

Digitálna knižnica a digitálny archív

prof. PhDr. Dušan Katuščák, PhD.

Osnova prednášky

1. Prerekvizity a osobný program
2. Metodologický a technologický koncept
3. Projekt Digitálna knižnica a digitálny archív a môj zápas s veternými mlynmi
4. Porovnanie plánu a skutočnosti
5. Konzervovanie a CBA (Cost Benefit Analysis)
6. Závery

Prerekvizity digitálneho príbehu a našej cesty 60.-70. roky

- 1968 – na začiatku – Matica slovenská, Miloš Kovačka, Dezider Kormúth, Peter Liba a článková bibliografia (1918-1938)
- Bibliografická práca – záujem o lepšie zaznamenanie obsahu popri popise (obsahová charakteristika...)
- Strojové spracovanie slovenskej národnej bibliografie (1975) (ing. Ján Mišík, dr. Štefan Ďurovčík)
- Formovanie metodologického mikrosystému –
- Štrukturalizmus, Nitra, Kabinet literárnej komunikácie, teória metakomunikácie, metatextu, komunikačný prístup k bibliografii, dôraz na informačnú funkciu bibliografie a knižníc, štátne výskumné programy P13 a P18

Prerekvizity 80. roky

- IKIS – pokusy o integráciu a kooperáciu (štátny program) – Dušan Katuščák, Štefan Kimlička, Ján Mišík, Miloš Kovačka, Soňa Makulová, Jela Šteinerová, Jaro Šušol, Štefan Ďurovčík a i.
- 1978 - metodológia informačných systémov, informačných jazykov, bibliografie, štandardizácie bibliografického popisu, formát ISO 2709, makrotezaurus;
- 1983 – SNK ako samostatný útvar v rámci Matice slovenskej
 - Regionálny informačný kooperačný systém
 - Zrod modernej koncepcie konzervovania dokumentov
 - Vybudovanie datacentra, získanie hardvéru, výmenný formát IKIS
- 2000, zákon 183/2000 Z.z. Zákon o knižniciach

Prerekvizity 90.roky

- Štandardizácia, integrácia, kooperácia
- Preklad a implementácia formát UNIMARC, Cena Mateja Bela (Lit. Fond)
- Implementácia štandardov ISBD a AACR2
- Projekt KIS3G a spoločný softvér VIRTUA (cena ITAPA, Európska komisia – jediný významný portál SK – manažment znalostí
- Záznamy vo formáte MARC21 implementované do Virtui, konverzie z lokálnych systémov (ca 1.5 mil.)
- Št. úloha – KNIHA SK, projekt základného výskumu – konzervovanie papiera (špičkové vedecké výsledky)
- Odlúčenie SNK od Matice slovenskej
- Univerzita Komenského (spoločný systém Virtua – 12 fakúlt, konverzie)

Rok 2000+

2000 Zákon 183/2000 Z.z. O knižniciach

Výberové konanie: Generálny riaditeľ (2000-2012)

Priority:

- Masová priemyselná digitalizácia (od 2002)

http://www.4digitalbooks.com/auto_dlmini.php

- Masové priemyselné konzervovanie

<http://www.knihask.eu/odborny.htm>

- 2006 – SNK v Európskej knižnici

- 2009 – SNK v Europeane

- **Projekt OPIS2** (180 mil eur – 2005-2015)

- Slovenská digitálna knižnica , OPIS2 - (Nórske fondy)

- Národný projekt: **Digitálna knižnica a digitálny archív (DIKDA)**

- Katedra mediamatiky a kultúrneho dedičstva 2002/2003 (prof. Milan Konvit),

- Centrum excelentnosti UNIZA (2010-2015)

Metodologický koncept bibliografie a dokumentácie

• Writing  Rewriting  Description  Digitisation

- Historickú vývojovú líniu vidíme takto:
- Bibliografia sa „vracia“ po dvoch tisícročiach ku svojej archetypálnej „podstate“
- Prešla od starovekého a stredovekého *písania* a *prepisovania* dokumentov (writing, rewriting) do 16. storočia,
- cez novoveké *popisovanie* kníh (description), ktoré teraz doznieva, po *digitalizáciu* (digitisation) od začiatku 21. storočia.
- To, čo bolo v minulosti v bibliografii dominantné a známe ako *prepisovanie* dokumentov, to je v súčasnosti pre bibliografiu *digitalizácia* ako **kvalitatívne nová fáza vývoja jej formy a spoločenského dosahu**

Globálna schéma (J. Dzvák)

Zbierky/objekty na policiach: 9.3 km (BIG Collections)

Prezentácia verejnosti (Big Data) Nová služba

ebooks



Čistenie v depozitoch

Textové analýzy (JULŠ,
IBM, HP, UPJŠ, Oracle...)



