

# Co znamená implementace DL4DH pro knihovny a knihovníky

AV ČR, Praha 7.12. 2022 (pst. ed.)

Mgr. Zdenko Vozár

National Library of the Czech Republic

# Úvod

- Zámer a výsledky riešenia DL4DH
- Akvizícia dát: Systém digitalizácie a štandardizácie v ČR
- Kurátor a metodik: Správa zbierok a kontrola kvality
- Analytik a bádateľ: Spracovanie a využitie dát
- Operativa a vývoj: Infraštruktúra a integrácia spracovania dát

# Zámery a riešenia DL4DH

Základné princípy

# Zámery

- 1. Sprístupniť dáta z digitalizácie textov vedcom a študentom pomocou postupov no-code / low-code
- 2. Zaviesť pokročilé kustomizovateľné obohacovanie dát
- 3. Implementovať variantné dátové formáty (a správu úložiska obohatených textov)
- 4. Voliteľný export dát
- 5. Správa relevantných datasetov so štandardizovanou kvalitou

- **1. Vytvorenie postupov a rozhraní**
  - Metodika: [Certifikovaná metodiky DL4DH](#)
  - Feeder: Dedikovaný GUI a REST API pre vedcov a datových špecialistov nad DK
  - Kramerius+: Administrácia BE procesov
  
- **2.-5. Stabilné a kustomizovateľné riešenie**
  - riadenie obohacovania a ukladania dát v BE [Kramerius+](#) , včítane indexovania v FE [Feeder](#)
  - voliteľné formátovane prostredníctvom interného formátu XML a konverzného mechanizmu [TEI converter](#)
  - rýchle exporthy textových výstupov a obohatených dát
  - [POC DL4DH vytvorenia a správy](#) odborných doménových kolekcí a ich vyťaženie



# Riešenie exportu dát

- Prostredie: GUI / REST API
- Princíp:
  - I. Agregácia
    - ad hoc
    - z doménových kurátorských kolekcí
    - z užívateľských datových sád
  - II. Prieskum výberu a ďalšia filtrácia
  - III. Iteratívne úpravy datasetu

Typ úlohy

- Všetchny
- Export JSON
- Export TEI
- Export CSV
- Export ALTO
- Export TEXT

# Akvizícia dát

System spracovania zdrojov pre DK

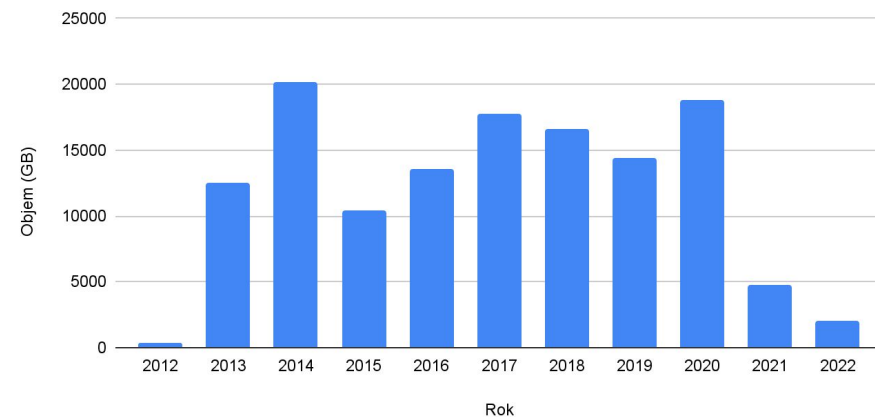
# Produkcia dát NDK.cz (typ UserCopy)

Total: 131 600 GB

Med: 13 598 GB /per y.

Max: 20 000 GB /per y.

