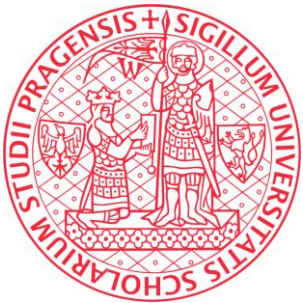


Digitální repozitář(e) na Univerzitě Karlově

Jakub Řihák

Oddělení repozitářů, digitalizace a depozitáře



UNIVERZITA
KARLOVA
Ústřední knihovna

jakub.rihak@ruk.cuni.cz

Prezentace je dostupná pod licencí Creative Commons, licence: [CC-BY-SA-4.0](https://creativecommons.org/licenses/by-sa/4.0/),
via <http://www.nusl.cz/ntk/nusl-367302>

[Twitter](#)
[LinkedIn](#)

Obsah

- Výchozí stav
- Základní principy
- Workflow...
- ... a jeho automatizace
- Aktuální stav repozitáře
- Krátkodobé plány
- Dlouhodobé plány

Výchozí stav

- Různé repozitáře pro různé typy digitalizovaného a digital-born obsahu
- Elektronické kvalifikační práce – **Repozitář závěrečných prací** – nadstavba nad SIS
 - Nedostatečné možnosti propojení
 - ostatní dokumenty (digitalizované i digital-born) – **DigiTool**
 - drahý, licenční poplatky založeny na počtu digitálních objektů, zastaralý

Výchozí stav

- poptávka po změně
 - Peníze věnovat na lidi, ne na SW (nebo ne tolik)
 - Stát se součástí komunity
 - Učit se z existujících řešení, těch která jsou **opravdu používána**
- **2014 / 2015**
 - hledání nového repozitářového systému
 - Otázka: „Na jakém typu dokumentů máme repozitář založit?“

Výchozí stav

- Jaký systém budeme využívat? DSpace!
- Na čem založíme jeho obsah? Elektronické kvalifikační práce!
- 1. polovina roku 2016
 - instalace / školení / experimenty
- 2. polovina roku 2016
 - Započetí prací – musí být hotovo do 1. 1. 2017

