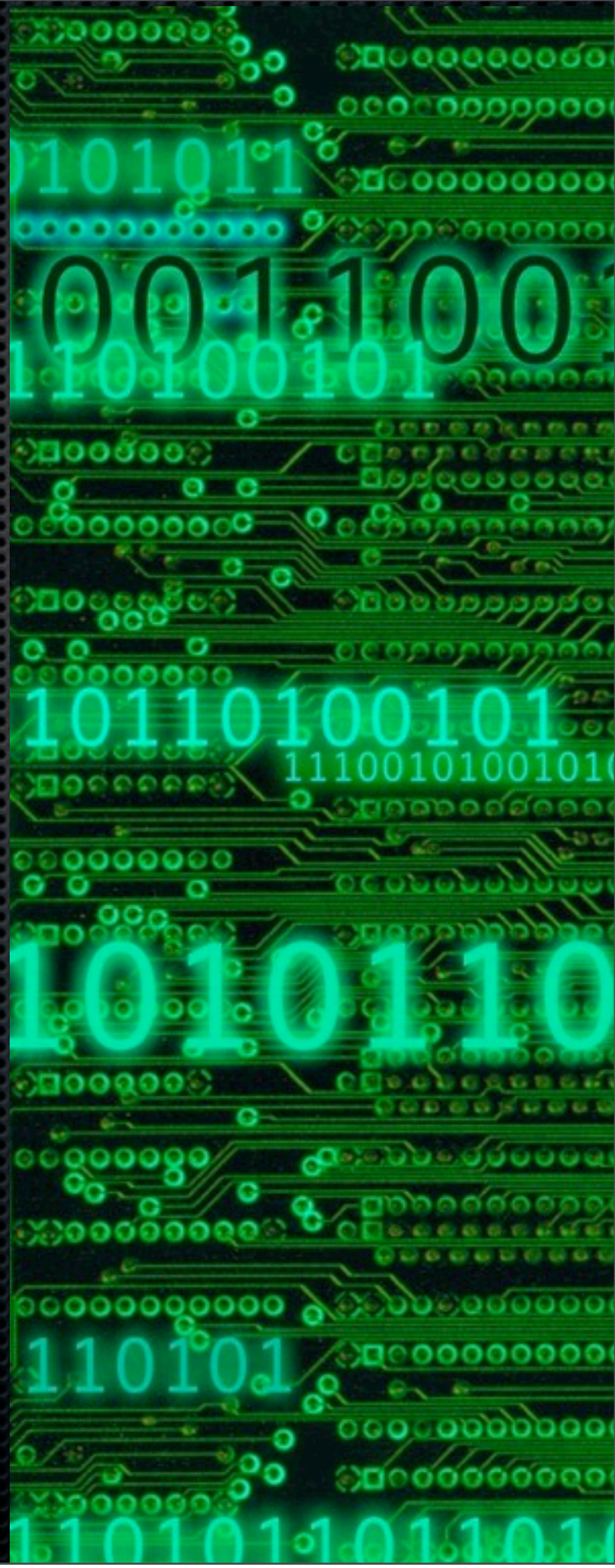
Malá digitalizace

Michal Denár

Městská knihovna Česká Třebová
KISK MU



O čem to bude?

- proč se snažíme o digiknihovnu
- 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-služeb