NDK: Objem (GB) per rok činnosti







# Implementácia riešenia

System spracovania zdrojov

## Aktuálna testovacia inštancia

- CPU: 8vCPU
- RAM: 16 GB
- HDD: 50 GB sys; 200 GB data
- System: Ubuntu 20.4 LTS
- IE dimm:
  - 5 MB ~ 30 MB, 160x
- Aktuální data:
  - 300 MB DB core
  - 110 MB index
  - 60 MB process DB

# Operačný technologický stack



Prostredie:	On prem virtualizácia / Cloud
OS:	Agnostické / Unix
Infra supp:	Docker Compose / Kubernetes
OS Docker:	Alpine /static files
WebServ:	NGINX / Apache HTTP
Middleware:	Active MQ, OpenJDK, React, Python - Flask
DB:	MongoDB, SOLR, PGSQL

## Dop. systémové zdroje:

- 8 vCPU
- 16-32 GB RAM
- SSD/SAS per DBs:
- TST: 150, 50, 50 GB
- PRD: 1000; 400; 150 GB

Rozhranie: REST API

## DL4DH Feeder



Klientská aplikácia (FE)

## TEI Converter



Datové konverzie (BE)

## Kramerius Plus



Adm. dát + obohacovanie (BE)

# Publikácia kódu (1/2023)

---

[LIBCAS/DL4DH](#)  
DL4DH – development of tools for effective utilization and mining of data from digital libraries to reinforce digital...  
☆ 8 GPL-3.0 license Updated 11 hours ago

---

[LIBCAS/DL4DH-Feeder](#)  
DL4DH Feeder  
● TypeScript Updated 13 days ago

---

[LIBCAS/DL4DH-TEI-Converter](#)  
DL4DH TEI Converter  
● Python Updated on May 20

---

[LIBCAS/DL4DH-Kramerius-plus](#)  
DL4DH Kramerius +  
● Java Updated on Oct 4

GIT - Průběžné nasazování:

- <https://github.com/LIBCAS/DL4DH>

- <https://github.com/LIBCAS/DL4DH-Feeder>

- <https://github.com/LIBCAS/DL4DH-Kramerius-plus>

- <https://github.com/LIBCAS/DL4DH-TEI-Converter>

# Exkurz k rozhraniu DL4DH

Feeder a Kramerius Plus

Národní knihovna České republiky  
DL4DH Feeder

Vyhledejte v DL4DH Feeder (základ slova nebo filtrujte výsledky)...