Konzervovanie

Sterilizácia
80 ton za rok

Deacidifikácia
20-60 ton za rok

Dolovanie znalostí
Spracovanie dát
Metadáta

KB/data Mining

MDS/DO associations



Pracovné úložisko



Skenovanie

Laboratórna
kontrola kvality

Projekt Digitálna knižnica a digitálny archív a Môj zápas s veternými mlynmi



Schválenie národného projektu

- **OP Informatizácia spoločnosti (OPIS)**
- **Schválenie 2006-2007 (RO, EK, vláda SR)**
- **Január 2010 OPIS 2 – Schválená Štúdia uskutočniteľnosti**
- **Vláda SR – uzn.192/2011 (24.7 mil. pro rata)**
- **31.3. 2011 Zoznam národných projektov OPIS na roky 2007 – 2013**
- **28.3.2011 – minister D. Krajcer - poverenie SNK pripraviť a implementovať projekt DIKDA (2011-2015)**
- **2010-2015 - Centrum excelentnosti FHV Žilina-SNK (4 mil.eur)**
- **16.11.2011 – hotový projekt – odovzdaný SORO**
- **2012 – politika zvíťazila nad verejným záujmom**



Národné projekty OPIS 2 Alokácia finančných prostriedkov

Názov projektu	Žiadateľ	Finančná alokácia €
Digitálna knižnica a digitálny archív	Slovenská národná knižnica	49 667 500,00 €
Digitálna galéria	Slovenská národná galéria	16 009 000,00 €
Digitálne múzeum	Múzeum SNP	27 612 000,00 €
Digitálny pamiatkový fond	Pamiatkový úrad SR	10 600 000,00 €
Digitálna audiovizia	Slovenský filmový ústav	24 200 000,00 €
Centrálna aplikačná infraštruktúra a registratúra	Národné osvetové centrum	23 797 000,00 €
Centrálny dátový archív	Univerzitná knižnica v Bratislave	33 193 918,00 €
Harmonizácia informačných systémov	Národné osvetové centrum	8 868 256,00 €
Dokumentačno-informačné centrum rómskej kultúry	Štátna vedecká knižnica v Prešove	1 924 000,00 €

Cieľ projektu DIKDA

- ▶ Digitalizácia a sprístupnenie písomného kultúrneho a vedeckého dedičstva všetkým občanom Slovenska, Európy a sveta
- ▶ Špecifický cieľ - *digitalizácia všetkých slovacikálnych dokumentov*
- ▶ *Slovacikálne písomné dokumenty, čiže diela, ktorých*
 - a) *autormi sú Slováci,*
 - b) *boli vydané, vyrobené, nájdené a nachádzajúce sa na území dnešného Slovenska,*
 - c) *sa obsahovo dotýkajú Slovenska alebo Slovákov alebo*
 - d) *sú napísané v slovenskom jazyku*
- ▶ Sociálny cieľ - poskytnúť všetkým občanom *nové služby*
- ▶ PFI už nebudú poskytovať len *záznamy* o dokumentoch ale *úplné texty samotných dokumentov*

Slovenská národná knižnica	Položka	počet objektov	str.na objekt	počet strán
	Knihy z depozitu SNK	500 000	350	175 000 000
	Historické dokumenty, manuskripty, staré a vzácne tlače, vybrané archívne pomôcky, textové prejavy tradičnej ľudovej kultúry a pod.	150 000	200	30 000 000
	Nové knihy (nové ročné prírastky slovacikálnych kníh SNK)	50 000	350	17 500 000
	Zborníky, Acta, zborníky z vedeckých a odborných konferencií atd.	10 000	350	3 500 000
	Zborníky, Acta, zborníky z vedeckých a odborných konferencií atd. 2010-2015	250	350	87 500
	Štúdie, vedecké a odborné state a články v zborníkoch	700 000	0	0
	Seriály(slovacikálne)	13 500		20 250 000
	Nové ročné prírastky titulov seriálov do SNK 2010-2015	5 000		7 500 000
	Viazané rukopisy, archívne pomôcky, plagáty,	10 250		10 250
Spolu	1 439 000		253 847 750	

