of the Paperwork.)

- „Some say that had books been invented after computers were, they would have been hailed as a great advance.“

Ian H.Witten (2002, How to Build a Digital Library.)

3.5 Přeměna knihoven v DL???

Proč to tak dlouho trvá?

1. o několik řádů složitější problém sdílení (globální infrastruktura)
2. velmi různorodá povaha digitálních objektů (variabilita typů a formátů)
 - *text* (nestructurovaný, strukturovaný, hypertext, ...)
 - *image* (kresba, obrázek, foto, mapa, RTG-snímek, ...)
 - *audio* (hlas, hudba, zvuky)
 - *video* (klip, film, video, animace)
 - *3D objekty, dynamické dokumenty, multimédia, datasey, ...*
3. VELMI velké objemy dat (tera-TB a peta-PB = 10^{15})
 - LoC: 23 mil svazků = 20 TB
 - CERN-LHC (Large Hadron Collider, 27km obvod):
15TB/den : přenést/uložit/zpracovat/zpřístupnit (LHC Grid, 10PB/ročně)
4. nepřipravené sociálně-společenské prostředí
 - chybějící legislativa a postupy
 - nová ekonomika procesů
 - nové možnosti x hrozba destabilizace zavedeného systému

Historie a současnost DL



4.1 Krátce z historie



- 1945 Vannevar Bush („As We May Think“, Memex)
 - 1965 J.C.R. Licklider („Libraries of the Future“)
 - 60. léta MARC, OPAC (LoC, OCLC)
 - 80. léta fulltexty
-
- poč. 90 **Computing + Communications + Contents**
(nízká cena, vysoký výkon, široká dostupnost)
 - 1994 Digital Library Initiative , www
 - nyní: budování infrastruktury a pokročilých systémů

4.2 Současný stav

- *teorie*: široký obecný výzkum, mapování klíčových problémů, návrhy prvků obecné infrastruktury
- *praxe*: digitalizace sbírek, izolované DL (tisíce projektů), ustanovení obecných infrastrukturních prvků
- **posun od experimentování (90.léta) k budování glob. infrastruktury:**
 - řada usazených „základních“ kamenů
(XML, RDF, Web Services, DC, OAI-PHM, ...)
 - budování „přízemí“
(METS, OpenURL, OpenDL, OAI, ...)
 - překryv s výzkumy v oblasti **E-commerce, GRIDy, Sémantický web**
(interoperabilita, bezpečnost, IPR, distribuovanost, velké objemy dat, strojové porozumění)
- řada komerčních řešení (IBM DL -> Content Manager, DigiTool)
- řada open-source řešení (DSpace, FEDORA, EPrints, Greenstone...)
- národní programy rozvoje DL (DLI-1,2 , i2010, VISK, ...)