Hledat v K< Procházet Štítky Informace Exporty Přijít do Kramáře test English

Výsledky: 1 - 15/ 27 Pokročilé vyhledávání

Řazení Abecedně Exportovat

**Dostupnost**

- Pouze veřejně 27
- Pouze neveřejně 0
- Vše 27

**Dbohaceni**

- Pouze obohaceni 0
- Pouze neobohaceni 27

**Typ dokumentu**

- Kniha 27

**Klíčové slovo**

- česká literatura 1

**Autor**

- Němcová, Božena 20
- Kašpar, Adolf 2
- Rožek, Karel 2

**Zobrazit více**

**Jazyk**

- čeština 15

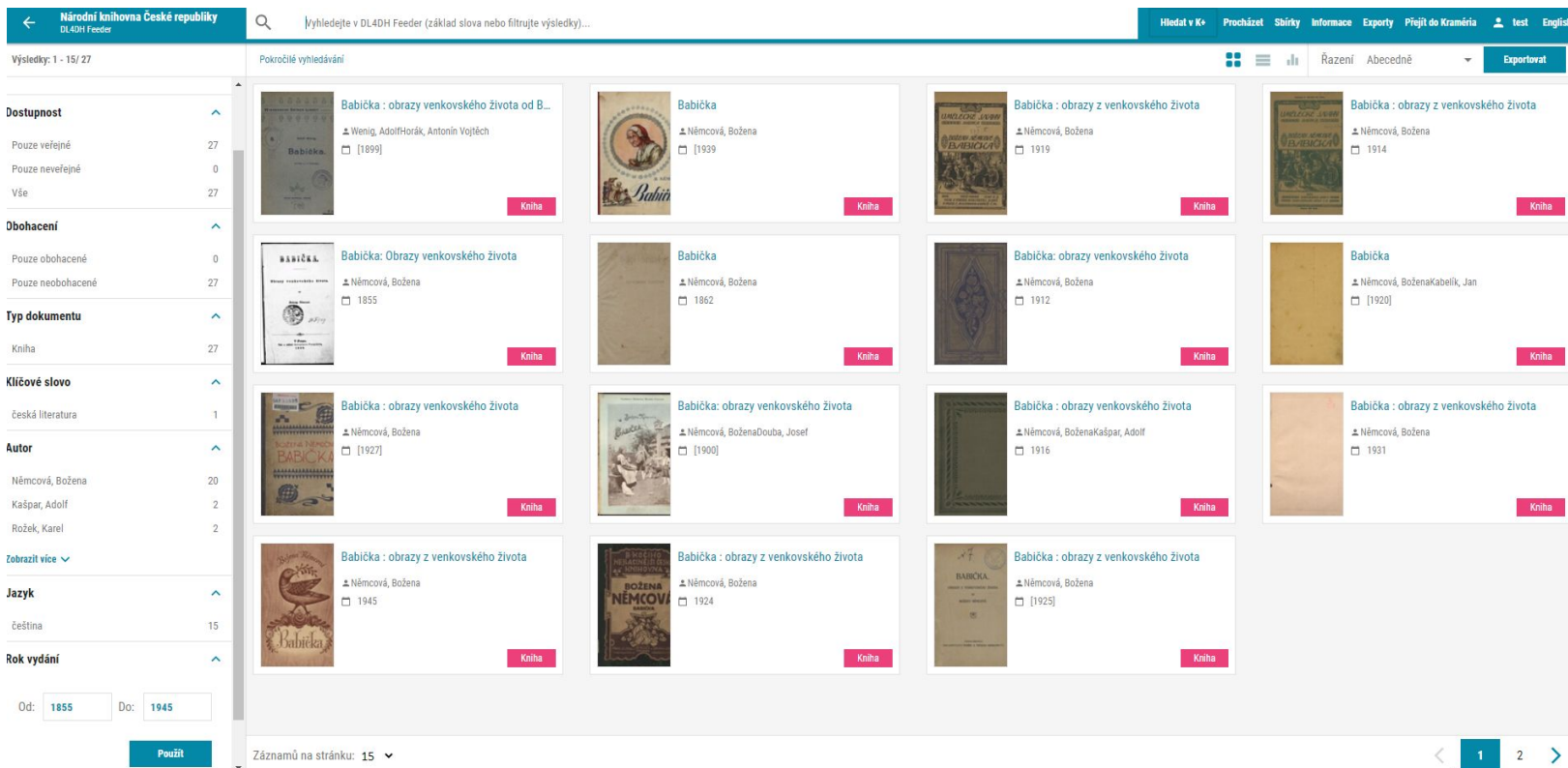
**Rok vydání**

Od: 1855 Do: 1945

Použít

Záznamů na stránku: 15

1 2



The screenshot displays a search results page for the book 'Babička'. The interface includes a search bar at the top with the text 'Vyhledejte v DL4DH Feeder (základ slova nebo filtrujte výsledky)'. Below the search bar, there are navigation options like 'Hledat v K<', 'Procházet', 'Štítky', 'Informace', 'Exporty', 'Přijít do Kramáře', 'test', and 'English'. The main content area shows a grid of 15 search results, each with a book cover image, the title 'Babička : obrazy venkovského života', the author 'Němcová, Božena', and the publication year. The results are filtered by availability (Dostupnost), enrichment status (Dbohaceni), document type (Typ dokumentu), keywords (Klíčové slovo), author (Autor), language (Jazyk), and publication year (Rok vydání). The year filter is set from 1855 to 1945. The page shows 15 results out of 27 total, with a 'Použít' button and a 'Záznamů na stránku: 15' dropdown menu. The bottom right corner shows page navigation with '1' and '2'.

