



# Národní digitální knihovna

Ing. Luděk Rašek

© CGI Group Inc. CONFIDENTIAL

**CGI**

Experience the commitment®

## Náplň projektu

- Masová digitalizace knižních fondů
- Vybudování dlouhodobého úložiště
- Zpřístupnění digitalizovaných dat

## Účastníci projektu

- Národní knihovna ČR
- Moravská zemská knihovna Brno

#### Předprojektové aktivy

- Sběr zkušeností
- Studie proveditelnosti
- Návrh standardů

#### Realizace

- Zahájení projektu – 10/2011
- Akceptace prováděcího projektu – 1/2012
- Dokončení realizace dodávek – 4/2012

#### Testování

- Zahájení testování – 7/2012
- Akceptace řešení – 9/2012
- Pilotní provoz – 10/2012

#### Provoz

- 2013 - 2014
- Digitalizace
- Optimalizace

## Cíle masové digitalizace

- 26 000 000 stránek z fondů NKP a MZK
- cca 54 000 stran denně

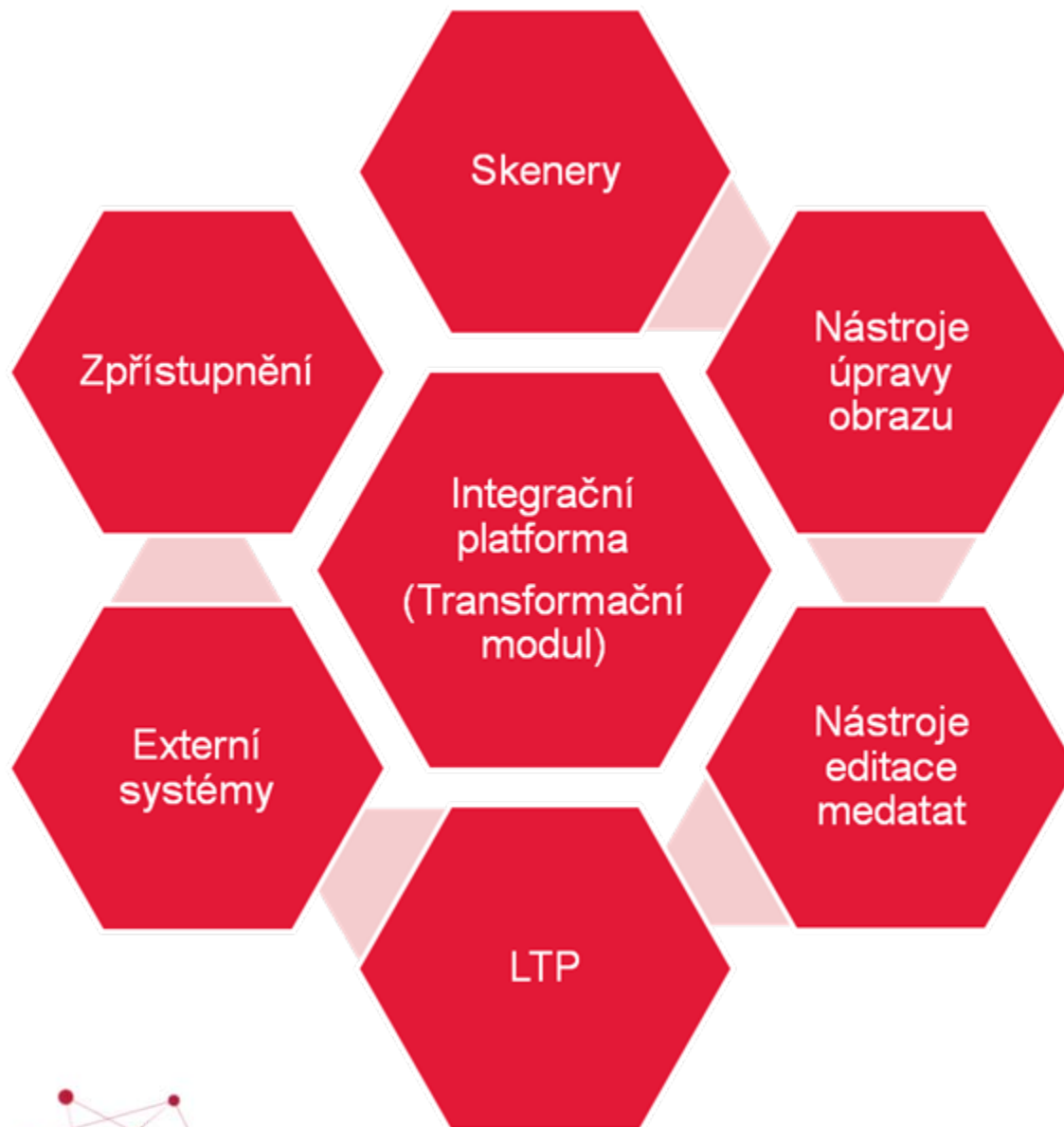
## Cíle dlouhodobé archivace

- Uložit 26 000 000 zdigitalizovaných stran
- Sklizeň projektu Webarchiv
- Data z minulých digitalizací a import externích zdrojů – 20 000 stran denně

