



Teorie a praxe nákladů na digitální archivaci

Zdeněk Hruška

zdenek.hruska@mzk.cz

Moravská zemská knihovna v Brně

30. 11. 2017, AKM 2017

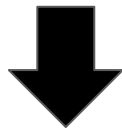
V knihovnách probíhá masivní digitalizace:

- ❑ Národní digitální knihovna (NDK): 45 000 000 stran
- ❑ Knihovna AV ČR (KNAV): 16 000 svazků;
- ❑ MěK Praha: 9 000 svazků;
- ❑ NTK: 3 000 svazků;
- ❑ Krajské digitalizační linky, univerzitní knihovny, městské knihovny, odborné knihovny, ...
- ❑ archivy, muzea, galerie, vědecké repozitáře, digital born data, ...

**“Digitální informace přetrvá věky -
nebo pět let, záleží, co přijde dřív.”**

Jeff Rothenberg

- ❑ HW + SW,
- ❑ el. energie,
- ❑ úložná média,
- ❑ zaměstnanci (know how),
- ❑ + rozvoj/obnova/výměna.



**Spotřeba zdrojů
(energie, čas, peníze)**



Náklady

Náklady:

„V penězích vyjádřené vynaložení ekonomických zdrojů uskutečněné za určitým předem stanoveným účelem.“

Zdroj: KRÁL, B. *Nákladové a manažerské účetnictví*. Str. 40.

Proč je nutné znát náklady?

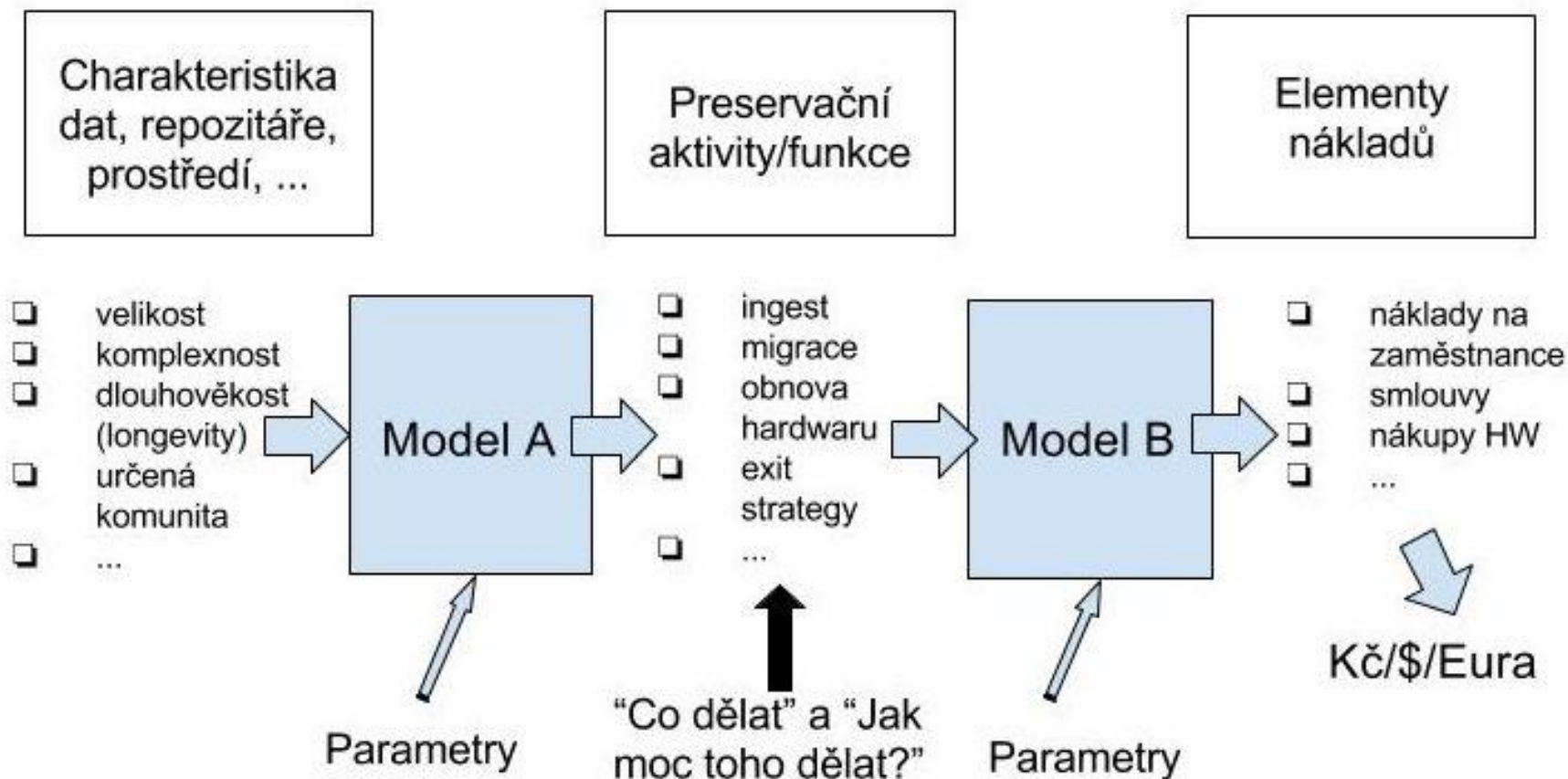
- ❑ Tvorba rozpočtů, dlouhodobé plánování a s tím spojené dlouhodobé financování;
- ❑ hodnocení alternativ a podpora rozhodování;
- ❑ advokacie, komunikace se stakeholdery;
- ❑ finanční plánování je indikátorem udržitelnosti a důvěryhodnosti repozitáře:
 - ❑ ČSN ISO 16 363: podkapitola 3.4
Ekonomická udržitelnost.