## UC Babička (1899): GUI - Hľadanie a filtrovanie

## Detail žádosti o export

[STÁHNOUT HROMADNÝ EXPORT](#)

Vytvořil [FEEDER](#)  
 Vytvořeno v [7. 12. 2022 8:07:01](#)  
 Formát [TEXT](#)  
 Název -

## Dílčí exporthy

UUID publikace	Název publikace	Stav úlohy	Akce
uuid:a1dde1c0-50c2-11e7-b03f-005056827e52	Babička : obrazy venkovského života od Boženy Němcové	COMPLETED	<a href="#">DETAIL</a> <a href="#">STÁHNOUT</a>

UC Babička (1899): Spracovanie požiadavky na export - Kramerius+



Národní knihovna České republiky  
DL4DH Feeder

Vyhledejte v DL4DH Feeder (základ slova nebo filtrujte výsledky)...

Vybrat položky vlevo

Hledat na stránkách Detail záznamu

Hledat v obsahu stránek

1 z 68 stránek

WIESNEROVA SBÍRKA LIBRET

8. Adolf Wenig:  
**Babička.**  
HUDBA A. V. HORÁKA.

MORAVSKÁ ZEMSKÁ KNÍHOVNA V BRNĚ

Požadavka na export byla úspěšně vytvořena. Její stav můžete zkontrolovat na podstránce Exporty.

Vybrat položky vpravo

Hledat na stránkách Detail záznamu

Babička  
obrazy venkovského života od Boženy Němcové

Místo uložení  
Moravská zemská knihovna v Brně  
Signatura: 1-0072.975,8

Fyzický popis  
Rozsah: 63 s.

Autor  
Wenig, Adolf

Nakladatelské údaje  
A. Wiesner, [1899]

Typ dokumentu  
Kniha

Jazyk  
čeština

Poznámky  
Adolf Wenig ; hudbu složil A. Vojtěch Horák

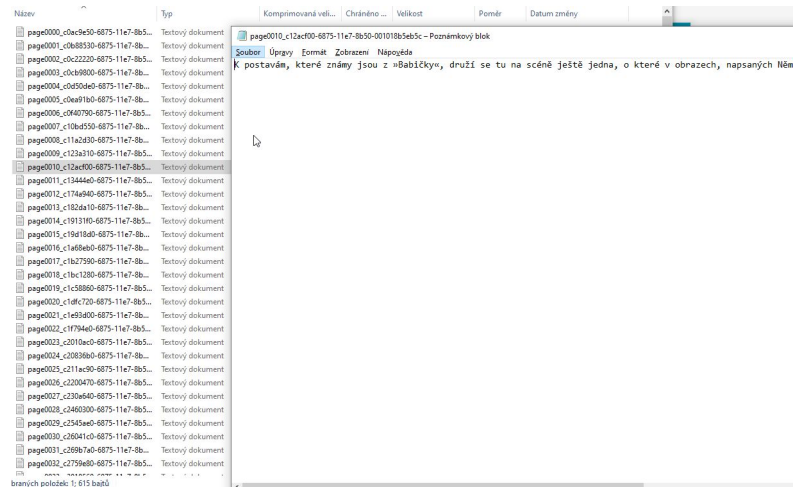
Kniha

UC Babička (1899): Zadané a spracované požiadavky na export



## Exporty


Název	Vytvořeno	Status	Formát	Akce
Babička : obrazy venkovského života od Boženy Němcové	07. 12. 2022	Úspěšně dokončen	TEXT	Stáhnout






The screenshot shows a file explorer window with a list of files on the left and a preview window on the right. The files are named with IDs like 'page0000\_c0ce9e50-6875-11e7-8b5...', 'page0001\_c0b68530-6875-11e7-8b5...', etc. The preview window shows the text content of a selected file: 'page0010\_c12ac000-6875-11e7-8b50-001018b5e5e - Poznámkový blok'. The text in the preview window reads: 'Soubor Úpravy Formát Zobrazení Nápověda' followed by 'K postavám, které známý jsou z »Babičky«, druží se tu na scéně ještě jedna, o které v obrazech, napsaných Něm'.