ILS

katalog

LTP

web

uživatel

digitální knihovna

OSOBNOSTI REGIONU

další

Požadavky

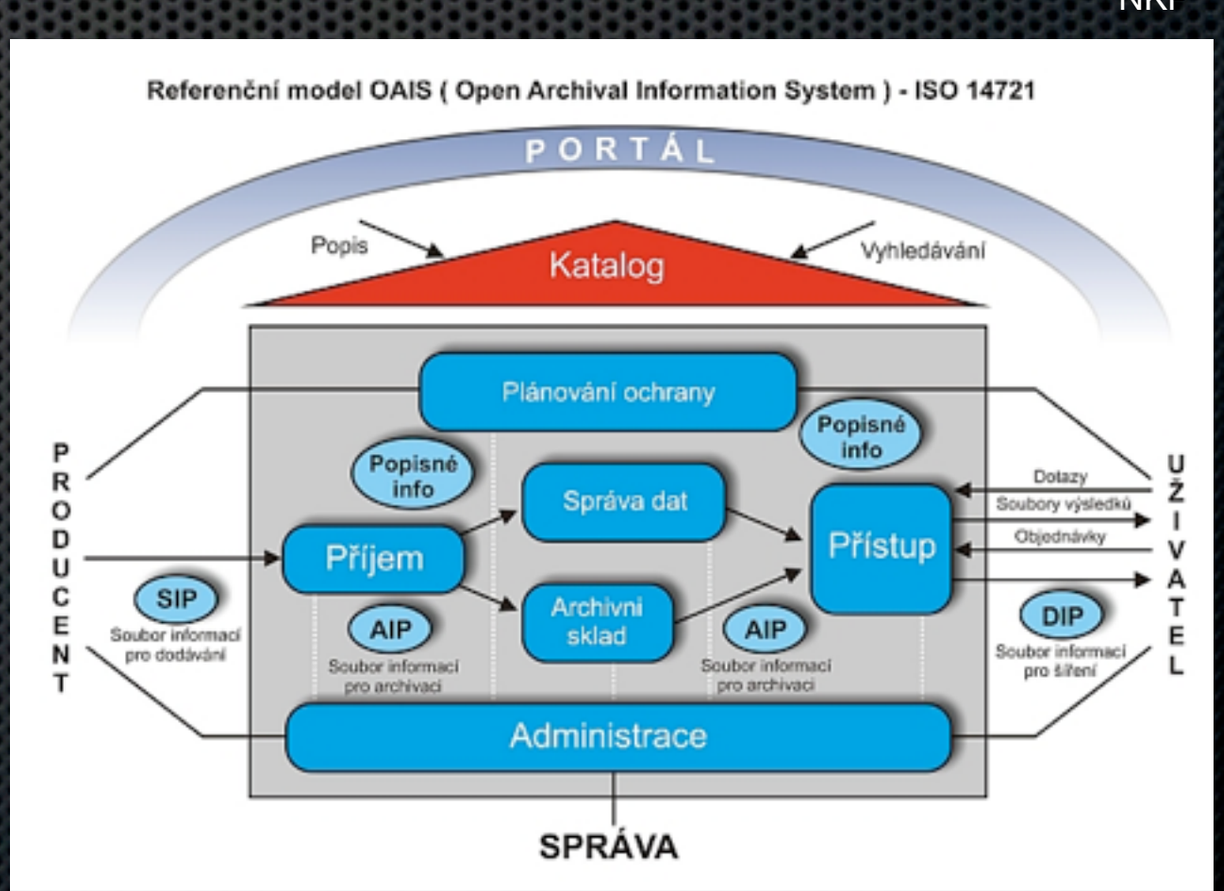
- open-source (cena, nezávislost)
- standardy NDK, možnost propojení
- metadata (ALTO XML)
- driver pro VuFind
- dostupné hw (virtualizace)