## Cíle zpřístupnění

- Zdigitalizovaná data zpřístupnit veřejnosti

# Projekt systémové integrace



operátor



manager



Digitalizační workflow

Řídicí a dohledové centrum

admin



Transformační modul

Pracovní  
prostor

Externí  
systémy

- Aleph
- Registr digitalizace
- URN:NBN resolver

- LTP
- NKP K4
- MZK K4
- Zpřístupnění

# Přístup k integraci a návrhu systému

## Podpora víceúrovňového řízení

- Procesní integrace pomocí digitalizačního workflow
- Technologická integrace
- Podklady pro management projektu

## Volná vazba komponent

- Integrace pomocí definovaného rozhraní
- Možnost úpravy funkcionality bez dopadu na okolí
- Možnost využití více různých implementací

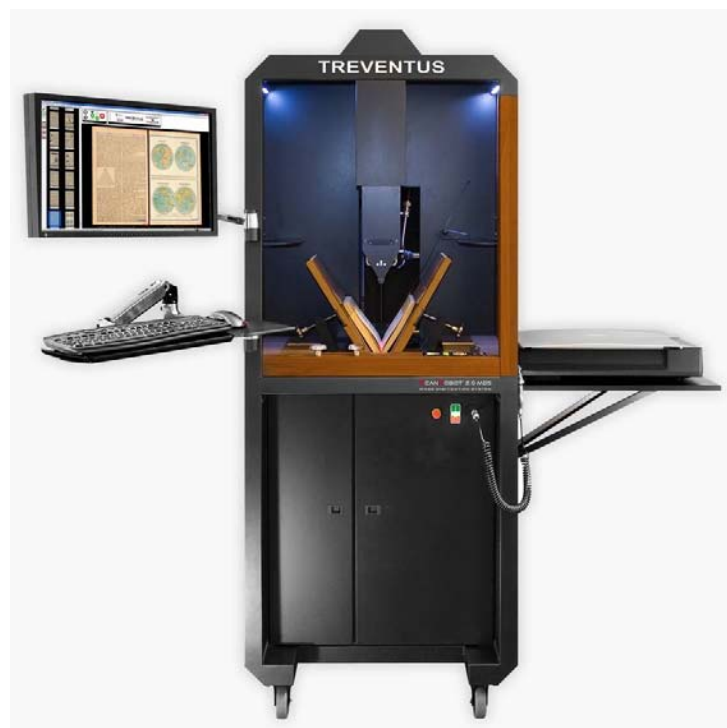
## Vysoká míra škálovatelnosti

- Distribuované zpracování
- Asynchronní zpracování požadavků
- Transparentní přidávání výpočetního výkonu

## Flexibilita při zpracování

- Vysoce integrované centralizované datové centrum
- Přístup nezávislý na lokalitě
- Přístup nezávislý na vybavení prostředí pracovníka

# Robotické a manuální skenery





# Využívané technologie

## Digitalizace

- **AIP Safe** – řídicí workflow (implementace)
- **LSA** – univerzální integrační utilita pro skenery (vývoj)
- **Scantailor** – nástroj pro postprocessing (opensource+úpravy)
- **Editační modul** – nástroj pro tvorbu a úpravu metadat (vývoj)
- **ABBYY Recognition Server** – OCR engine (produkt)
- **Integrované systémy** – Aleph, Registr digitalizace, URN:NBN resolver, Kramerius 4

## Transformační modul

- **JBoss JBPM** – řízení transformačních procesů (opensource)
- **ActiveMQ** – komunikační platforma (opensource)
- **Mule ESB** – integrační platforma (opensource)
- **Vývoj** – konkrétní výkonné moduly (vývoj)
- **Integrované nástroje** – Kakadu, JHOVE, ... (různý původ)

## LTP - dlouhodobé úložiště

- **AIP Safe LTP** – řízení OAIS procesů
- **Transformační modul** – separátní instance pro zajištění procesů v LTP
- **IBM Information Archive** – dlouhodobé datové úložiště

## Zpřístupnění

- **Vufind** – uživatelské rozhraní (opensource)
- **SOLR** – full text index (opensource)
- **Harvester** - těžení datových zdrojů (opensource+vývoj)

Řídicí a dohledové centrum (MS BI)

## Zkušenosti integrace

- Nezralé prostředí digitálních knihoven
- Není přesně vymezena role jednotlivých systémů
- Nerovnoměrná dostupnost integrovaných systémů
- Nedostatečné sladění životního cyklu systémů

## Zkušenosti digitalizace

- Standardy kvality na pracovištích musí odpovídat cílům projektu
- Vyšší náročnost na pracovišti zpracování obrazu a metadat, než na pracovišti skenování
- Zaškolení nových pracovníků je rychlé a brzy podávají dostatečný výkon
- Velký dopad mají procesy s fyzickou předlohou (prostory, mezisklad, příprava předloh)

## Zkušenosti LTP

- Připravené standardy byly příliš striktní
- Nutno přizpůsobit při zachování vysoké míry homogenity úložiště

## Rozvoj digitalizace

- Zpřístupnění technologie jako služby

## Rozvoj LTP

- Zveřejnění validátoru formátu NK 2012 (ke stažení zdarma)
- Centrální dlouhodobé úložiště pro knihovnické subjekty
- Příjem dat z krajských digitalizací

Děkuji za pozornost



**CGI**

Experience the commitment®