# UC Babička (1899): Formát TXT - Export dát

## Vytvoření nového plánu obohacení

Publikace Definice množiny publikací na obohacení Konfigurace Konfigurace spouštěných úloh 

- 1 Inicializace dat**  
Přepsat: Ne 
- 2 Obohacení externími nástroji**  
Přepsat: Ne  
Stratégia při chybějícím ALTO: Přeskočit 
- 3 Obohacení TEI**  
Přepsat: Ne 

INICIALIZACE DAT

OBOHACENÍ EXTERNÍMI NÁSTROJI

OBOHACENÍ NDK

OBOHACENÍ TEI

Název Definice názvu vytvářených úloh (nepovinné) 

VYTVORIT

UC Kramerius+ Babička(1899) - Obohatenie metadát

výsledky: 1 - 15 / 149

Pokročilé vyhledávání

Razení Abecedně Exportovat

**titvní filtry** ✕

Sbírka: Česká esoterická literatura  
Vše

**stupnost** ↑

ouze veřejné 78  
ouze neveřejné 71  
še 149

**ohaceni** ↑

ouze obohacené 36  
ouze neobohacené 113

**p dokumentu** ↑

niha 143  
oviny a časopisy 6

**čové slovo** ↑

piritismus 30  
kultismus 22  
arapsychologie 8

**razit více** ∨

**itor** ↑

ezemský, Karel 14  
leinfurter, Karel 12  
ouček, Rudolf 10

**razit více** ∨

<p><b>Adept, aneb, Výchova v záračného žáka</b> ▲ Arnold, Hans ☐ [1904]</p> <p>Knihna</p>	<p><b>Alchymie, její teorie a praxe</b> ▲ Lasenic, Pierre de ☐ 1936</p> <p>Knihna</p>	<p><b>Animismus a spiritismus: kritické prozkoumání</b> ▲ Aksakov, Aleksandr Nikolajevič ☐ 1903</p> <p>Obohacená Knihna</p>	<p><b>Bible ve světle mystiky. Řada IV. Evangelium</b> ▲ Weinfurter, Karel ☐ 1930</p> <p>Obohacená Knihna</p>
<p><b>Bratři růžového kříže (roseae crucis) v zemích</b> ▲ Volf, Josef ☐ 1908</p> <p>Obohacená Knihna</p>	<p><b>Co bude po smrti?</b> ▲ Lombroso, Cesare ☐ [1908]</p> <p>Knihna</p>	<p><b>Co jest Spiritismus?</b> ▲ Kardec, Allan J. K. ☐ [188-?] ]</p> <p>Obohacená Knihna</p>	<p><b>Co jest spiritismus?</b> ▲ Allan Kardec ☐ [1908]</p> <p>Knihna</p>
<p><b>Co víme o duši?</b> ▲ Souček, Rudolf ☐ 1943</p> <p>Knihna</p>	<p><b>Děj skutečnosti : rozmluva čtyř mužů, před...</b> ▲ ☐ 1925</p> <p>Knihna</p>	<p><b>Duch Estely : příklad mediální materialisace</b> ▲ Owen, Robert Dale ☐ 1894</p> <p>Knihna</p>	<p><b>Duchovní vůdce, který odpoutává duši : pra...</b> ▲ Molinos, Miguel deWeinfurter, Karel ☐ 1929</p> <p>Obohacená Knihna</p>
<p><b>Elektrické vidění na dálku : elektrické přen...</b> ▲ Sanholzer, Vilém ☐ 1928</p> <p>Knihna</p>	<p><b>Encyklopedie okultismu, mystiky a všech t...</b> ▲ Čapek, Otakar ☐ 1940</p> <p>Knihna</p>	<p><b>Fysikální mediumita : přehled vědeckých b...</b> ▲ Mikuláška, Viktor ☐ 1923</p> <p>Knihna</p>	