Základní principy

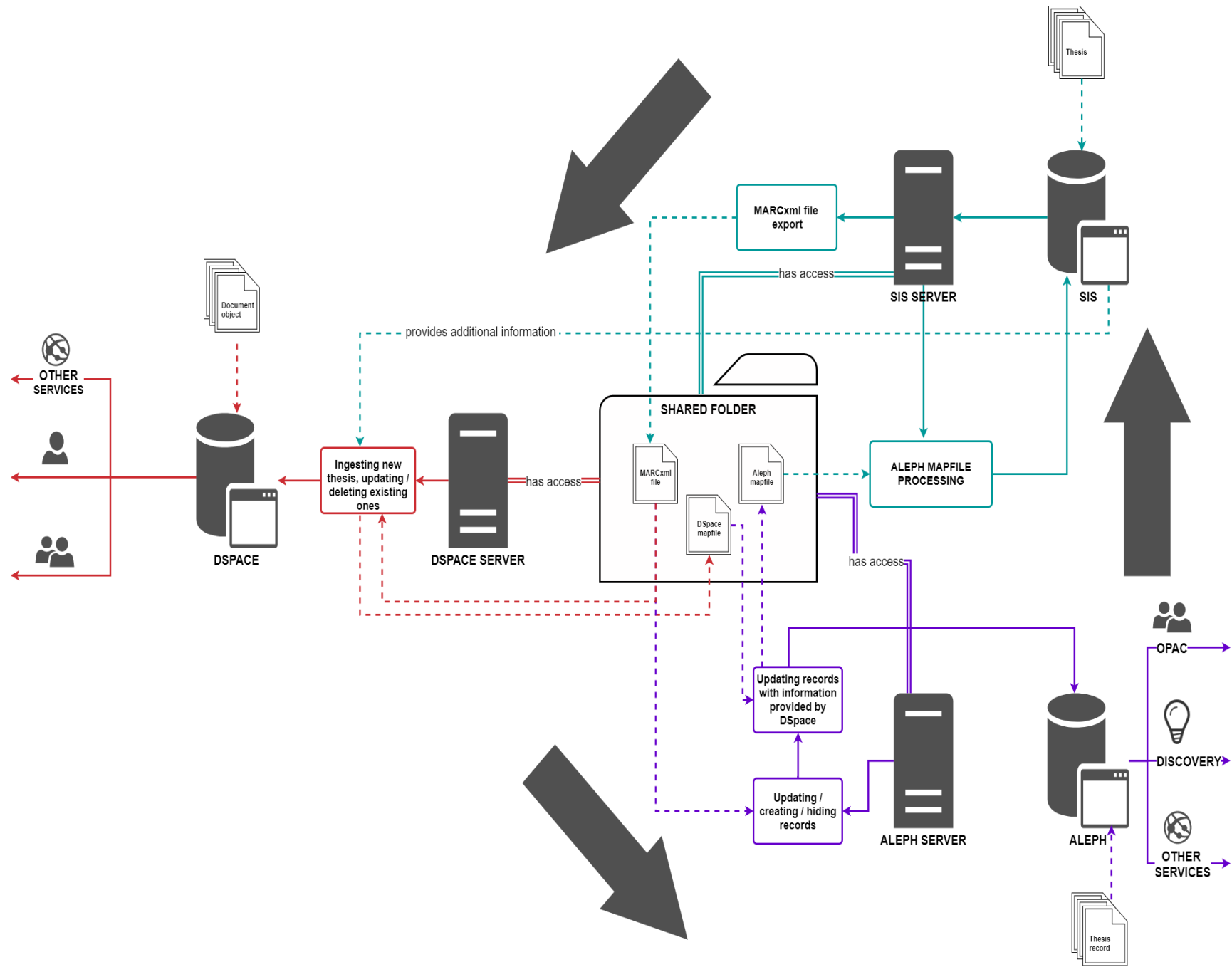
- Kvalifikační práce musí být do repozitáře vkládány přímo ze Studijního informačního systému
- Bez zbytečných zásahů ze strany uživatele
- Vložené práce musí mít trvalé odkazy a identifikátory
- Musí být přístupné z:
 - Knihovního katalogu
 - Discovery systému

Základní principy

- **Struktura DSpace** (komunity, sbírky, atd.) by měly kopírovat strukturu univerzity
 - → Fakulty (komunity)
 - → typy dokumentů (sbírky)
 - → objekty / jednotky
- Důraz na **automatizaci** (velké množství kvalifikačních prací každý rok)

Definice workflow

- použití existujícího workflow
 - Místo přímého přístupu k SIS
- **jednodušší** a rychlejší na zavedení do praxe
- Využití existujícího workflow pro napojení repozitáře na knihovní systém Aleph pomocí **existujících metod**
- Kvalifikační práce jsou dostupné z katalogu i discovery systému



Automatizace workflow

- Kvalifikační práce by měly být zpracovávány alespoň jednou denně
- Vkládání by mělo probíhat pomocí dostupných nástrojů DSpace nebo pomocí DSpace API
- Automatizovaný proces vkládání by měl využívat existující zdroje, pokud je to možné
- Ale jak přesně to udělat? Co přesně potřebujeme?
 - metadata, soubory, metodu vkládání

Metadata

- DSpace využívá formát **Dublin Core**
- V základu poskytuje 2 metadatová schémata
- Existující schémata mohou být rozšířena nebo mohou být vytvořena nová schémata
 - To byl náš případ → potřebujeme uložit nestandardní metadata, pro rozšíření funkcionality o další **fasety** a **vyhledávací filtry**



Metadata registry

The metadata registry maintains a list of all metadata fields available in the repository. These fields may be divided amongst multiple schemas. However, DSpace requires the qualified Dublin Core schema. You may extend the Dublin Core schema with additional fields or add new schemas to the registry.

ID	Namespace	Name
1	http://dublincore.org/documents/dcmi-terms/	dc
<input type="checkbox"/>	http://purl.org/dc/terms/	dcterms
<input type="checkbox"/>	http://dspace.org/eperson	eperson
<input type="checkbox"/>	http://cuni.cz/schema/uk-results-schema/	uk
<input type="checkbox"/>	http://www.ndltd.org/standards/metadata/etdms/1.0/	thesis

Delete schema

Add a new schema

Namespace: *

Namespace should be an established URI location for the new schema.