= Kramerius

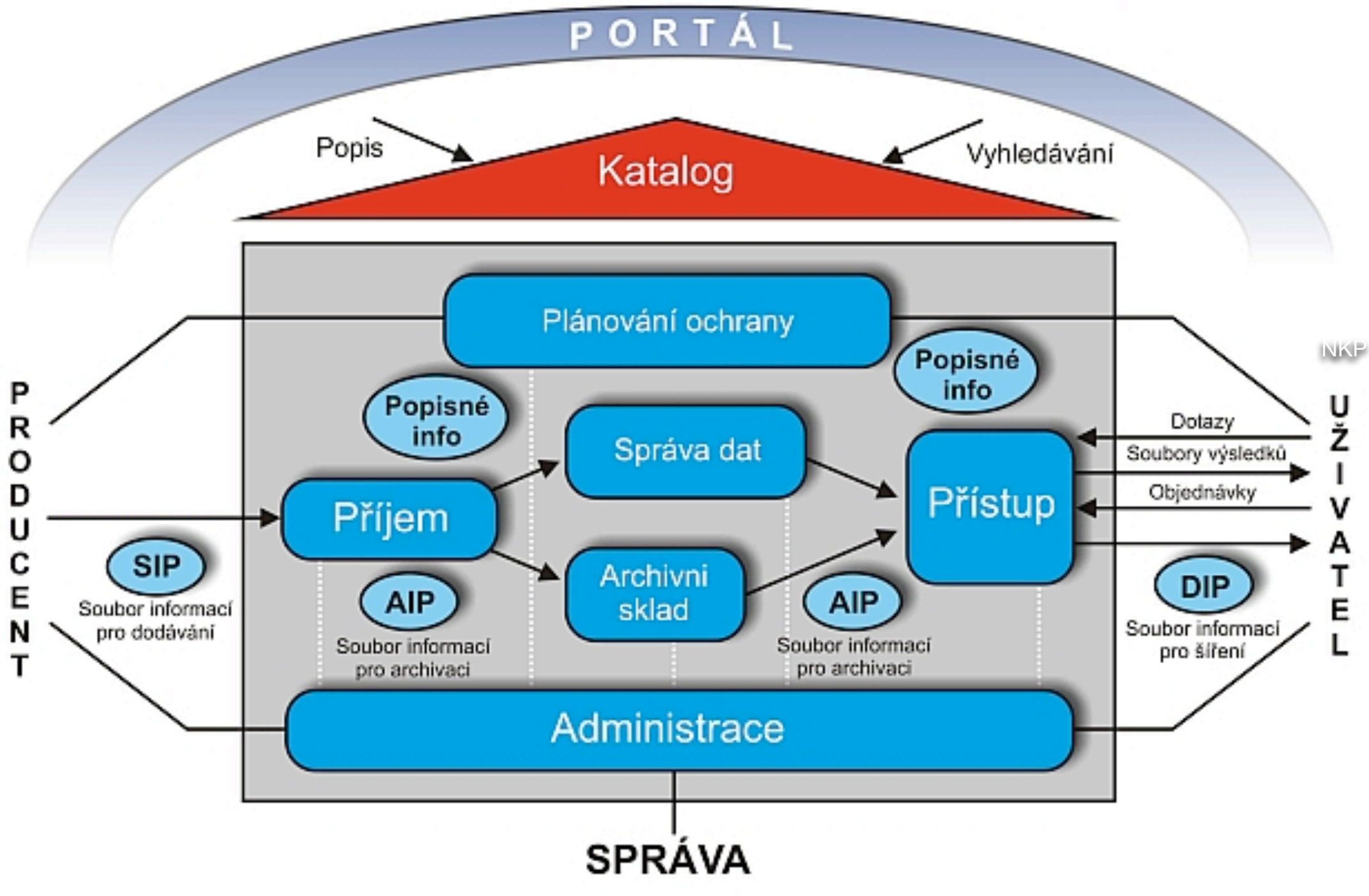
Konceptuální model OAIS

- funkční a důvěryhodný repozitář
- SIP
- AIP
- DIP

NKP



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

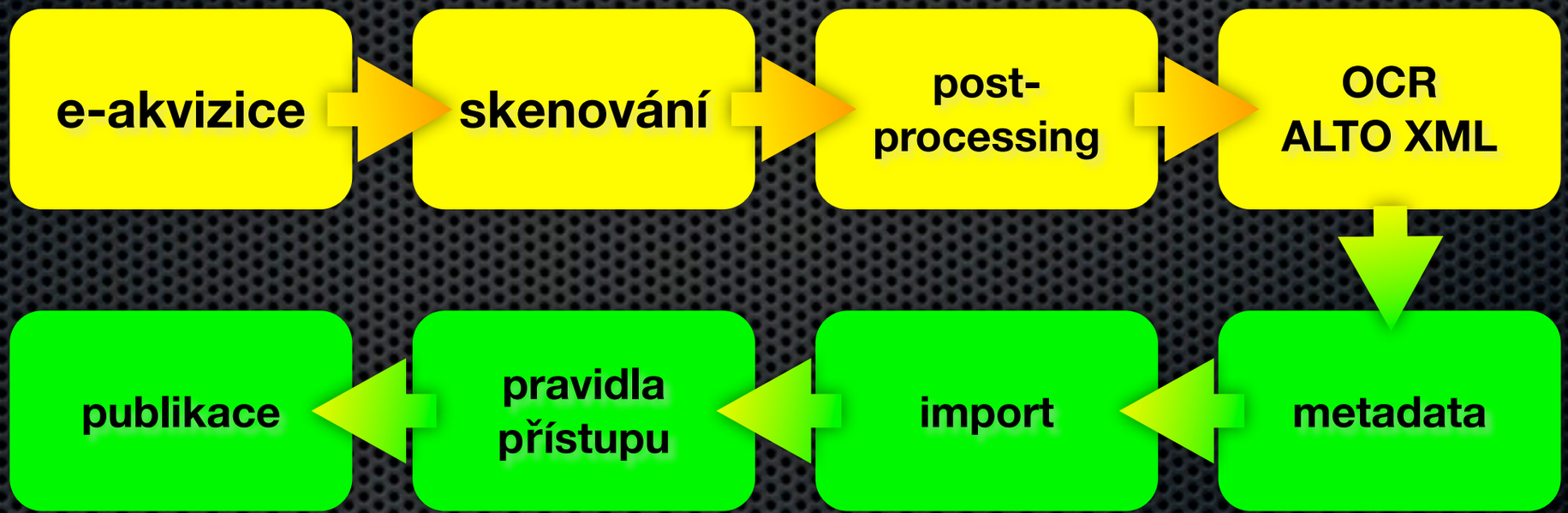


