

Malá digitalizace

Michal Denár

Městská knihovna Česká Třebová
KISK MU

O čem to bude?

- proč se snažíme o digitální knihovnu
- co tam chceme publikovat
- jak digitalizujeme
- jaké nástroje používáme
- jaké problémy řešíme



O čem to bude?

- služby
- technologie
- procesy
- software
- standardy
- autorské právo



Národní digitální knihovna

Národní digitální knihovna

Projekt je financován částkou cca 300 mil. Kč v rámci Výzvy č. 07 Integrovaného operačního programu "Elektronizace služeb veřejné správy", prioritní osa 1: Modernizace veřejné správy, oblast intervence 1.1 Rozvoj informační společnosti ve veřejné správě.

Národní digitální knihovna

Národní digitální knihovna

Z důvodu omezení Zákonem č. 121/2000 Sb. (autorský zákon) může být plný text zobrazen pouze v prostorách knihovny nebo instituce, která dokument zapůjčila k digitalizaci. V případě zájmu jsou poskytovány elektronické kopie částí knih nebo sborníků v max. rozsahu 20 stran na jednu objednávku a kopie článků z periodik (v pdf formátu).

Národní digitální knihovna

Z důvodu omezení Zákonem č. 121/2000 Sb. (autorský zákon) může být plný text zobrazen pouze v prostorách knihovny nebo instituce, která dokument zapůjčila k digitalizaci. V případě zájmu jsou poskytovány elektronické kopie částí knih nebo sborníků v max. rozsahu 20 stran na jednu objednávku a kopie článků z periodik (v pdf formátu).

Digitalizované dokumenty podléhající ochranné lhůtě Autorského zákona jsou dostupné pouze v knihovně, jinak jsou zpřístupňovány pouze popisné údaje (metadata).

Malá digitalizace?

- knihovna v e-době
- digitalizace=zpřístupnění
- regionální témata a díla
- regionální autoři
- komerčně nedostupná díla
- e-publikování
- ekosystém služeb

Ekosystém e-slужeb



Požadavky

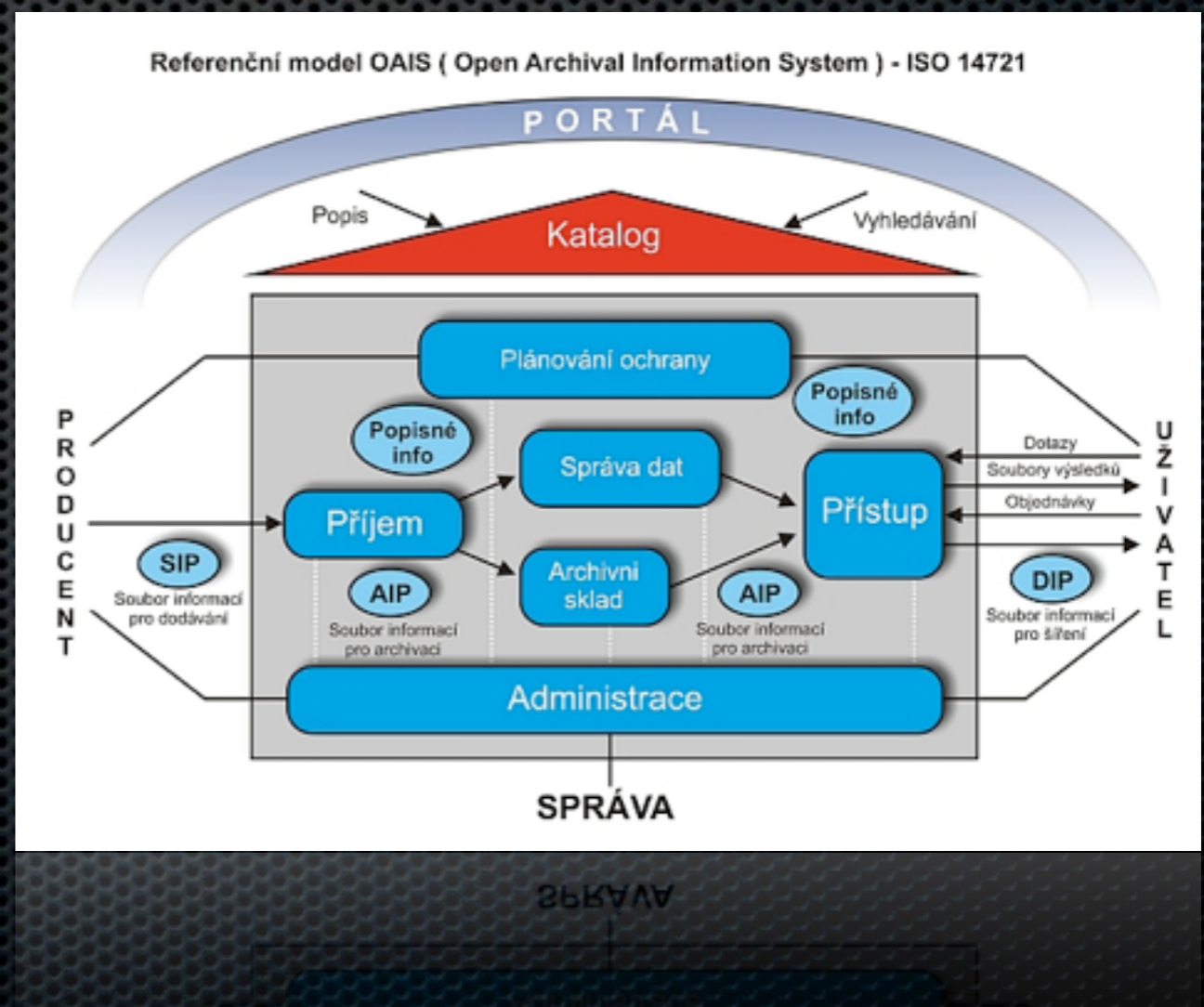
- open-source (cena, nezávislost)
- standardy NDK, možnost propojení
- metadata (ALTO XML)
- driver pro VuFind
- dostupné hw (virtualizace)