LINKS

Useful links



BROWSE

Whole repository



MY ACCOUNT

Logout

Profile

Submissions

ADMINISTRATIVE

Control Panel

Statistics

Curation Tasks

Access Control



Content Administration





Hledejte v Digitálním repozitáři UK

Hledat

Sbírky fakult Univerzity Karlovy

Vyberte fakultu k procházení jejích sbírek.

1. lékařská fakulta [2641]

First Faculty of Medicine

2. lékařská fakulta [1038]

Second Faculty of Medicine

3. lékařská fakulta [2114]

Third Faculty of Medicine

Evangelická teologická fakulta [1133]

Protestant Theological Faculty

ODKAZY

[Užitečné odkazy](#)



PROCHÁZET

[Celý repozitář](#)



MŮJ ÚČET

[Přihlásit se](#)

PROHLÍZENÍ

[Fakulta](#)



1. lékařská fakulta (2634)

2. lékařská fakulta (1037)

3. lékařská fakulta (2114)

Evangelická teologická fakulta
(1133)

Fakulta humanitních studií (4332)

```
<subfield code="a">Univerzita Karlova.</subfield>
<subfield code="b">Katedra fyzikální a makromol. chemie</subfield>
</datafield>
```

```
<datafield tag="850" ind1=" " ind2=" ">
```

```
<subfield code="a">PRF</subfield>
</datafield>
```

```
<datafield tag="IDS" ind1=" " ind2=" ">
```

```
<subfield code="a">149396</subfield>
</datafield>
```

```
<controlfield tag="repId">193071</controlfield>
<controlfield tag="didId">193071</controlfield>
<controlfield tag="func">insert</controlfield>
<controlfield tag="ds_dateAccepted">31-08-2017</controlfield>
<controlfield tag="ds_workType">Rigorózní práce</controlfield>
<controlfield tag="ds_academicTitle">RNDr.</controlfield>
<controlfield tag="ds_facultyName_cs">Přírodovědecká fakulta</controlfield>
<controlfield tag="ds_facultyName_en">Faculty of Science</controlfield>
<controlfield tag="ds_facultyAbbr">PřF</controlfield>
<controlfield tag="ds_publication_place">Praha</controlfield>
<controlfield tag="ds_finalGrade_cs">Prospěl</controlfield>
<controlfield tag="ds_finalGrade_en">Pass</controlfield>
<controlfield tag="ds_studyLevel">rigorózní řízení</controlfield>
<controlfield tag="ds_studyField_cs">Modelování chemických vlastností nano- a biostruktur</controlfield>
<controlfield tag="ds_studyField_en">Modeling of Chemical Properties of Nano- and Biostructures</controlfield>
<controlfield tag="ds_studyProgram_cs">Chemie</controlfield>
<controlfield tag="ds_studyProgram_en">Chemistry</controlfield>
<controlfield tag="ds_departmentName_cs">Katedra fyzikální a makromol. chemie</controlfield>
<controlfield tag="ds_departmentName_en">Department of Physical and Macromolecular Chemistry</controlfield>
<controlfield tag="ds_keywords_cs">molekulární dynamika, simulace spekter, kvantová chemie, chiralita, optická aktivita</controlfield>
<controlfield tag="ds_keywords_en">molecular dynamics, spectra simulations, quantum chemistry, chirality, optical activity</controlfield>
<controlfield tag="ds_work_availability">V</controlfield>
</record>
```