DL a vědecké publikování

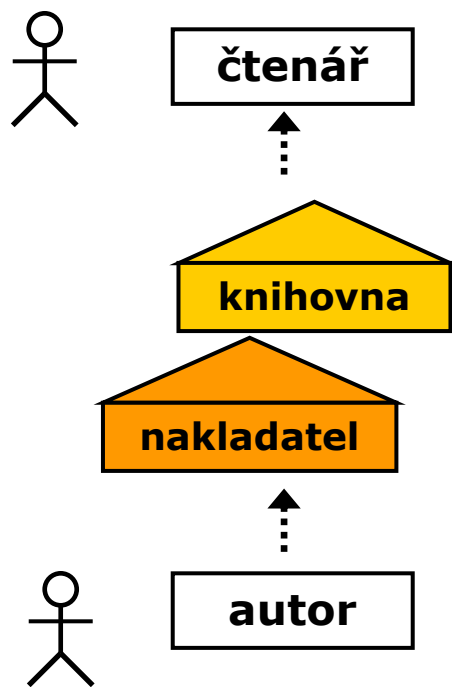


5.1 DL a oblast STI

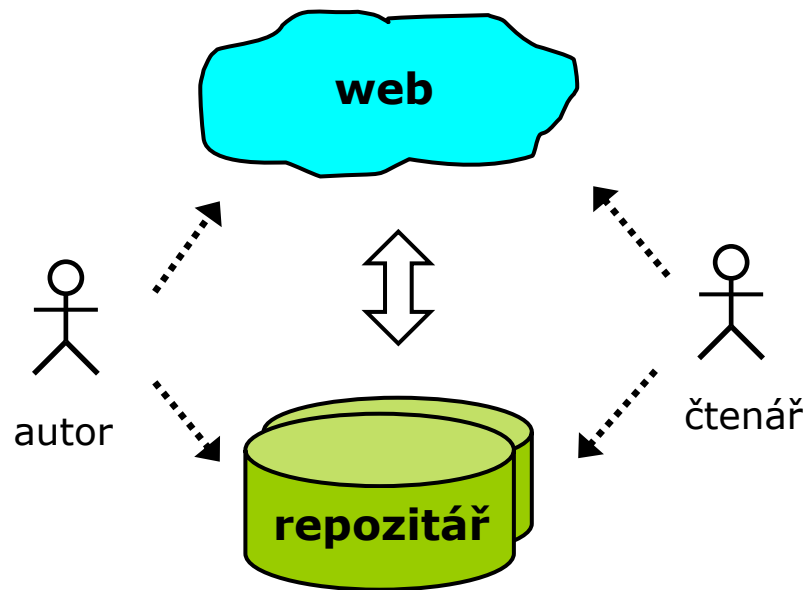
- **Scientific & Technical Information**
- Internet – z potřeby efektivnějšího šíření STI
- DL – další krok v tomto směru
 - STI – experimentální oblast nových technologií a přístupů z oblasti DL, **předvoj**
 - hlavním hnacím motorem pro aplikace DL v STI není ekonomika (ve smyslu generovat profit)
 - následné ekonomické zhodnocení – motor pro široké komerční nasazení
- STI je jen malý výsek současných aplikací DL (ale velmi důležitý)