= Kramerius

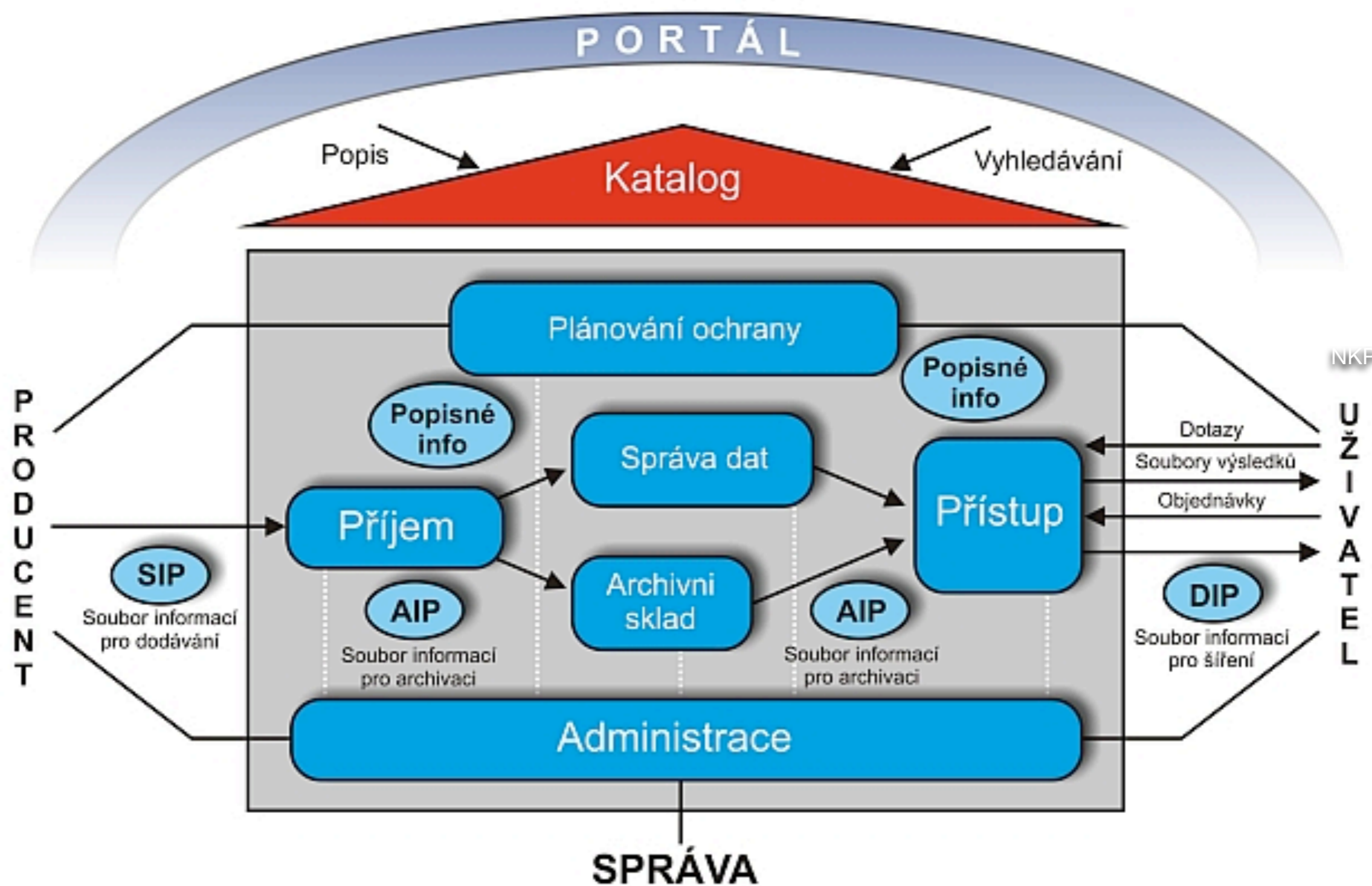
Konceptuální model OAIS

- funkční a důvěryhodný repozitář
- SIP
- AIP
- DIP

NKP



Referenční model OAIS (Open Archival Information System) - ISO 14721



Technické řešení

Kaila

2 TB HDD
24 GB RAM