NKP

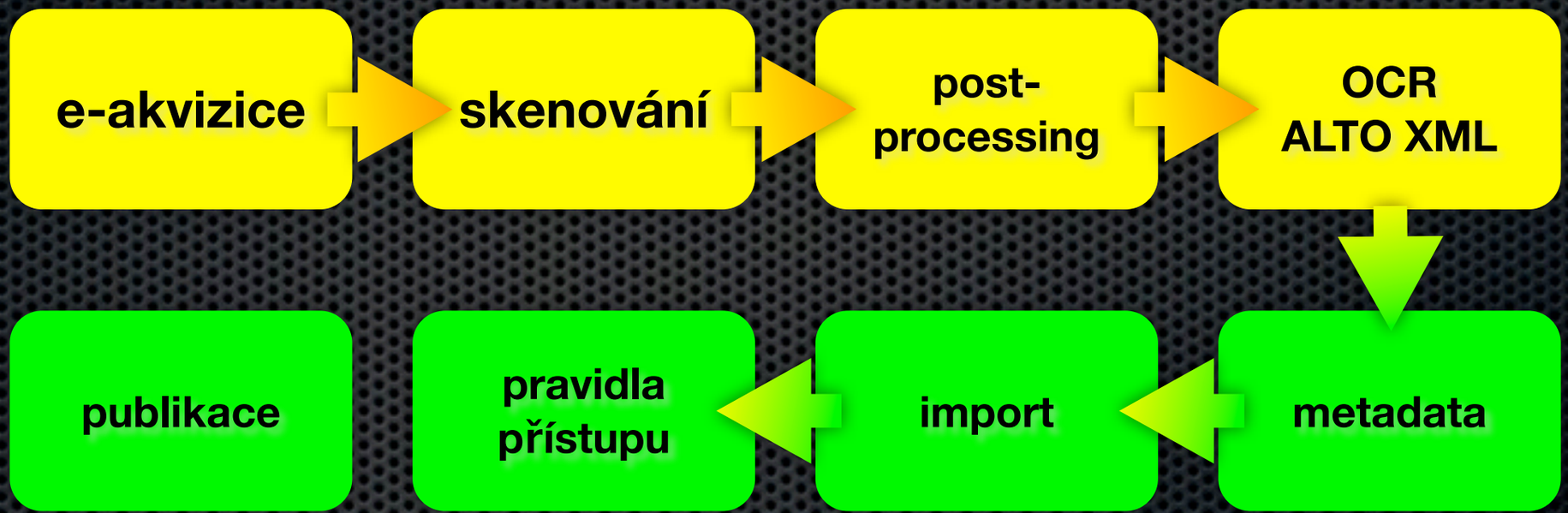
Technické řešení

Kaila		Keira
2 TB HDD		4 TB HDD
24 GB RAM		24 GB RAM
2, 66 Ghz		2, 66 Ghz
4 jádra		4 jádra
RAID		RAID
Debian		VMware
Ubuntu		Essentials

