

# Digitální knihovny

## Úvod



Miroslav Bartošek  
MU Brno, Ústav výpočetní techniky, C212  
bartosek@ics.muni.cz

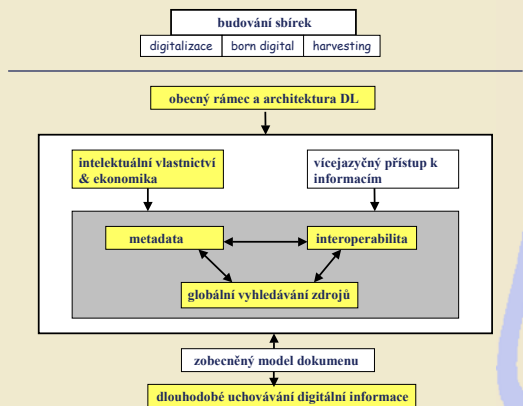
## Přístup k problematice DL

DL = *velmi široké téma!*

- ~~• technický přístup (standards, formáty, digitalizace)~~
- ~~• informační věda (knihovnicko-„sociální“ pohled)~~
- ~~• zoologie projektů (project-centric přístup)~~
- **obecný popis hlavních oblastí ve výzkumu a praxi DL (z pohledu počítačníků)**
  - charakteristika základních přístupů
  - příklady praktických řešení (zákl.kameny)
  - přístup „shora-dolů“

## Témata předmětu „DL“

Co jsou to Digitální knihovny  
Obecný rámec a architektura  
Jména a identifikátory  
Metadata  
Interoperabilita  
Globální vyhledávání zdrojů  
Ekonomika a legislativa (IPR)  
Problém dlouhodobého uchování  
Zajímavé programy a projekty



1. přednáška

## Co jsou to Digitální knihovny



## Obsah 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 Co je to DL ? (DL = Digital Library)

7

### Problém vymezení:

- ✓ pojem se vyvíjí spolu s inf. technologiemi  
(knihovna bez zdi, elektronická knihovna, virtuální knihovna, digitální knihovna)
- ✓ rozdílná profesní chápání
  - databázista: DL = systém federativních databází
  - IS-pracovník: DL = aplikace nad webem
  - knihovník: DL = další vývojová etapa knihoven
- ✓ jiné významy  
(knihovny procedur, systémy na správu dokumentů, ...)

## 1.2 Definice DL (1)

8

### počítačnický pohled:

- Digitální knihovna je **správaná 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)

## 1.2 Definice DL (2)

9

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

## 1.3 Obecné znaky DLs

10

- klíčová je **organizace** digitální sbírky
- DL **není jedna** uzavřená entita (DLs)
- inf.zdroje: **heterogenní, dynamické a multimediální**
- technologie pro **propojení** autonomních inf. celků
- **transparentní propojení**
- **koherentní přístup** bez ohledu na formu, formát, způsob a místo uložení informací

## DL x Web



## 2.1 DL x Web

12

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

## 2.2 Web není DL

13

WEB = DL? **ne!** (1995)

- "Although the Internet provides access to an enormous amount of information, the current state-of-the-art falls far short of what is commonly viewed as a *library service* – that is, relatively easy navigation of and access to a set of documents that are part of a collection. The notion of a *collection* is important in that it implies that the set of documents was not selected haphazardly, but by some trusted intermediary. Current users of the Internet confront an information space where the quality of documents is far from reliable, facilities for locating documents are primitive, and access to a specific document frequently means wading through a Tower of Babel of architecture dependencies and file formats."



Carl Lagoze, Cornell University  
„Dienst: Implementation Ref.Manual“, 1995

## 2.3 Web se blíží DL, ale ...

14

WEB = DL? **možná, ale...** (2002)

- obrovské množství informací, snadný přístup
- společná technologie
- pokroky v oblasti vyhledávání (Google) (přesnost vs odezva)
- autor sám sobě nakladatelem
- harvested DL (WebBase)

**Ale:**

- rights management
- preservation
- authenticity
- quality



## 2.4 Rozdíly DL x Web

15

- 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)
- 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 DL x knihovny - společné rysy

17

Mnohé společné rysy:

- 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 uchování* (v knihovnách - staletí!)