2,66 Ghz
4 jádra
RAID

Debian
Ubuntu



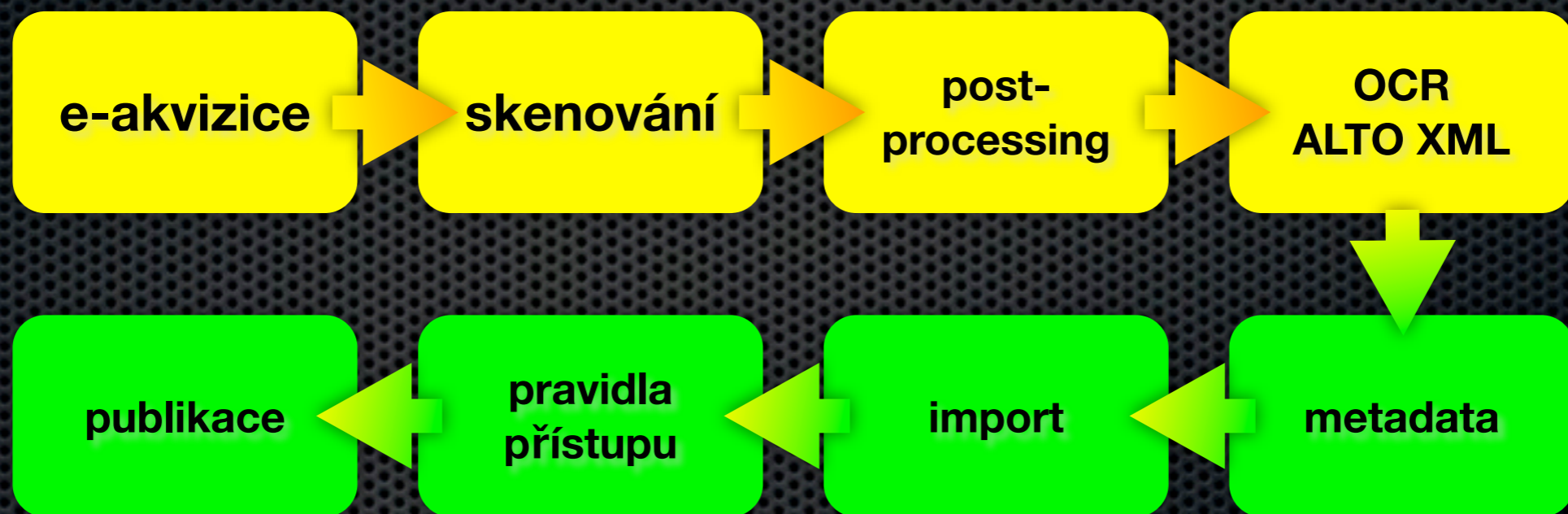
Keira

4 TB HDD
24 GB RAM
2,66 Ghz
4 jádra
RAID

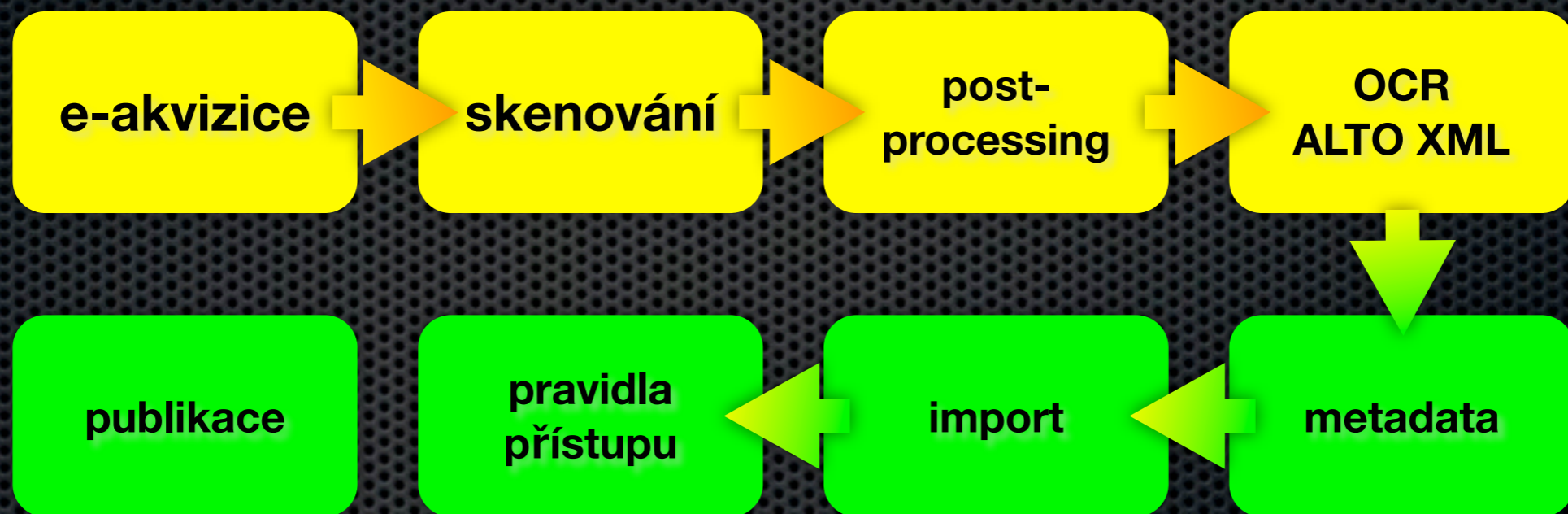
VMware
Essentials



Digitalizační workflow



Digitalizační workflow