Metoda vložení

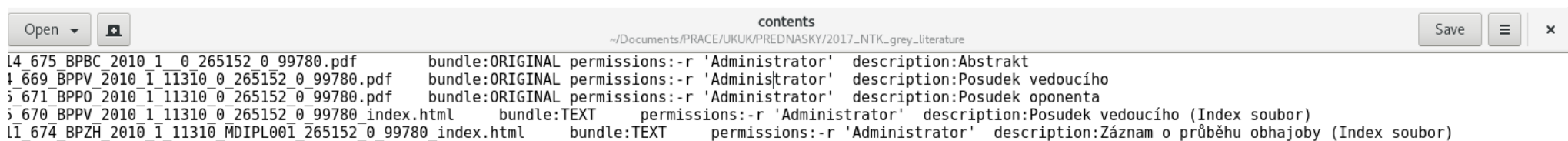
- Množství dostupných metod (uživ.rozhraní, API, CSV, balíčky,...)
- **Simple Archive Format balíček** + import pomocí příkazové řádky
- Jednoduchý způsob, jak kontrolovat chyby v balíčku, jeho strukturu a obsah – nápomocné během vývoje

```
archive_directory/  
  item_000/  
    dublin_core.xml      -- qualified Dublin Core metadata for metadata fields belonging to the dc schema  
    metadata_[prefix].xml -- metadata in another schema, the prefix is the name of the schema as registered with the metadata registry  
    contents             -- text file containing one line per filename  
    collections         -- text file that contains the handles of the collections the item will belong two. Optional. Each handle in  
                        -- Collection in first line will be the owning collection  
                        -- files to be added as bitstreams to the item  
    file_1.doc  
    file_2.pdf  
  item_001/  
    dublin_core.xml  
    contents  
    file_1.png  
    ...
```

Metoda vložení

```
<dublin_core>
  <dcvalue element="title" qualifier="none">A Tale of Two Cities</dcvalue>
  <dcvalue element="date" qualifier="issued">1990</dcvalue>
  <dcvalue element="title" qualifier="alternative" language="fr">J'aime les Printemps</dcvalue>
</dublin_core>
```

(Note the optional language tag attribute which notifies the system that the optional title is in French.)