## 3.2 Knihovny - východisko DL

18

- knihovny = inspirace a východisko DL (postupy a přístupy)
- tradiční knihovny (zaměřené na *fyzické objekty*)
- hybridní knihovny (současné fyzické i digitální sbírky)
- digitální knihovny (plně digitalizované)
  - výhody z odstranění fyzické podstaty a bariér (viz dále)
  - nevýhody -dtto-

- DL = plně digitalizovaná knihovna + glob.interoperabilita (obecně ale DL ↔ digitalizovaná knihovna)
- řada dalších výchovných a sociálních rolí tradičních knihoven

### 3.3 Knihovny - přednosti

19

- staletá tradice v organizaci/správě/zpřístupňování informací (evropský versus americký přístup)
- jednotné standardy 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 Knihovny - přínosy digitalizace

20

#### a) zlepšení klasických knihovnických služeb

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

#### b) nové možnosti

- globální integrace
- nové formy/formáty informací
- zcela nové typy služeb a modely šíření informace (viz STI)

### 3.5 Knihovny - v digit. prostředí

21

Proč je to tak složité

- o několik řádů vyšší problém sdílení (globální infrastruktura)
- velmi různorodá povaha digitálních objektů (variabilita typů a formátů)
  - *text* (nestrukturovaný, strukturovaný, hypertext, ...)
  - *image* (kresba, obrázek, foto, mapa, RTG-snímek, ...)
  - *audio* (hlas, hudba, zvuky)
  - *video* (klip, film, video, animace)
  - *3D objekty, dynamické dokumenty, multimedia, datasety, ...*
- správa **VELMI** velkých sbírek (tera-TB a peta-PB = 10<sup>15</sup>)
  - LoC: 23 mil svazků = 20 TB
  - CERN-LHC (2005): x-TB/den : přenést/uložit/zpracovat/zpřístupnit
- nepřipravené sociálně-společenské prostředí
  - chybějící legislativa a postupy
  - nová ekonomika procesů
  - nové možnosti x hrozba destabilizace zavedeného systému

### 3.5 Knihovny - v digit. prostředí

22

Proč je to tak složité ?

- „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.)

## Historie a současnost DL



### 4.1 Krátce z historie

24

- 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: od experimentování k budování infrastruktury

## 4.2 Současný stav

- *teorie*: velmi široký obecný výzkum, mapování klíčových problémů, návrhy prvků obecné infrastruktury
- *praxe*: digitalizace sbírek, izolované DL (tisíce projektů), pokusy o ustanovení obecných infrastrukturních prvků
- žádné jednotné řešení, ale již zřetelný posun od experimentování (90.léta) k budování globální infrastruktury:
  - řada usazených „základních“ kamenů (XML, RDF, DC, OAI-PHM, ...)
  - budování „přízemí“ (METS, OpenURL, OpenDL, OAI, ...)
  - překryvy s výzkumy v oblasti E-commerce (interoperabilita, bezpečnost, IPR)
- několik komerčních řešení (IBM DL → Content Manager)
- řada open-source řešení (Greenstone, OAI, FEDORA, ...)
- národní programy rozvoje DL (DLI-1,2, ...)

## Literatura, zdroje info



## 5.1 Literatura

### monografie

- Witten I.H.: **How to Build a Digital Library.** (Morgan Kaufmann Publishers, July 2002)
- Arms W.Y.: **Digital Libraries.** (MIT Press, 2000)
- Lesk M.: **Practical Digital Libraries.** (Morgan Kaufmann Publishers, 1997)

### články

- M.Bartošek: *Digitální knihovny*. In sborník konference DATAKON 2001 (též na <http://www.ics.muni.cz/mba/dl-datakon01.pdf>)
- povinná literatura zadaná v rámci jednotlivých přednášek

## 5.2 Další zdroje info (1)

### časopisy

- D-Lib <http://www.dlib.org>
- spec.číslo CACM 95/4, 98/4, 01/5
- RLG DigiNews <http://www.rlg.org/preserv/diginews/>
- ICSTI forum <http://www.icsti.org/forum/index.html>

### projekty na Inetu

- obrovské množství info (musím ale vědět, co hledám)

## 5.2 Další zdroje info (2)

### Konference (sborníky - viz LNCS-online)

- Advances in DLs (IEEE) 1996-2000
- ACM Conference on DLs 1996-2000
- Joint Conference on DLs od 2001
- European Conference on Research and Advanced Technologies for DLs od 1997