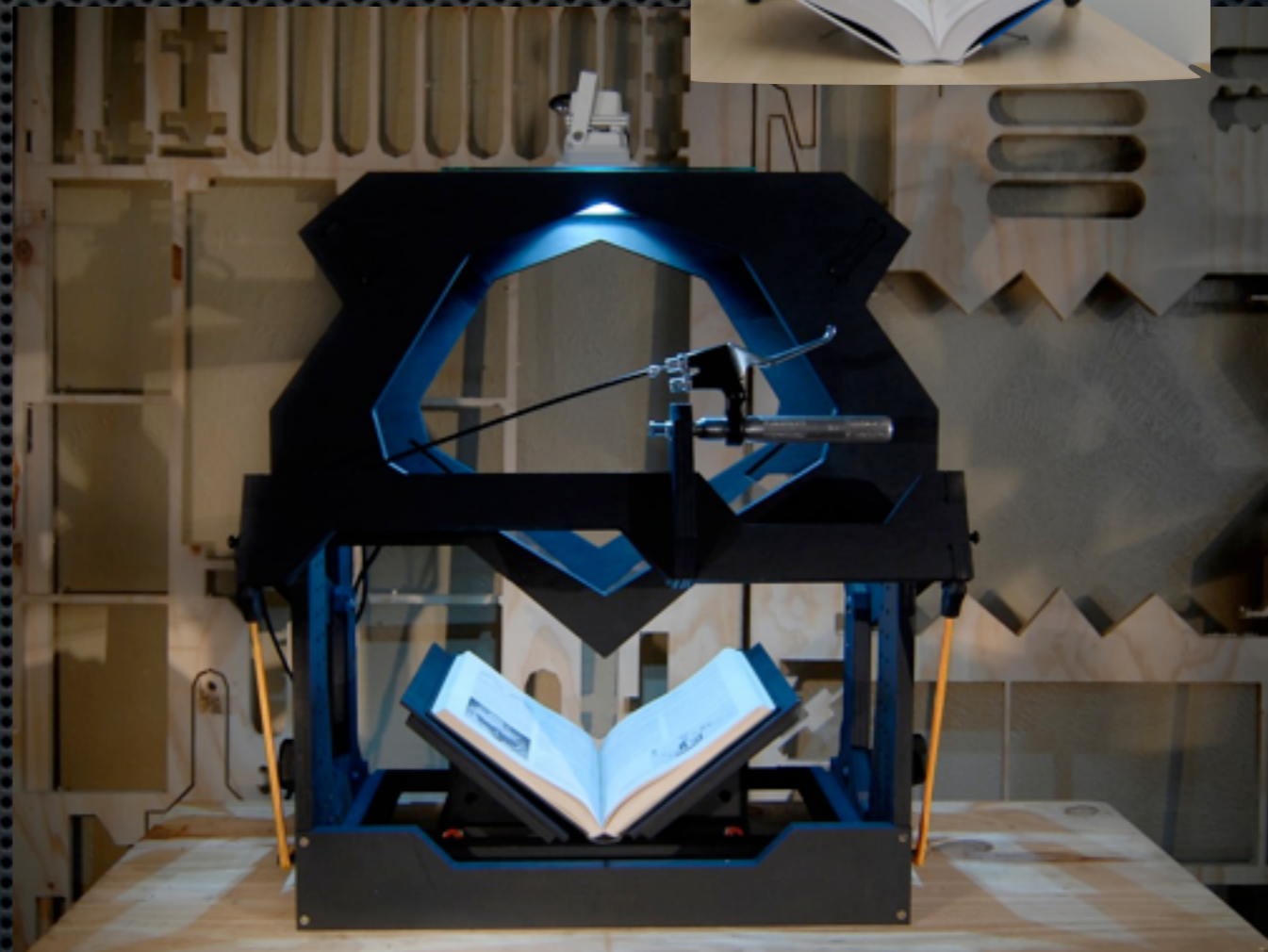
E-akvizice

- výběr a vyhledávání dokumentů
- vyjednávání s autory a vydavateli
- nový AZ: osiřelá a nedostupná díla
- spolupráce: lidé, paměťové instituce
- born digital: vlastní produkce

Skenování

- stolní skener