Digitalizačné roboty SNK DL3000



17. 5. 2018

Brno, Masarykova univerzita, Filozofická fakulta,
Blok expertov

16



Kamaldulský kódex (1756)

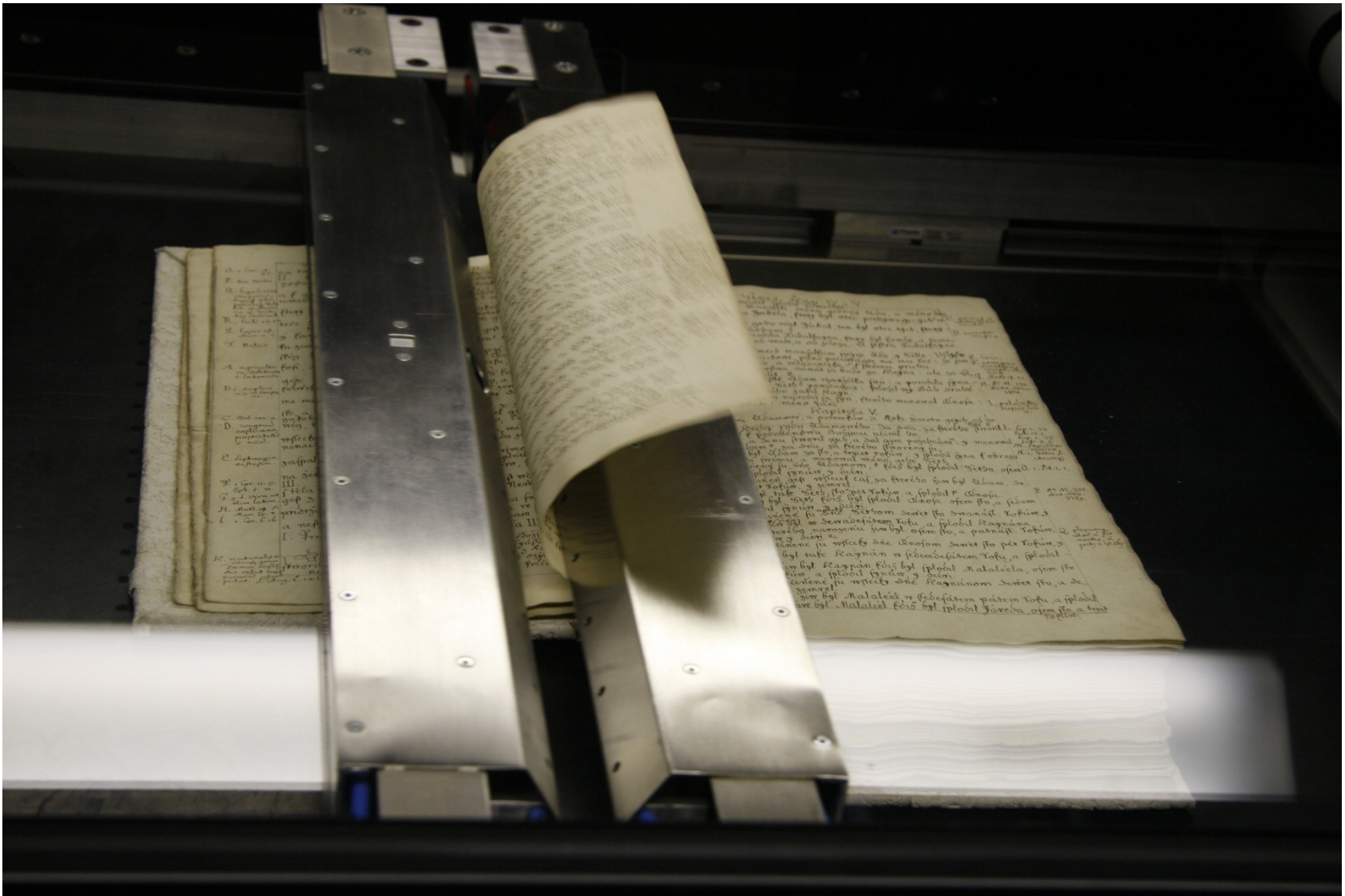
- Rukopis začal vznikať opisovaním Nového zákona 6. mája 1756; opisovanie Starého zákona začalo 4. marca 1757.
- Rukopis má dva zväzky. Prvý zväzok má 920 strán a druhý 780 strán.
- Súčasťou prekladu bol aj Latinsko-slovenský slovník, ktorý sa zachoval v rukopise z roku 1763.
- Romulad Hadbavný, Kamaldulský rád, Červený kláštor