Záznamů na stránku: 15

1 2 3 4 5 6 7 8 9 10

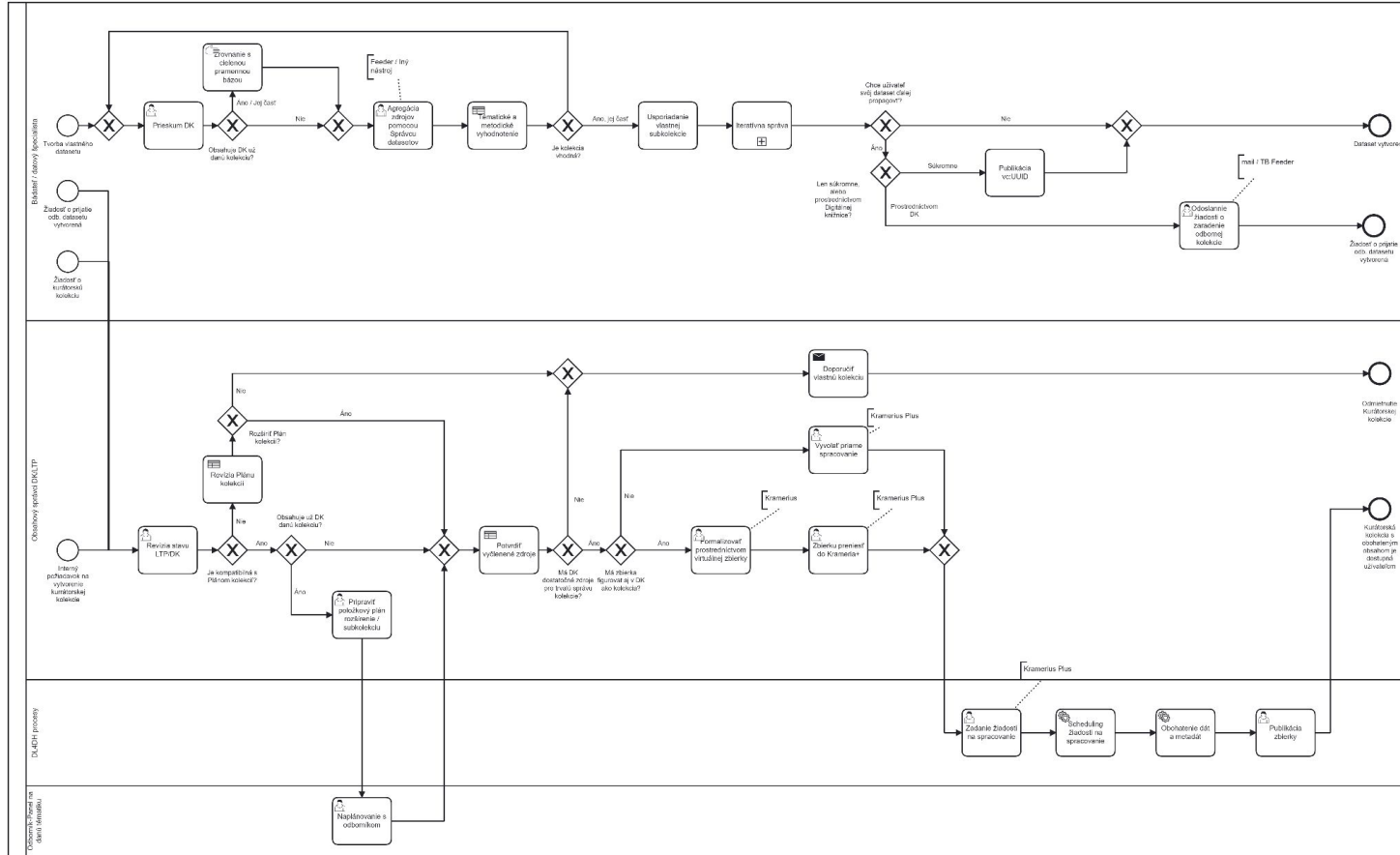
UC Zbierka: Česká esoterická literatura

# Procesy správy zbierok a obohacovania dát

TRS / TOBE

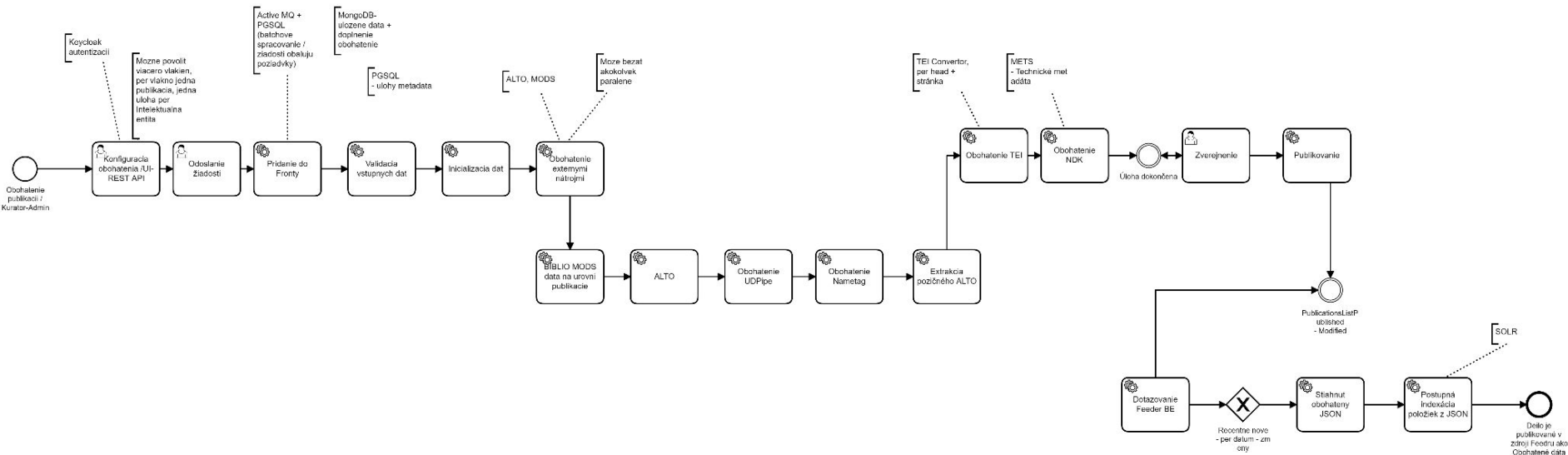
# Proces tvorby datasetov a kurátorských kolekcí

## 2nd level pohľad



# Proces obohacenia diela o obsah

## Detailný pohľad



# **Dopady na knihnice**



## Prevádzka riešenia:

- Infraštruktúrne prostriedky
- Operatívna podpora riešení
- Rozvoj a inovácia
- Správa operácií
- Správa užívateľov / int. CIAM

### - Pozície:

- Systémový administrátor
- Aplikačný správca
- Vývojár / UIX designér
- IT Security manager



## Obsah riešenia:

- Podpora obsahu samotného DK
- Správa kvality a metadát
- Tvorba a správa zbierok
- Jednanie s užívateľmi
- Pozície:

- Metodik
- Kurátor LTP
- Kurátor DK
- Metadatový špecialista
- Doménový špecialista
- Podpora DH



## Podpora uživatelův řešení:

- Začlenenie do online komunikácie knižníc
- Poučenie personálu študovní a referenčných centier
- Užšie prepojenie technických a obsahových pracovníkov knižnice
- Příprava na individuálne otázky datovo-analytickej povahy
- Příprava podpory väčších tímov
- Posudzovanie vybraných zbierok odborníkov ako potenciálnych kurátorských kolekcíí
- Priebežný zber a analýza požiadaviek na rozšírenie riešenia



# Ďakujem za pozornosť !

Mgr. Zdenko Vozár

[zdenko.vozar@gmail.com](mailto:zdenko.vozar@gmail.com)

National Library of the Czech Republic