## 5.2 Další zdroje info (3)

### courseware

- Fox E.: DL - Virginia Tech Courseware
- Nelson M.L.: Intro to DLs, Uni. N.Carolina  
<http://buckets.dsi.internet2.edu/classes/inls110/>
- a mnohé další

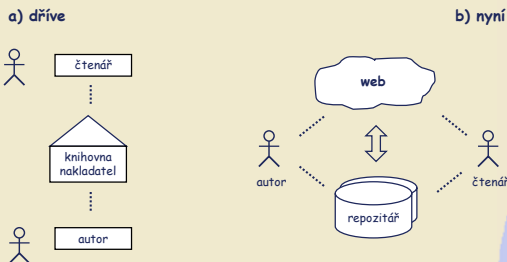
## DL a vědecké publikování



## 6. DL a oblast STI

- **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)

### 6.1 Vědecké publikování



### 6.2 Co je to STI ?

- soubor informačních materiálů (bez ohledu na jejich formu a formát) pro potřeby vědy, výzkumu a vývoje (zprávy, datasety, obrázky, videa, software, ...)
- zahrnuje:
  - **bílou literaturu** (knihy, časopisy, důležité sborníky)
    - autor a vydavatel jsou obvykle různé subjekty
    - obvyklé je posuzování kvality předcházející publikování
  - **šedou literaturu**: typ publikací nedostupných běžnými prodejními kanály, produkovány často v omezeném množství, s nízkou či žádnou inzercí a s omezeným využitím (disertace, výzkumné zprávy, články na Inetu)
    - el.publikování - velký impuls pro tento druh informací

### 6.3 Proč se starat o "Grey"

- **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“*

### 6.4 Pyramida STI

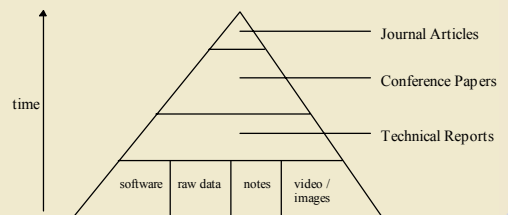


Figure 2: Pyramid of Publications Rests on Unpublished STI

## 6.5 Řešení dle disciplin

37

- **fyzika** preprinty (arXiv.org)
- **matematika** preprinty
- **computer science** technické zprávy  
sborníky konferencí
- **chemie** stále časopisy,  
ale zásadně *review*

## 6.6 Ekonomika STI

38

- 20.000 vědeckých časopisů celosvětově
- 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, 17.400 USD dnes)
- stagnující či klesající rozpočty knihoven
- pokud jeden zruší předplatné, cena pro ostatní stoupne
- 2017 - knihovny nebudou moci nic nakupovat ☹
- *časopisy - jen zlomek STI informací*

## 6.7 Jak z toho ven ?

39

- potřeba nových modelů pro šíření STI
- nové ekonomické modely

Šance a motivace i pro DL

## Literatura k přednášce



## Povinná literatura

41

- Vannevar Bush. *As We May Think*. Atlantic Monthly, July 1945  
<http://www.isg.sfu.ca/~duchier/misc/vbush/vbush-all.shtml>  
Bush je prvním průkopníkem v oblasti DL a to co se odehrává v této oblasti v dnešní době není v zásadě nic jiného, než praktická implementace vizí, které Bush vyslovil před více jak 50 lety
- Sandra Esler, Michael Nelson. *Evolution of Scientific and Technical Information Distribution*. JASIS 49(1), 1998.  
<http://techreports.larc.nasa.gov/ltrs/PDF/1998/jp/NASA-98-jasis-sle.pdf>  
Úvodní přehledový článek o digitálních knihovnách z oblasti šíření vědeckých a technických informací. Oblast STI bývá jednou z prvních experimentálních oblastí pro ověřování nových technologií a přístupů DL

## Povinná literatura

42

- Jaroslav Pokorný. *Digitální knihovny: principy a problémy*. In: Automatizace knihovnických procesů (8), duben 2001.  
<http://knihovny.cvut.cz/akp/clanky/O3.pdf>
- Miroslav Bartošek. *Digitální knihovny*. In Proc. Datakon 2001. Těž na <http://www.ics.muni.cz/mba/dl-datakon01.pdf>