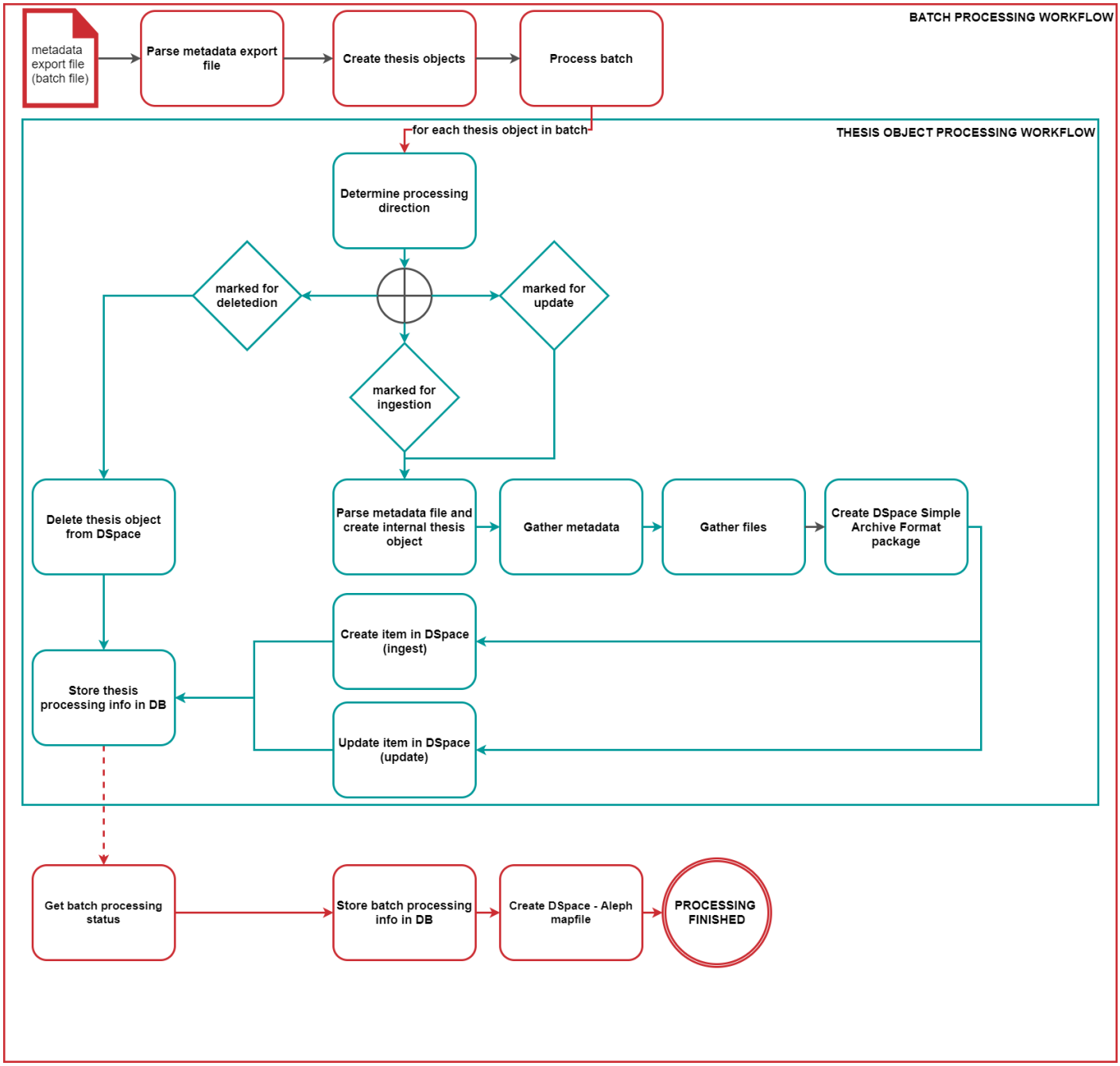


The screenshot shows a file manager window titled "contents" with the path "~/Documents/PRACE/UKUK/PREDNASKY/2017_NTK_grey_literature". The window contains a list of files with their metadata, including file names, bundle types, permissions, and descriptions.

File Name	Bundle	Permissions	Description
l4_675_BPBC_2010_1_0_265152_0_99780.pdf	bundle:ORIGINAL	permissions:-r 'Administrator'	description:Abstrakt
l_669_BPPV_2010_1_11310_0_265152_0_99780.pdf	bundle:ORIGINAL	permissions:-r 'Administrator'	description:Posudek vedoucího
l_671_BPP0_2010_1_11310_0_265152_0_99780.pdf	bundle:ORIGINAL	permissions:-r 'Administrator'	description:Posudek oponenta
l_670_BPPV_2010_1_11310_0_265152_0_99780_index.html	bundle:TEXT	permissions:-r 'Administrator'	description:Posudek vedoucího (Index soubor)
l_674_BPZH_2010_1_11310_MDIPL001_265152_0_99780_index.html	bundle:TEXT	permissions:-r 'Administrator'	description:Záznam o průběhu obhajoby (Index soubor)