https://www.researchgate.net/publication/313350851_Kamaldulsky_kodex -
[symbolicky zaciatok masovej digitalizacie na Slovensku Narodny projekt digitalna kniznica a digitalny archiv](#)

(Chemical Papers, Krásy Slovenska, Matriky, ... Ca 2 mil. strán experimentov)







Sprístupnenie objektov - repozity

Europeana (51 mil objektov)

<https://www.europeana.eu/portal/sk>

Slovakiana

<https://www.slovakiana.sk>

Web umenia

<https://www.webumenia.sk>

MEDIAINFO (digi zbierky)

<http://mi.memoria.sk/#>

FEZ/FEDORA (open source) (digi zbierky)

<http://fez.ceps.uniza.sk/view/FEZ:106>

Virtua (KIS3G_LIS - knižničný systém)

<http://www.memoria.sk>

<https://www.kis3g.sk>

(Hydra, Islandora, Invenio, Dspace, Eprint ...)

Projekt v číslach

- Rozpočet projektu: 50 000 000 Eur
- 2 800 000 objektov => 250 000 000 strán
- 250 000 000 strán => 17 PB dát
- 2 pracovné kópie potrebujú asi 34 PB dát
- Produkcia 43 TB/deň
- Preniesť 6 Gbit dát za sekundu
- Trojsmenná prevádzka jún 2012-jún 2015
- 78 nových pracovných miest
- 30 % dostupných objektov cez internet
- Udržateľnosť minimálne do roku 2020
 - Ca 24 000 000 CD (700 MB)
 - Ca 3 600 000 DVD (4,7 GB)
 - Ca 9 000 pásov LTO 5 (1,6 TB)