5.2 Změny publikování v STI

a) dříve



b) nyní



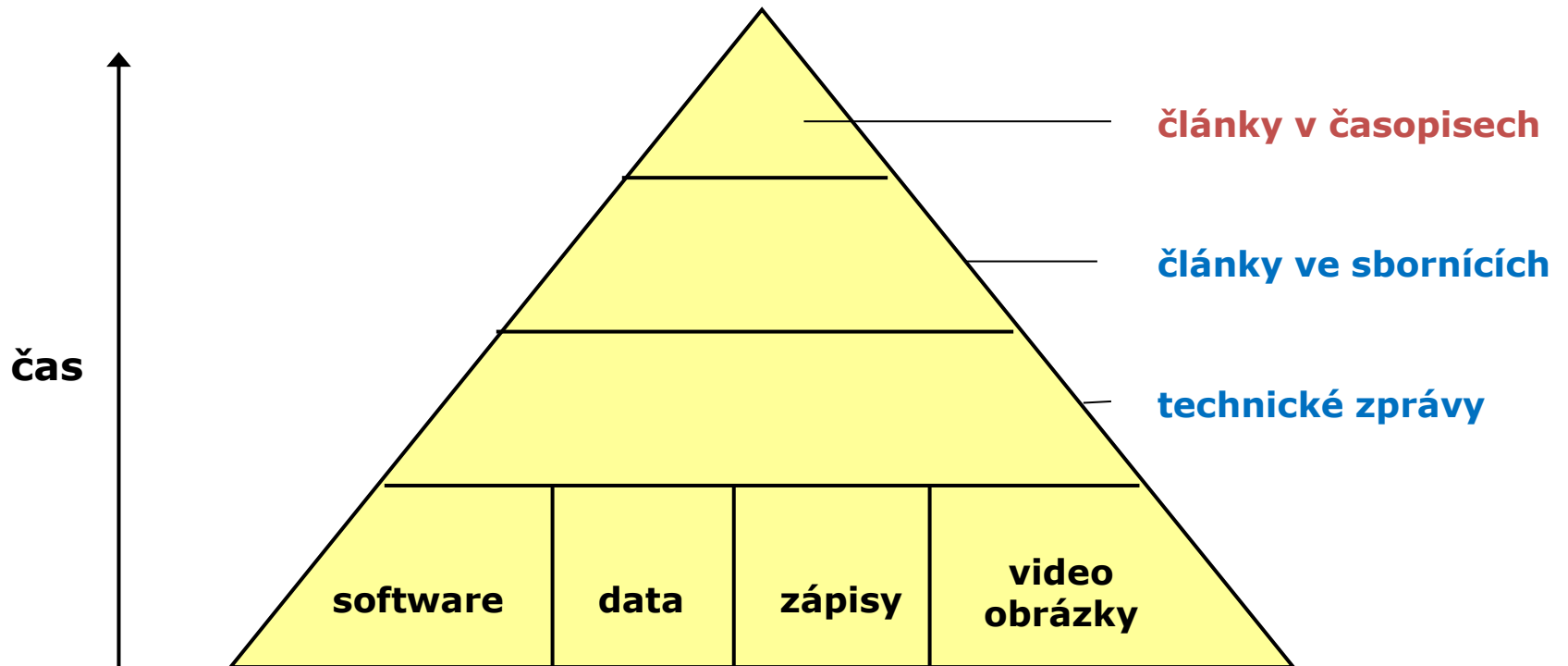
5.3 Literatura v STI

- STI - soubor informačních materiálů (bez ohledu na jejich formu a formát) pro potřeby vědy, výzkumu a vývoje (články, zprávy, experimentální data, měření, obrázky, videa, software, ...)
 - **bílá literatura** (knihy, časopisy, důležité sborníky)
 - autor a vydavatel jsou obvykle různé subjekty
 - posuzování kvality předcházející publikování (**peer review**)
 - **šedá literatura** (grey literature)
publikace nedostupné běžnými prodejními kanály, produkované často v omezeném množství, s nízkou či žádnou inzercí a s omezeným využitím (disertace, výzkumné zprávy, články na Inetu)
 - el. publikování – velký impuls pro „šedivku“

5.4 Proč se starat o šedou lit?

- bílá literatura
 - + vysoký stupeň důvěryhodnosti
 - + snadno dostupná
 - dlouhý čas mezi napsáním a publikováním
 - často jen výtah z podstatně většího materiálu
- *některé technologie mohou zastarat dříve, než informace o nich pronikne z literatury „šedé“ do literatury „bílé“*

5.5 Pyramida STI



Pyramida publikovaných info leží na základech nepublikované STI

5.6 Ekonomika STI

- 25.000 vědeckých časopisů celosvětově (peer-reviewed)
 - 1.5 mil recenzovaných článků / rok
 - 2000 vydavatelů (giganti Elsevier, Springer)
- počet vědeckých publikací se zdvojnásobuje každých 10-15 let (adekvátně roste i počet časopisů)
- velikost čtenářské obce konstantní, specializace
- rostoucí cena vědeckých časopisů
(Chemical Abstract: desítky USD v 1950, 31.600 USD dnes)
- stagnující či klesající rozpočty knihoven
 - [serials crisis](#)
- ***přítom časopisy = jen zlomek STI informací***

5.7 Jak z toho ven?

- potřeba nových modelů pro šíření STI
 - nové technologie (digitální knihovny, repozitáře)
 - ekonomické modely (open access)

šance a motivace pro DL

Literatura a DÚ



Literatura

Základní literatura

- I.H.Witten: **How to Build a Digital Library**. Morgan Kaufmann Publ. 2002, 2010
- M.Bartošek: **Digitální knihovny – teorie a praxe**. Národní knihovna 2004, 15/4
<http://full.nkp.cz/nkkr/NKKR0404/0404233.html>
- Miroslav Bartošek. **Technologie digitálních knihoven**. In Proc. INFORUM 2006.
http://www.inforum.cz/pdf/2006/Bartosek_Miroslav.pdf

Praktický úkol – vyzkoušet si

- knihovny: **JSTOR, American Memory, Europeana**
- Portál EIZ MU <https://ezdroje.muni.cz>

Zajímavost

- Vannevar Bush. **As We May Think**. Atlantic Monthly, July 1945
<http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881/>