Digitalizační workflow



Digitalizační workflow



E-akvizice

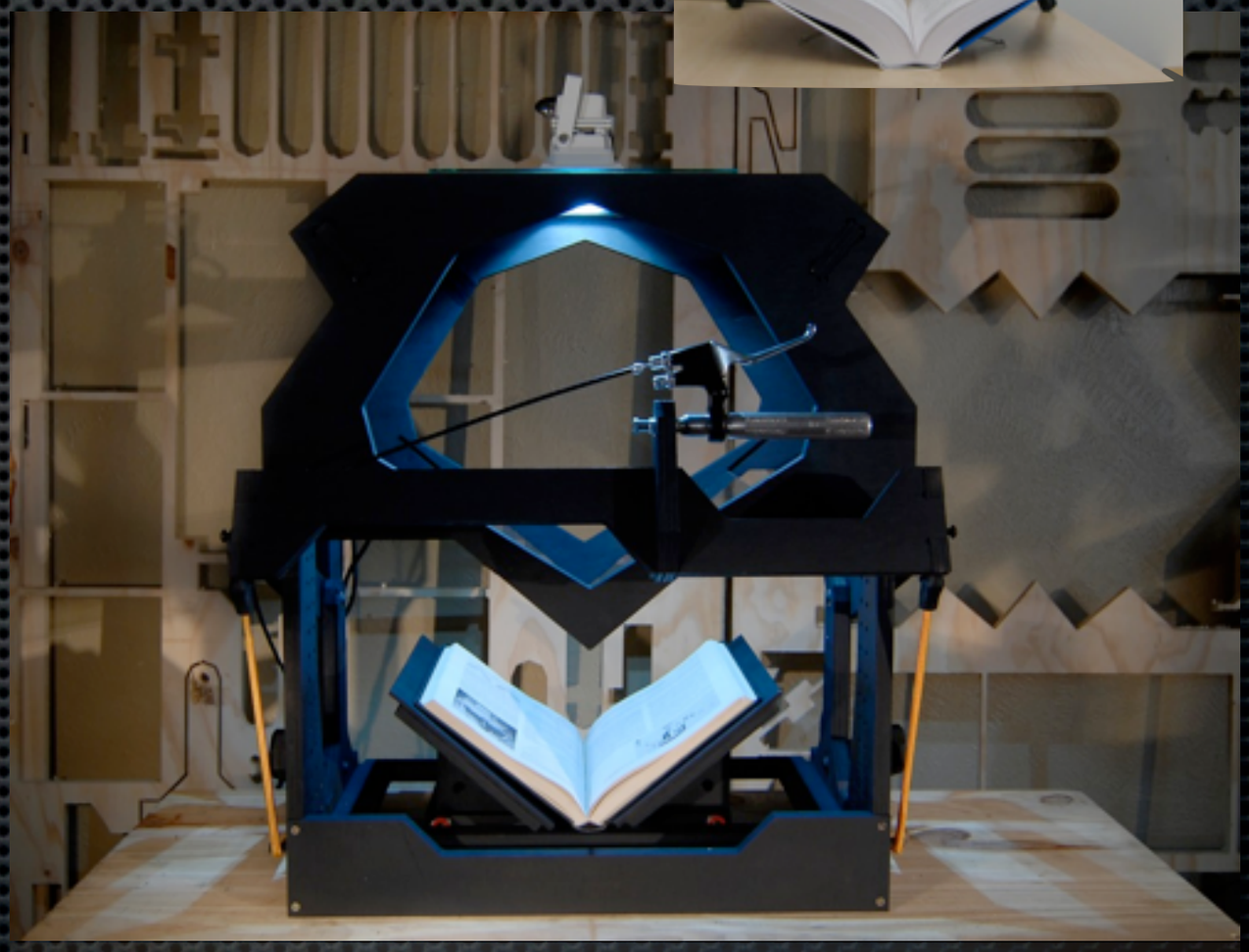
- výběr a vyhledávání dokumentů
- vyjednávání s autory a vydavateli
- nový AZ: osiřelá a nedostupná díla
- spolupráce: lidé, paměťové instituce
- born digital: vlastní produkce