Modely pro DPC = Digital Preservation Costs

“Reprezentace zdrojů, jako jsou investice nebo práce, použitých v digitálním kurátorství. Jejich využití může být měřeno, aplikováním ceny na jejich použití mohou být tyto zdroje ohodnoceny peněžním vyjádřením a tím stanoveny jejich náklady.”

Zdroj: STOKES, Paul (ed.). *D3.1—Evaluation of Cost Models and Needs & Gaps Analysis*.

Obecný rámec DPC modelů



Zdroj: KAUR, Kirnn et. al. APARSEN. D32.1 Report on Cost Parameters for digital Repositories. 2013. Dostupné z: http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D32_1-01-1_0_incURN.pdf

- ❑ Chybí jednotící standard -> roztržitost.
- ❑ V počátcích vývoje DPC modelů malá spolupráce.
- ❑ Granularita modelů: obecný vs. podrobný:
 - ❑ KRDS vs. NASA CET
- ❑ Zdroje dat z případových studií:
 - ❑ metodika sběru dat,
 - ❑ ochota se o data dělit.

11 modelů hodnocených v projektech 4C a APARSEN:

4C Project: Collaboration to Clarify the Costs of Curation

STOKES, Paul (ed.). *D3.1—Evaluation of Cost Models and Needs & Gaps Analysis*. 4C Project, 2014. Dostupné z:

http://www.4cproject.eu/documents/D3.1_final_report_10May2014-v1.02.pdf

APARSEN Project:

KAUR, Kirnn et. al. APARSEN. *D32.1 Report on Cost Parameters for digital Repositories*. 2013. Dostupné z:

http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D32_1-01-1_0_incURN.pdf

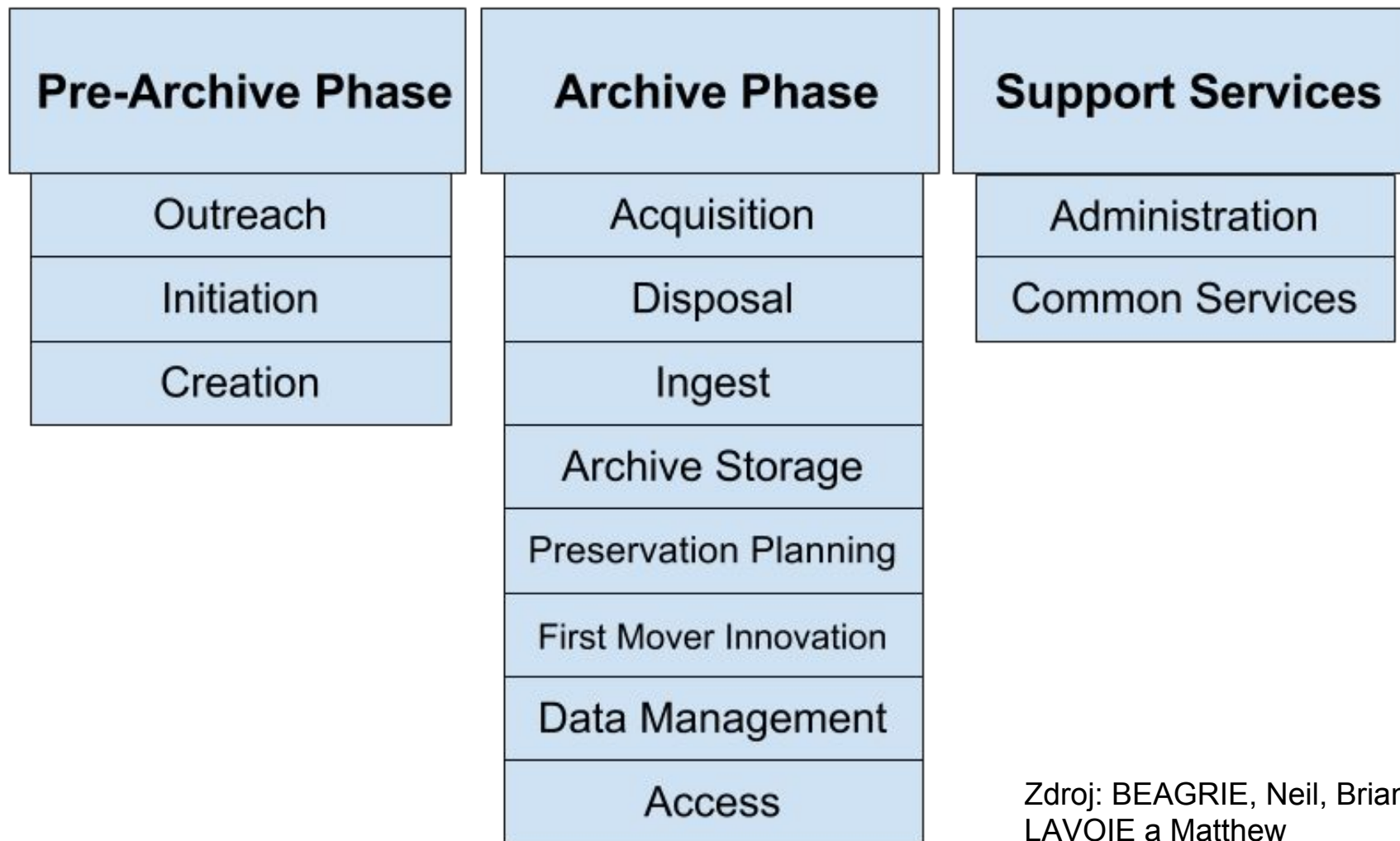
Aktivity a náklady dle modelu LIFE3



Pořízení	Akvizice	Deponování	Ochrana Bit-streamu	Ochrana obsahu	Přístup
Digitalizace	Výběr	Zajišťování kvality	Správa úložiště	Sledování	Poskytování přístupu
...	Dohoda s vydavatelem	Metadata	Zajištění úložiště	Plánování ochrany	Řízení přístupu
...	Práva duševního vlastnictví a licence	Uložení	Obnovování, renovační migrace	Ochranná činnost	Podpora uživatelů
...	Objednávání a fakturování	Update záznamů	Zálohování	Opětovné deponování	
	Získávání	Referenční odkazy	Inspekce	Odstranění	
	Kontrola				

Zdroj: VYCHODIL, Bedřich a Zuzana KVAŠOVÁ. *Projekt LIFE: Návrh metodiky pro sledování nákladů životního cyklu v rámci Národní knihovny v Praze za využití softwaru vyvinutého v rámci projektu LIFE*. Str. 8.

Aktivity a náklady dle modelu KRDS2



Zdroj: BEAGRIE, Neil, Briane LAVOIE a Matthew WOOLLARD. KEEPING RESEARCH DATA SAFE 2.

Uložení v cloudu?



Pokles cen - Amazon cloud:

- ❑ 2013 – 0.095 \$/GB/měsíc,
- ❑ 2016 – 0.03 \$/GB/měsíc,
- ❑ 2017 – 0.0245 \$/GB/měsíc pro prvních 50 TB

Další poplatky:

- ❑ 0.10 \$ za milion zkontrolovaných objektů za měsíc

Přenos dat:

- ❑ Do 10 TB v měsíci: 0.09 \$ za GB
- ❑ Dalších 40 TB v měsíci: 0.085 \$ za GB

- ❑ Ingest a uložení dat, zpřístupnění i archivace - včetně migrace
- ❑ 1 až 10 TB – 12 350 € (cca 315 000 Kč) za rok pro 1. TB v Amazon S3 a 1 580 € (cca 40 000 Kč) za TB/rok v Amazon S3, nebo 600 € (16 000 Kč) za TB/rok v Amazon Glacier.
- ❑ Více než 10 TB – bez uvedené ceny, až 50% sleva
30 TB/rok -> 1 - 2 mil. – dle úvazků v instituci