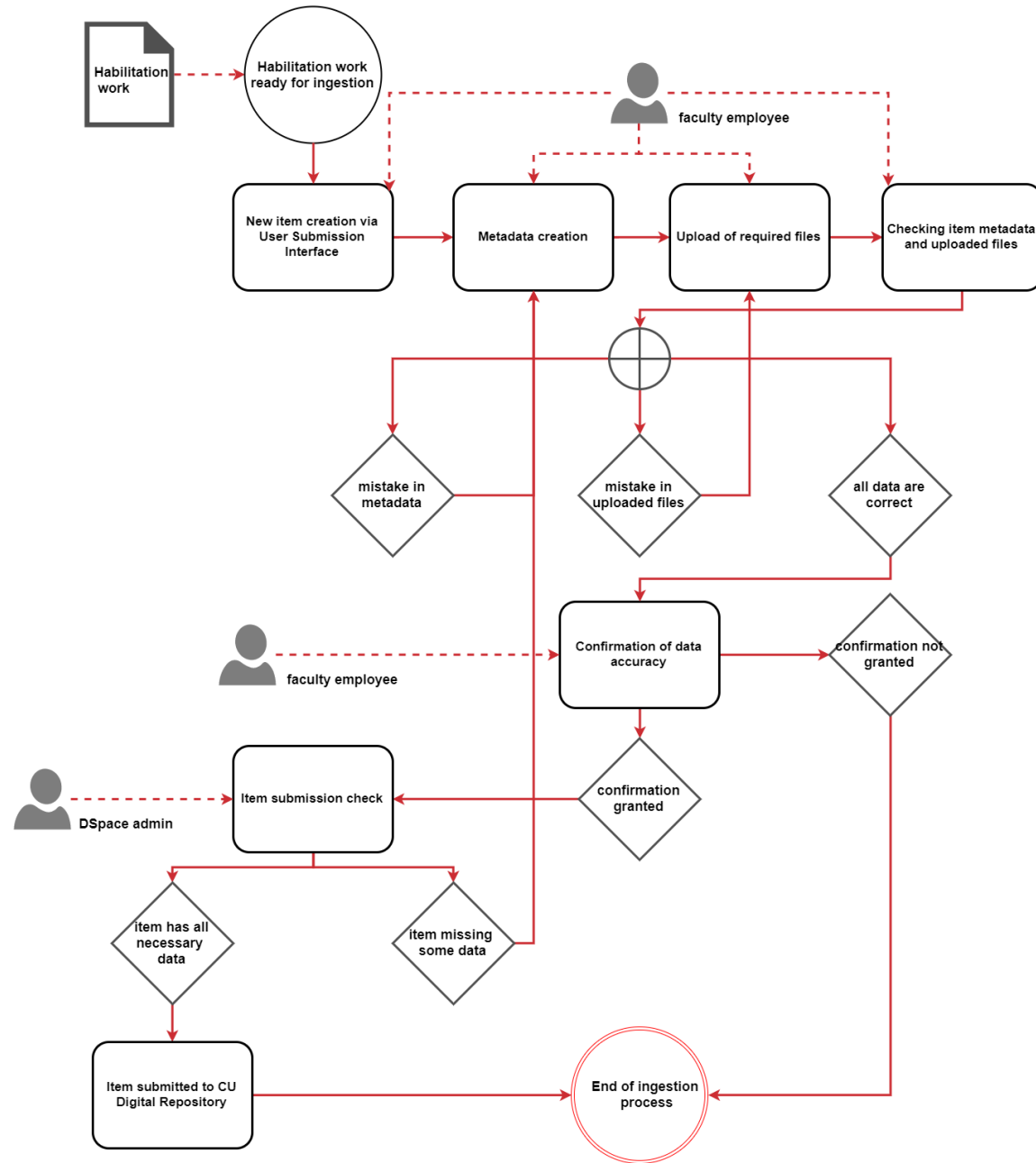
Automatizační nástroj

- Python3 + PostgreSQL databáze
- Exportovaný metadatový soubor ze SIS → 'dávka'
- Kontroluje nové dávky každých 15 minut
 - V základu pouze jeden export prací denně
- **Workflow zpracování do DSpace**



Aktuální stav repozitáře

- Dostupný na <https://dspace.cuni.cz>
- Přes 91k objektů
- Napojený na Národní úložiště šedé literatury
 - OAI-PMH protokol
- Zahrnutý v OpenDOAR
- Workflow pro vkládání habilitačních prací
- Nelze je vkládat automaticky (prozatím)
 - Využití formuláře pro vložení
- Aktuálně nová funkcionality: automaticky generované citace



Krátkodobé plány

- Autentizace uživatelů přes Shibboleth
- Horizont2020 – zpřístupňování vědeckých publikací
- Elektronické knihy pro studenty se speciálními potřebami
- Převod obsahu z repozitáře DigiTool
- Vytvoření Centrální Digitální knihovny pro historické knihy, periodika a mapy
 - Kramerius

Dlouhodobé plány

- **Centrální přístupový bod** k digitalizovaným a digital-born dokumentům Univerzity Karlovy
 - Analýza současného stavu (probíhá)
 - → strategie vývoje a rozvoje Centrálního přístupového bodu a možnosti **dlouhodobého uchování** digitalizovaného a digital-born obsahu

Závěr

- 6 měsíců vývoje
 - Díky důrazu na automatizaci
- Přístup k rostoucímu množství dokumentů a typů dokumentů
- Napojení na národní a mezinárodní služby
- Digitální repozitář pro digital-born obsah
 - Kvalifikační práce jsou pouze začátek
- Bude existovat separátní digitální knihovna pro prezentaci digitalizovaných historických dokumentů
- Integrace do Centrálního přístupového bodu

Děkuji za pozornost