Miesto a spôsob realizácie digitalizácie

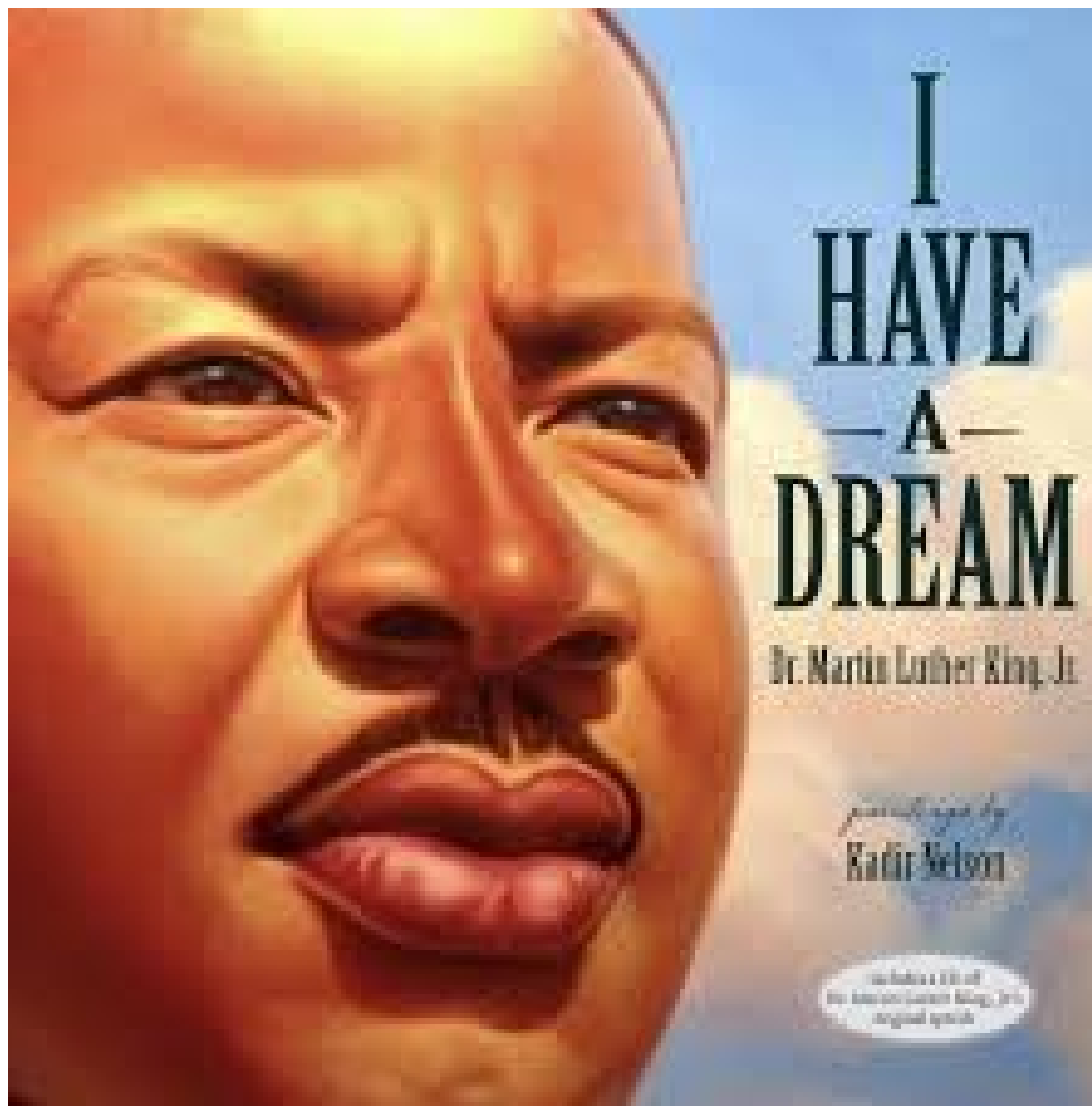
- ▶ *Masová priemyselná digitalizácia*
- ▶ 80% digitalizácie sa uskutočnili s použitím *robotov*
- ▶ 20 % sa uskutočnili s použitím manuálnych skenerov
- ▶ *Integrované konzervačné a digitalizačné centrum*
- ▶ Digitalizácia a konzervovanie ako *jeden sofistikovaný technologický proces*
- ▶ *Kritický masív slovacikálnych objektov sa nachádza v Slovenskej národnej knižnici v Martine a v Slovenskom národnom archíve v Bratislave*
- ▶ *Dočasne 213 nových zamestnancov (trvalo zostáva 76)*

Martin Luther King 1963

Mali sme sen :-))

Aj sny a projekty sa dajú
ukradnúť a deformovať





Porovnanie plánu a skutočnosti

<http://www.itapa.sk/data/att/3467.pdf>

(plagiát J. Kováčik SNK, referuje o projekte bez uvedenia jeho autorstva)

Podľa informácie Týždenníka Trend

„Celkový počet strán, ktoré sa mali zdigitalizovať v národnej knižnici, tým klesol z pôvodných 253,8 milióna na 40,3 milióna, čiže šesťnásobne. Jednak sa znížil počet objektov a tiež sa uprednostňujú písomnosti menej náročné na spracovanie na úkor náročnejších kníh. Napríklad články v novinách, príspevky v zborníkoch a podobne. Podľa pôvodnej zmluvy o nenávratnom finančnom príspevku mala knižnica zdigitalizovať celkovo 700-tisíc kníh. Tento cieľ sa znížil na 100-tisíc monografií.“

05.10.2015

Náklady na digitalizáciu (nie len skenovanie)

- infraštruktúru,
- investície - stavby, materiál,
- IKT,
- skenery,
- softvér,
- archivovanie,
- prístup
- hardvér,
- laboratórne vybavenie,
- konzervovanie
- energie,
- mzdy,
- odvody,
- poistné,
- zamestnanci,
- služby,
- slovom **všetko!**

Zdroj podvodu (podľa môjho názoru)

- [Správa](#) Úradu pre verejné obstarávanie hovorí:

*"Pre každú každú položku aj pre každú kategóriu použil paretovu analýzu na stanovenie koeficienta váhovosti na ocenenie služieb a súvisiacich tovarov potrebných pre implementáciu národných projektov, definovaných v OPIS 2, t.j. **určil váhu ceny pre každú kategóriu tak, aby dosiahol čo najefektívnejšie a najhospodárnejšie vynakladanie finančných prostriedkov za daný predmet zákazky.**"*

V skutočnosti Ministerstvo kultúry a Úrad pre verejné obstarávanie nevykonali žiadny cenový prieskum. Len akceptovali návrhy cien budúceho dodávateľa v rámcovej zmluve.