Skenování

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

Video

Decapod project

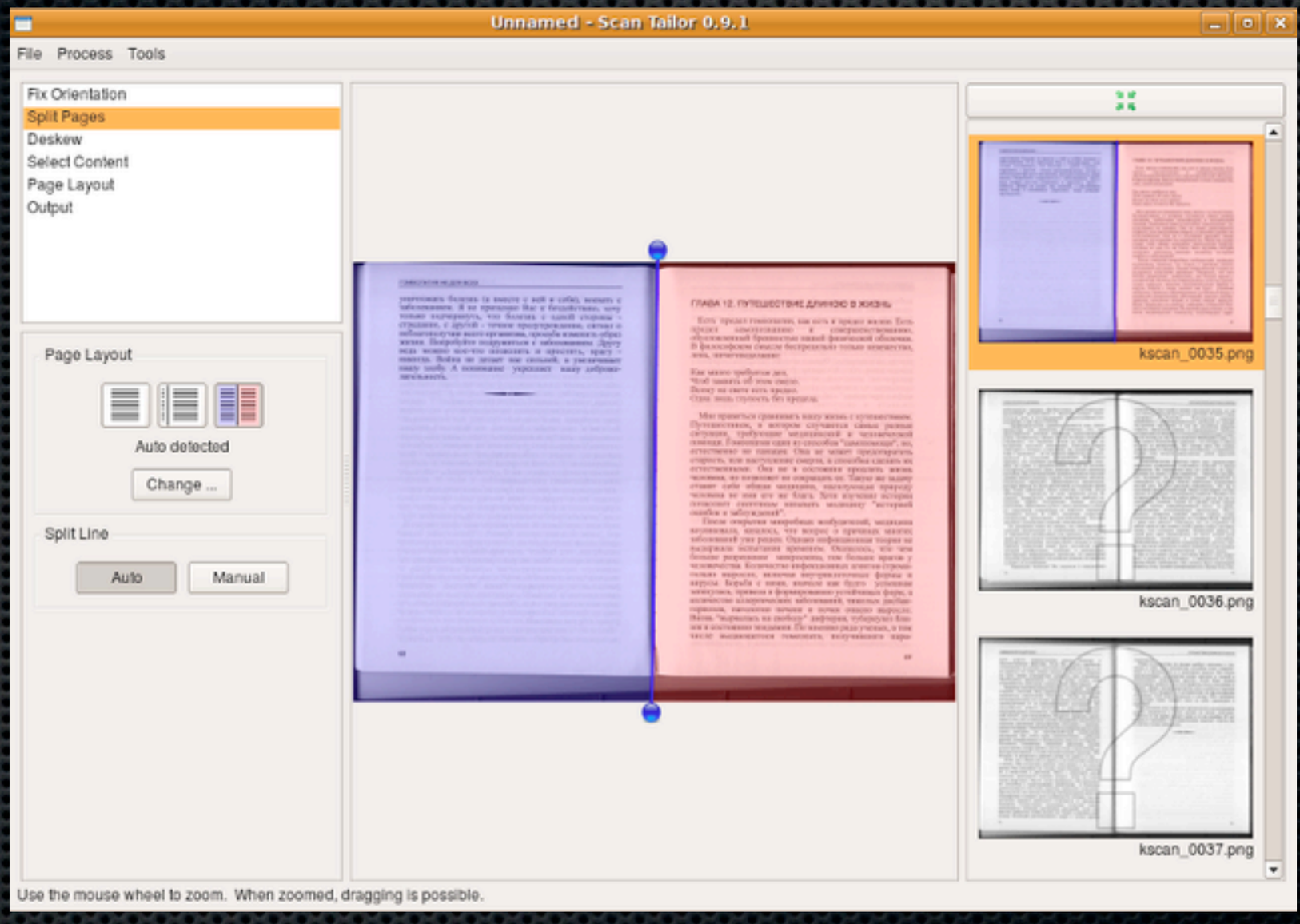


diybookscanner.org

Postprocessing

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

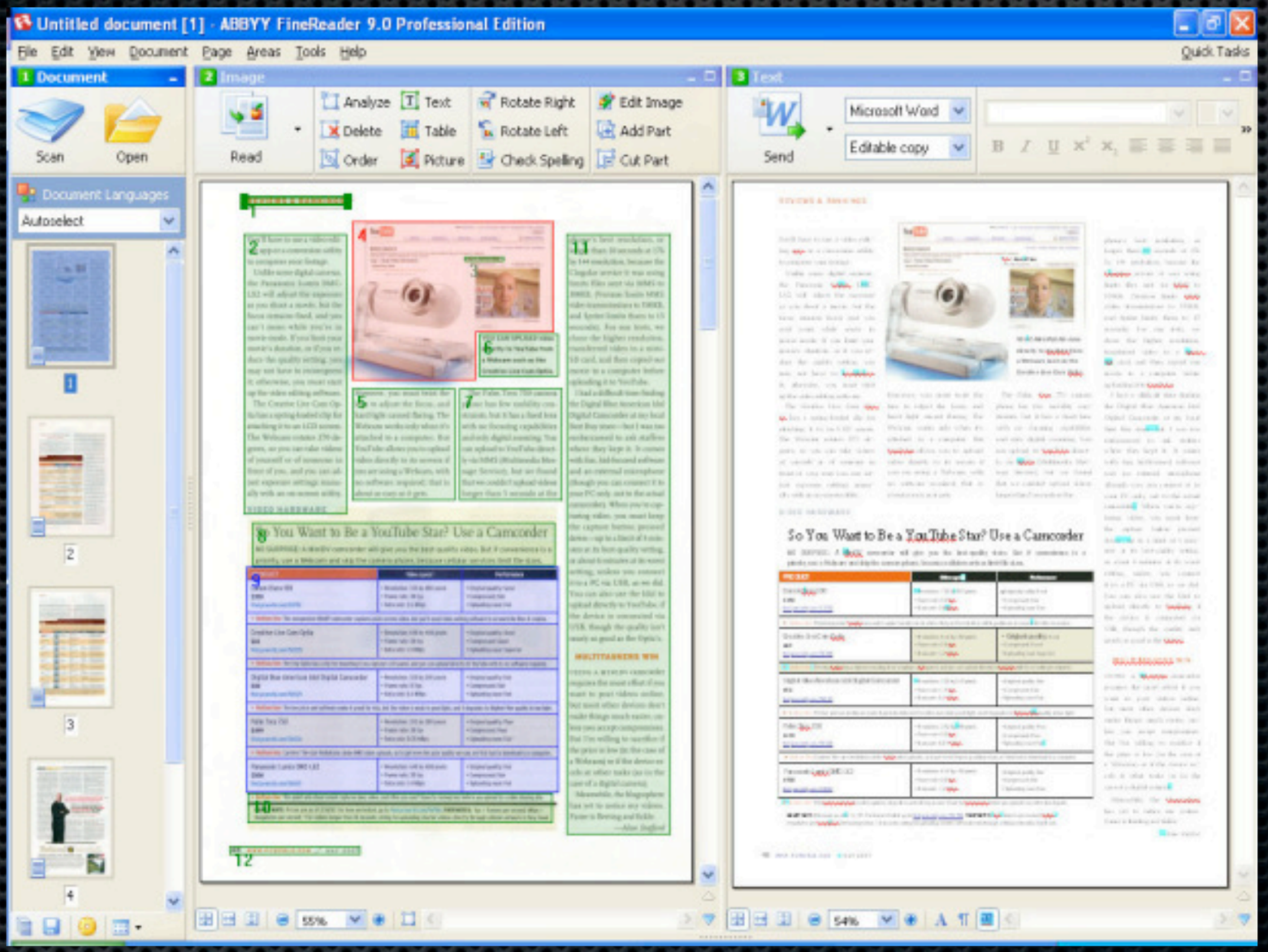
ScanTailor


Unnamed - Scan Tailor 0.9.1
File Process Tools
Fix Orientation
Split Pages
Deskew
Select Content
Page Layout
Output
Page Layout
Auto detected
Change ...
Split Line
Auto
Manual
kscan_0035.png
kscan_0036.png
kscan_0037.png
Use the mouse wheel to zoom. When zoomed, dragging is possible.