- **DIY skener**
- služba

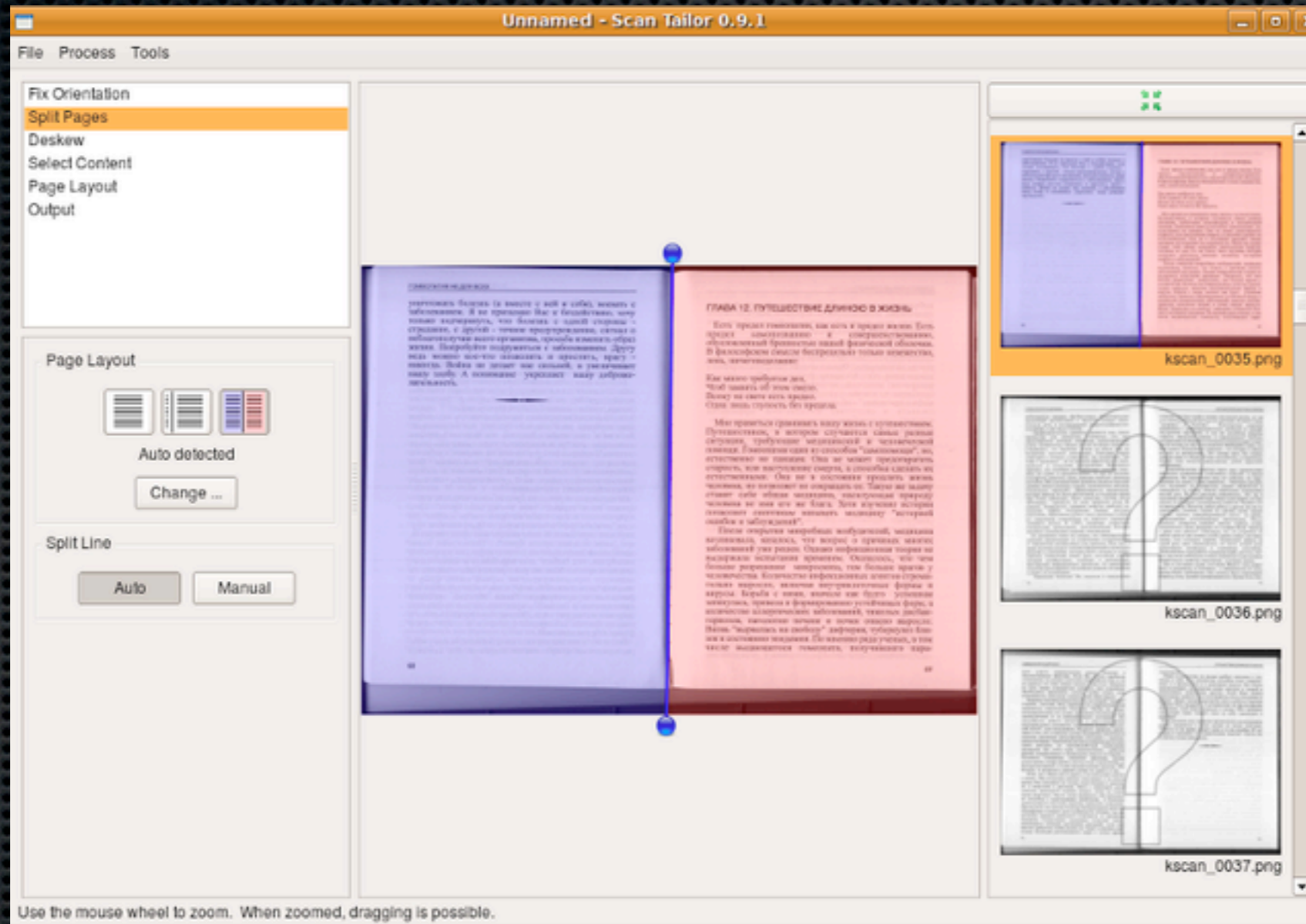
Video



Postprocessing

- ScanTailor (open-source)
- rozdělení na jednotlivé strany
- vyrovnání stran
- stejné okraje
- korekce kontrastu, čištění
- zdrojové soubory pro OCR a import