Plán a skutočnosť - porovnanie

Objekty (zmenený plán SNK) Strany	A (plán) Projekt DIKDA/obje kty	B(zmena) Objekty/p očet zmena SNK	(C) Rozdiel oproti projektu/objekty
1.Monografie a zväzky viacdielných monografií zväzkov	700 000	100 000	-600 000 kníh
2. Seriály (čísla) výtlačkov	28 000	350 000	+322 000
3. Články a príspevky v zborníkoch článkov	0	800 000	+800 000
4. Mapy, plagáty, pohľadnice kusov	10 250	10 000	+250
Počet strán	253 847 750	40 300 000	-213 547 750
Celkom objektov	1 439 000	1 260 000	-179 000

Počítajme ďalej

- Podľa koncepcie MKSR nemá SNK digitálny obsah dlhodobo **archivovať** sama, ale prostredníctvom **Centrálneho dátového archívu**, ktorý je vybudovaný v rámci projektu OPIS2 v Univerzitetnej knižnici v Bratislave.
- Tento archív má kapacitu **25 000 Terabytov** (=25 PB). Náklady na archív v UKB boli ca **33 000 000 €**.
- Z toho je pre digitálny obsah SNK plánovaná a obstaraná kapacita **17 000 TB** (=17 PB)
- Celkový počet TB v apríli 2018 je **4162,21 TB (zo 17 inštitúcií)**

Centrálny dátový archív Univerzitná knižnica v Bratislave

Názov zmluvy: **Zmluva o poskytnutí nenávratného finančného príspevku**

- Číslo zmluvy: **MK-40/2012-M**
- Dodávateľ: **Univerzitná knižnica v Bratislave**
Michalská 1, 814 17 Bratislava
Slovenská republika
IČO: 00164631
- Odberateľ: **Úrad vlády Slovenskej republiky**
Námestie slobody 1, 813 70 Bratislava 1
Slovenská republika
IČO: 00151513
- Celková hodnota predmetu zmluvy, ak ju možno určiť (vrátane DPH): **33 184 470,48 €**
- Dátum uzatvorenia zmluvy: **13. marec 2012**
- Kapacita CDA (archív) **25 PB (25 000 TB)**

Počítajmeďalej!

Archivovanie digitálneho obsahu

- Od roku 2012 do konca projektu v roku 2015, SNK do archívu nedala **17 000 terabytov**, ale len približne 80 **terabytov**, čo nie je ani len jedno percento plánovaného objemu.
- Tých **17 PB** v archíve UKB, ktoré sú určené pre SNK, predstavuje **67%** nákladov na archivovanie digitálneho obsahu z projektu SNK.
- Znamená to, že sumu **40 282 685 €**, ktorú použila na digitalizáciu SNK, **treba zvýšiť o 67% zo sumy 33 000 000 €.** (67% z 33 000 000 je **22 110 000 €.**)
- **Suma na projekt SNK je takto zvýšená na hodnotu 62 392 685 €.**

Počítajme ešte ďalej!

Sprístupnenie digitálneho obsahu

- **K sume 62 392 685 €** je potrebné pripočítať alikvotnú čiastku aj z projektu Národného osvetového centra, ktoré malo od konca novembra 2015 sprístupňovať digitálny obsah cez systém **Slovakiana**
- **Slovakiana** stála (projekt OPIS2 - Centrálna aplikačná infraštruktúra) ca **23 000 000 €**.
- **Väčšina digitálneho obsahu mala byť zo SNK.**
- Zo sumy 23 000 000 je preto vhodné dedikovať na sprístupnenie digitálneho obsahu **odhadom ca 15 000 000 €**
- **Vývoj nového softvéru nebol potrebný!!!**