- ❑ DuraSpace - Archivematica + DuraCloud
- ❑ 1 TB – 9 999 \$ za rok (cca 215 000 Kč), možnost dokoupit další úložný prostor za 825 \$ za 1 TB/rok (cca 18 000 Kč).
- ❑ Při větším množství dat - individuální nacenění
30 TB/rok -> 1 - 2 mil. – dle úvazků v instituci

Digital Preservation Network (DPN)

- ❑ Členský poplatek 20 000 \$ (cca 430 000 Kč) + 5 TB
- ❑ Další data: 137,50 \$ (cca 3 000 Kč) za TB/rok.
- ❑ 3 kopie; 5 lokalit v USA
- ❑ Možnost zaplatit za uložení dat na dvacet let za 2 750 \$ (cca 60 000 Kč) za TB.

MetaArchive Cooperative

- ❑ Nový člen - 5 500 \$ (cca 118 000 Kč) + část úvazku na správce/digitálního archiváře
- ❑ Různé úrovně členství - možnost ovlivňovat chod konsorcia
- ❑ Členský poplatek + poplatek za HW = 5 000 \$ (cca 108 000 Kč)
- ❑ 585 \$ (cca 14 000 Kč) za TB/rok

Použité zdroje



- BEAGRIE, Neil, Briane LAVOIE a Matthew WOOLLARD. KEEPING RESEARCH DATA SAFE 2. In: [Http://www.webarchive.org.uk](http://www.webarchive.org.uk) [online]. 2010 [cit. 2015-05-02]. Dostupné z: <http://www.webarchive.org.uk/wayback/archive/20140615221405/http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>
- HRUŠKA, Zdeněk. *Náklady na dlouhodobou ochranu digitálních dokumentů v českém prostředí na příkladu Moravské zemské knihovny*. Brno, 2017. Magisterská diplomová práce. Filozofická Fakulta Masarykovy univerzity. Vedoucí práce Petr Žabička. Dostupné z: https://is.muni.cz/th/217895/ff_m/
- KAUR, Kirnn et. al. APARSEN. *D32.1 Report on Cost Parameters for digital Repositories*. 2013. Dostupné z: http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D32_1-01-1_0_incURN.pdf
- STOKES, Paul (ed.). *D3.1—Evaluation of Cost Models and Needs & Gaps Analysis*. 4C Project, 2014. Dostupné z: http://www.4cproject.eu/documents/D3.1_final_report_10May2014-v1.02.pdf
- THE BRITISH LIBRARY. *Life3_ver50*. [List Microsoft Excel 97–2003]. Ver. 50. London, 2009. Dostupné z: http://www.life.ac.uk/3/docs/life3_ver50.xls
- VYCHODIL Bedřich a Zuzana KVAŠOVÁ. *Projekt LIFE: Návrh metodiky pro sledování nákladů životního cyklu v rámci Národní knihovny v Praze za využití softwaru vyvinutého v rámci projektu LIFE*.

Použité zdroje



Subscription plans. *ArchivesDirect* [online]. [cit. 2017-06-10]. Dostupné z: <http://www.archivesdirect.org/pricing>

Summary of Cost Models. *4C Project* [online]. c2013 [cit. 2016-11-22]. Dostupné z: <http://4cproject.eu/summary-of-cost-models>

Curation Costs Exchange [online]. 2014 [cit. 2017-06-29]. Dostupné z: <http://www.curationexchange.org/>

How Preservica Works. *Preservica* [online]. c2016 [cit. 2017-06-10]. Dostupné z: <http://preservica.com/preservica-works/>

Cloud Edition Subscription Plans. *Preservica* [online]. c2016 [cit. 2017-06-10]. Dostupné z: <http://preservica.com/pricing/eu-eeur/>

Amazon S3 Pricing. *Amazon Simple Storage Service (S3)* [online]. c2017 [cit. 2017-05-29]. Dostupné z: <https://aws.amazon.com/s3/pricing/>

About DPN. *The Digital Preservation Network* [online]. c2017 [cit. 2017-06-15]. Dostupné z: <http://dpn.org/about>

Costs. *MetaArchive* [online]. c2017 [cit. 2017-06-10]. Dostupné z: <https://www.metaarchive.org/costs>

Děkuji za pozornost!

Zdeněk Hruška

zdenek.hruska@mzk.cz

Moravská zemská knihovna v Brně

www.mzk.cz