ScanTailor



OCR

- předpoklad pro fulltextové hledání
- spolu s ALTO XML: výsledkem hledání je konkrétní část textu na stránce
- chybovost výstupu=crowdsourcing
- ručně psané dokumenty

Meta-editor

- indexace a popis digitálních objektů
- kompletní proces importu
- nastavení náhledů
- přiřazení OCR a ALTO ke stranám
- pravidla přístupu
- Dublin Core, Mods
- FOXML, METS

Medit

MORAVSKÁ ZEMSKÁ KNIHOVNA

Vstupní fronta

ID	Jméno
monograph	
ANL	
testNewOcr	
000830944	Quintetto ex G mol F
000831081	
000831091	
2610203052	Diagnostika naruše
2610279426	
2610402007	
2610435937	
2610466147	
2610469032	
2610469033	
2610469034	
2610469035	
2610469036	
2610471250	
2610472025	
2610473470	
2619203969	Verse : (1892-1896)
2619204685	Magyarische Sagen
2619207421	
2619210181	Zákon o používání a
2619210260	
2619211646	Aristotelova kniha o
2619211959	Luhačovice, alkalick
2619211976	
2619211986	
2619212044	
2619212444	
2619212506	
2619212617	

Vytvořit podstrukturu

Podstruktury

Metadata

Najít metadata protokolem OAI-PMH

URL poskytovatele :

Prefix identifikátoru :

Báze :

SYSNO :

Hledej

Najít metadata protokolem Z39.50

Hledat podle :

Čárový kód :

Hledej

Výsledky hledání

Název	Vydavatel	Datum
Aristotelova kniha o duši ; přel. a vysvětliv. opatř. P. J. Vych	Brno : Papežská knihtiskárna	1895

Detail

Název:	Aristotelova kniha o duši ; přel. a vysvětliv. opatř. P. J. Vychodil
Identifikátor:	http://kramerius.mzk.cz/search/handle/uid:eca75ae1-0562-11e1-ad36-0050569d679d
Vydavatel:	Brno : Papežská knihtiskárna benediktinů rajhradských,
Datum:	1895
Popis:	Dostupné též v digitalizované podobě v aplikaci Kramerius
Jazyk:	cze
Předmět:	

Další Pokračovat bez metadat

Publikace

- vlastní webové rozhraní
- vyhledávání
- prohlížení
- možnost stažení, tvorba PDF (omezeno AZ)
- persistentní URL: na dokument, stranu
- integrace s VuFindem