Náklady na DIKDA spolu

- Projekt SNK: **40 282 685 € (-9 mil)**
- Archív CDA v UKB pre SNK: **22 110 000 € (67% kapacity)**
- Sprístupnenie v Slovakiane: **15 000 000 € (odhad alikv. časť)**
- **Digitalizácia písomného dedičstva (DIKDA)**
- **Výsledok 40 300 000 strán (z plánovaných 250 000 000 strán)**

Σ 77 392 685 €

Pri kurze 25 CZK za euro (ca 1,5 miliardy CZK)

Ca 2 € za stranu (plánované 0,19 €/strana)

Konzervovanie (nerealizované podľa projektu)

Ciele IKDC:

Projekt KNIHASK (štátny plán základného výskumu SAV, STU, SNK, SNA): <http://kniha.chtf.stuba.sk>

- Technologický rozvoj najlepších dostupných konzervačných technológií (BAT) na Slovensku, alebo zmluvne potvrdené zabezpečenie zapojenia Slovenska do ich rozvoja (Nitrochemie)
- Udržateľný rozvoj technológie a zamestnanosti vysokoškolsky kvalifikovaných ľudí na Slovensku.
- Znižovanie merných nákladov na ochranu kultúrneho dedičstva
- Zvyšovanie úžitkových vlastností KAD
- Vývoj a optimalizácia technologických postupov ochrany KAD so zreteľom na udržanie ekologickej kvality životného prostredia a pracovného prostredia

Technologické centrum KNIHASK



UniLab S



Navrhované procesy a výkony prevádzky IKDC

Proces	Spracované množstvo KAD (kg/r)
Deacidifikácia	20 000
Digitalizácia	50 000
Sterilizácia	80 000
Čistenie	80 000
Sušenie	20 000
Individuálne úpravy	2 000
Lyofilizácia	bez určenia množstva

CBA=Cost Benefit Analysis

- Citát z CBA, povinnej prílohy projektu DIKDA: „Pri použití tejto metódy **pre potreby SNK** sa vychádza z nasledujúcich údajov (prepočítané na euro kurzom 1 EUR=1,3 USD):
- Počet dokumentov v knižniciach SR je **3 500 000** (od roku 1900 po rok 2010).
- Cena jedného dokumentu je odhadnutá na 76,9 EUR.
- Celková cena dokumentov je 269 230 769 EUR.
- Ročné straty sú 4,66% [ročná degradácia], čo vo finančnom vyjadrení predstavuje **12 546 154 EUR.**“

CBA v SNK (priemyslové konzervovanie sa nekoná)

- V rámci projektu sa malo chemicky ošetrovať **cca 500 000** knižničných dokumentov a **1,4 mil. dokumentov z národného archívu**, čo počas troch rokov realizácie predstavuje **1 900 000 kusov**.
- Touto metódou ročne zabráni stratám 4,66% z počtu 633 333 dokumentov.
- Cena jedného dokumentu je odhadnutá na **76,90 EUR** (LC 100 USD)
- Celková cena dokumentov je **48 703 308 EUR**.
- Ročné straty sú 4,66%, čo vo finančnom vyjadrení predstavuje **2 269 574 EUR**.
- Prínos zo záchrany dokumentov je trvalý a tak do prínosov projektu je možné započítať každoročnú záchranu opotrebovania knižničného fondu o 4,66%.
- Návratnosť nákladov na projekt 10 rokov podľa CBA



Závery

Okrem osobnej frustrácie aj profesionálne závery:

1. Do vedy, výskumu a vzdelávania patria aj poznatky o negatívnych javoch a skúsenostiach!
2. Nevyužitá historická príležitosť kvalitatívnej zmeny služieb informačných inštitúcií (fragmentovaný prístup, divoká digitalizácia...)
3. Plytvanie verejnými zdrojmi
4. Zlyhanie európskych a národných kontrolných orgánov a ÚVO
5. Zneužitie odborníkov politikmi a politizácia programu informatizácie
6. Nekompetentné riadenie na úrovni ministerstva a inštitúcie
7. Nedostatočný záujem odbornej informačnej komunity o zásadnú inováciu prostredníctvom masovej priemyselnej digitalizácie a konzervovania
8. Komunita na Slovensku takmer nereaguje na inovácie a výzvy (Big data, Digital Humanities, scientizácia LIS, READ...)

dusankatuscak@gmail.com

Ďakujem za pozornost