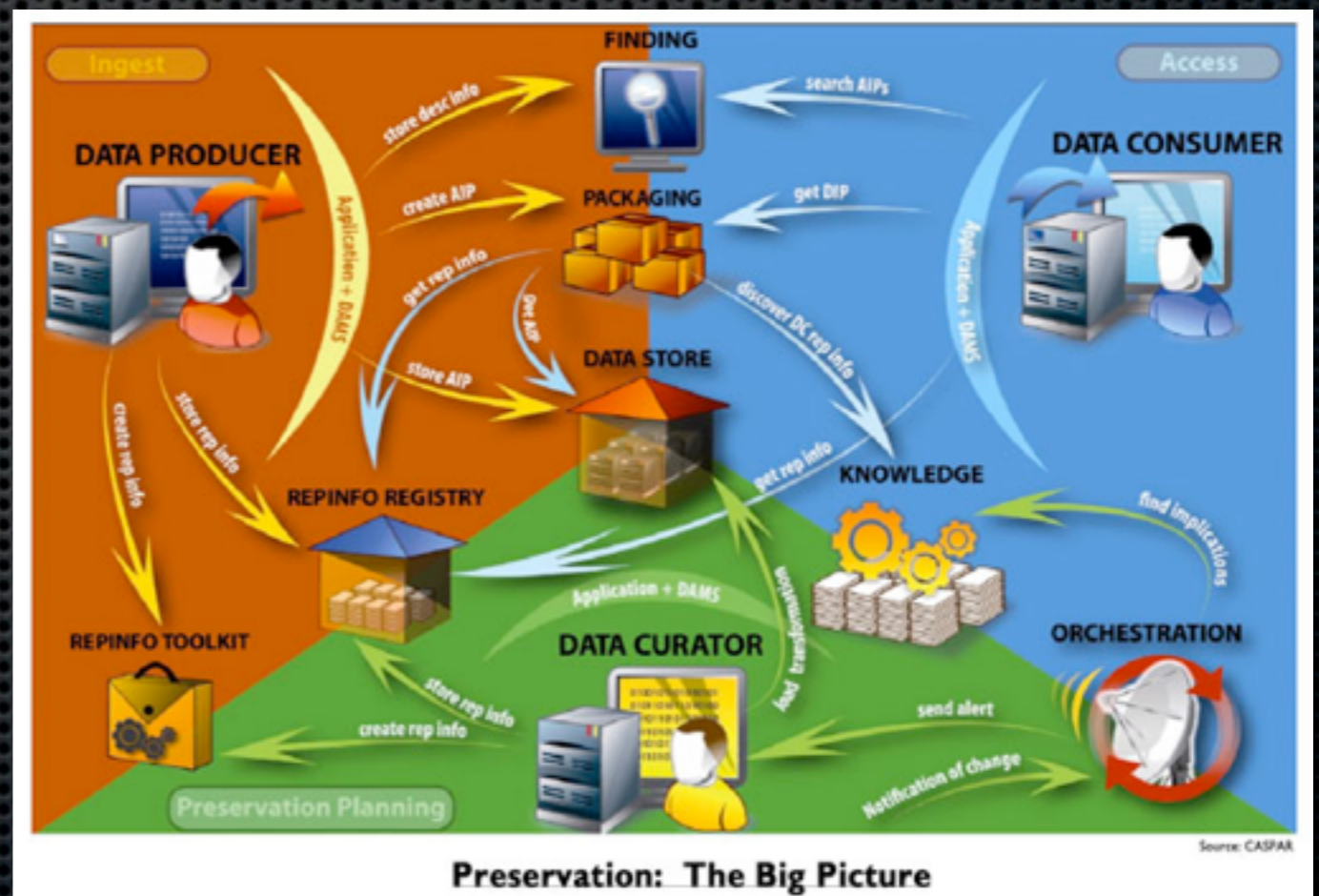
Kramerius 4

The screenshot displays the Kramerius 4 digital library interface. At the top left is the logo 'K4' in orange. To its right, the text reads 'Digitální knihovna obsahuje 2933 monografií'. A search bar with a magnifying glass icon and the text 'Pokročilé vyhledávání' is positioned to the right of the logo. Below the search bar are navigation links: 'Domů', 'Nápověda', 'Přihlášení', 'Registrace', 'english', and 'česky'. The main content area is divided into two columns. The left column features a 'Časová osa' (Timeline) with a vertical axis and a slider. The timeline is divided into decades: '1960 - 1969', '1970 - 1979', '1980 - 1989', '1990 - 1999', '2000 - 2009', and '2010 - 2019'. Above the timeline, there are input fields for 'Od: 01.01.1953' and 'Do: 31.12.2012'. The right column contains a grid of book covers and thumbnails. The first row includes a book cover titled 'XML' and five thumbnails of historical maps. The second row contains six thumbnails of historical maps. The third row contains six thumbnails of historical maps. At the top of the main content area, there are tabs for 'Vybrané', 'Nežádanější', 'Nejnovější', 'Navigace', 'Tituly', 'Autoři', 'Virtuální sbírky', and 'Informace'.

Long Term Preservation

- ochrana
- kompatibilita
- použitelnost
- vyhledatelnost

CASPAR

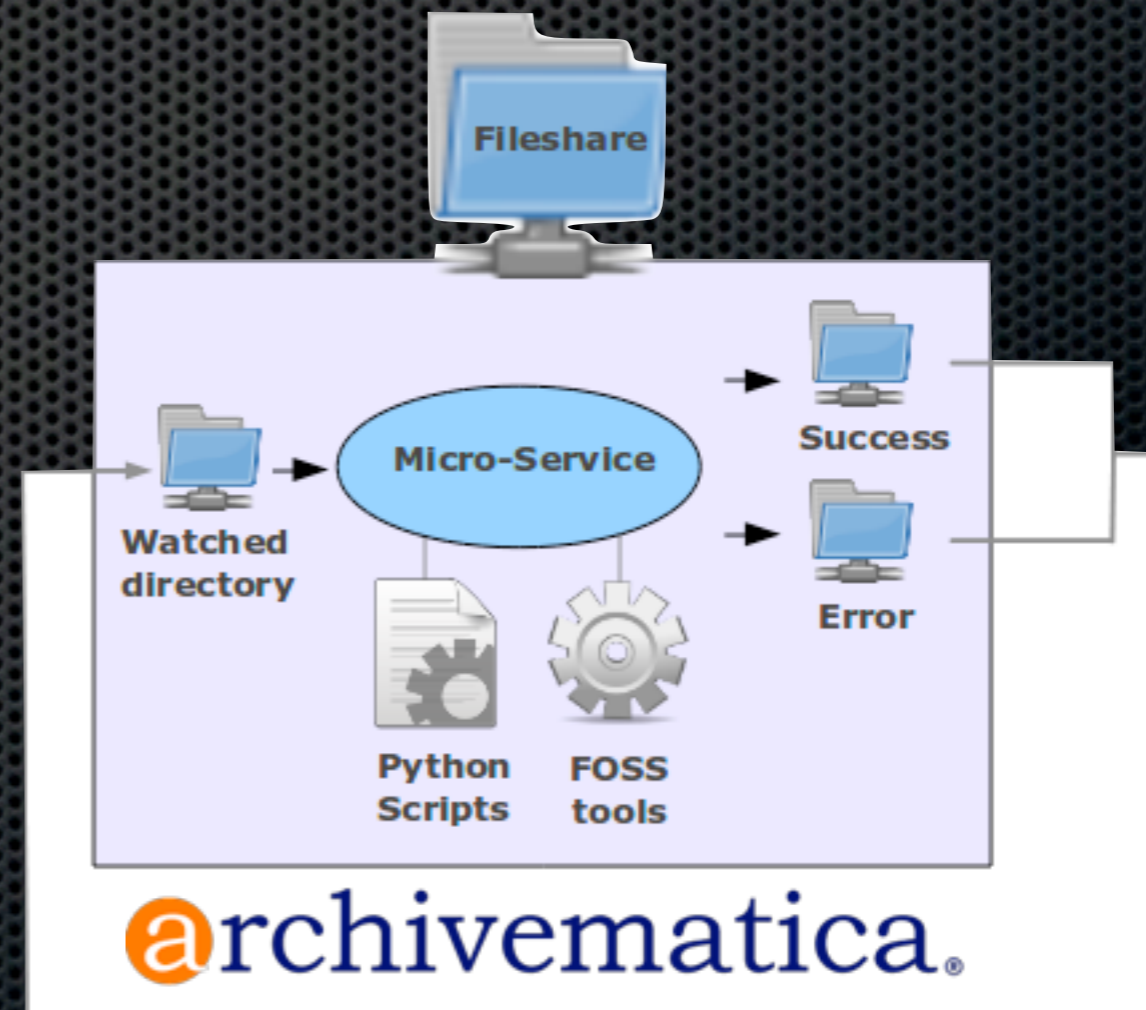


Preservation: The Big Picture

PRESEVATION: THE BIG PICTURE

Archivematica

- dokument
- tabulka
- obraz
- video
- zvuk
- další



Archivematica

The screenshot shows the Archivematica Dashboard interface in a Mozilla Firefox browser window. The browser's address bar displays 'localhost/transfer/#'. The page header includes the Archivematica logo and navigation tabs for 'Transfer', 'Ingest', 'Archival storage', 'Preservation planning', 'Access', and 'Administration'. The user is logged in as 'demo' and is connected.

Below the header, there is a form for starting a transfer. It includes a dropdown menu for 'Type' (set to 'Standard'), a text input for 'Transfer name', a file selection button labeled 'Browse' with the path '/home/demo/sampledata/Samp', and a green 'Start transfer' button.

The main content area displays a table of transfer logs. The table has columns for 'Transfer' and 'Transfer start time'. A single transfer is listed with the name 'test' and a UUID, starting on 2013-03-20 at 05:17. The transfer status is 'Completed successfully'.

Transfer	Transfer start time
test <input type="text" value="UUID"/> <ul style="list-style-type: none">Micro-service: Scan for virusesMicro-service: Extract packagesMicro-service: QuarantineMicro-service: Generate METS.xml documentMicro-service: Verify transfer checksumsMicro-service: Assign file UUIDs and checksumsMicro-service: Include default Transfer processingMCP.xmlMicro-service: Rename with transfer UUIDMicro-service: Verify transfer complianceMicro-service: Approve transfer Approve transfer [?]	2013-03-20 05:17

Kam pro další informace

- digital-preservation-cz.blogspot.cz
- code.google.com/p/kramerius/
- code.google.com/p/meta-editor/
- diybookscanner.org
- sites.google.com/site/decapodproject/
- scantailor.sourceforge.net
- archivematica.org
- opensource.knihovna.cz

**Chcete se
na něco zeptat?**

Děkuji za pozornost.

michal.denar@mail.muni.cz