OCR

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

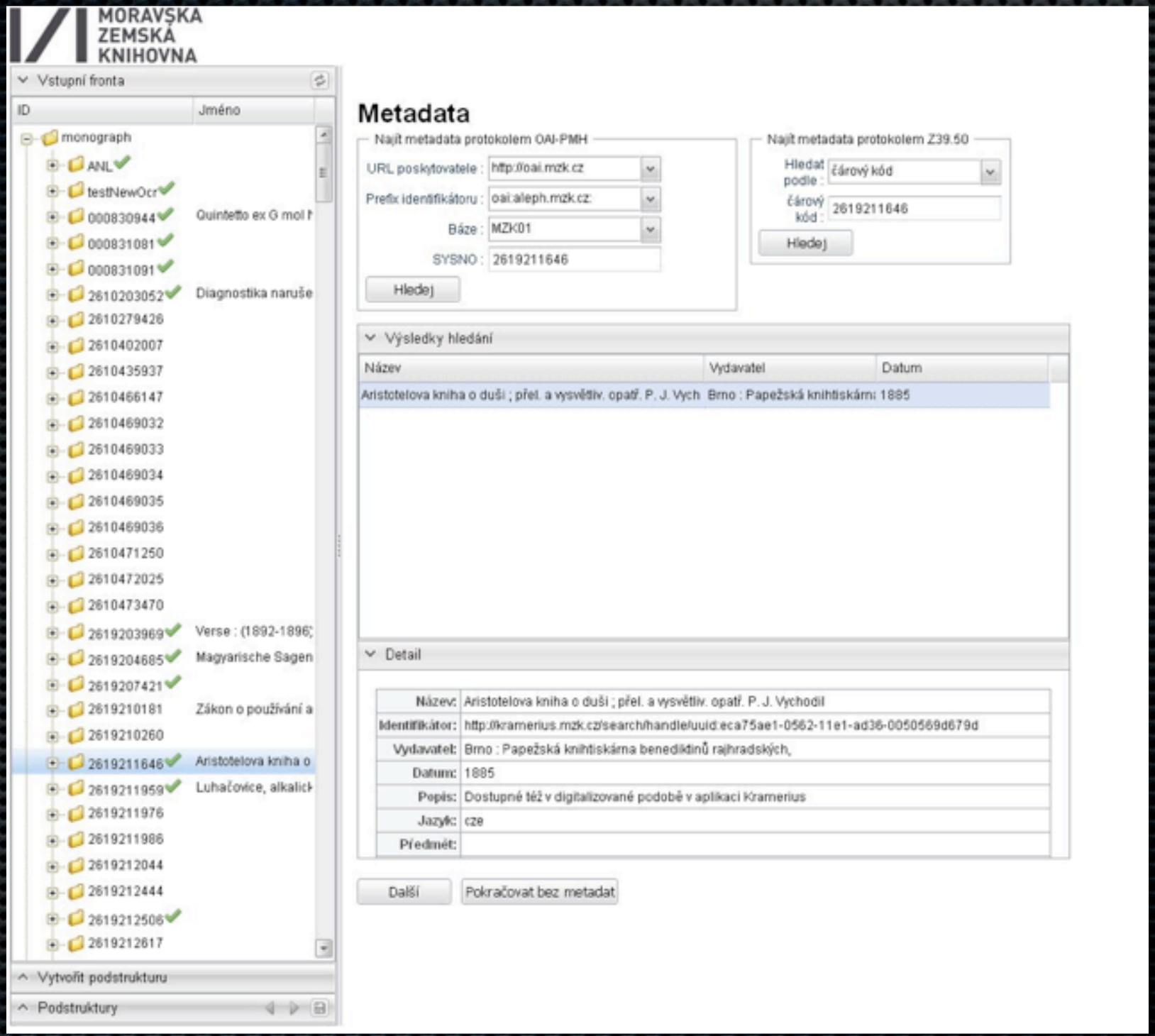
Abby FineReader



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

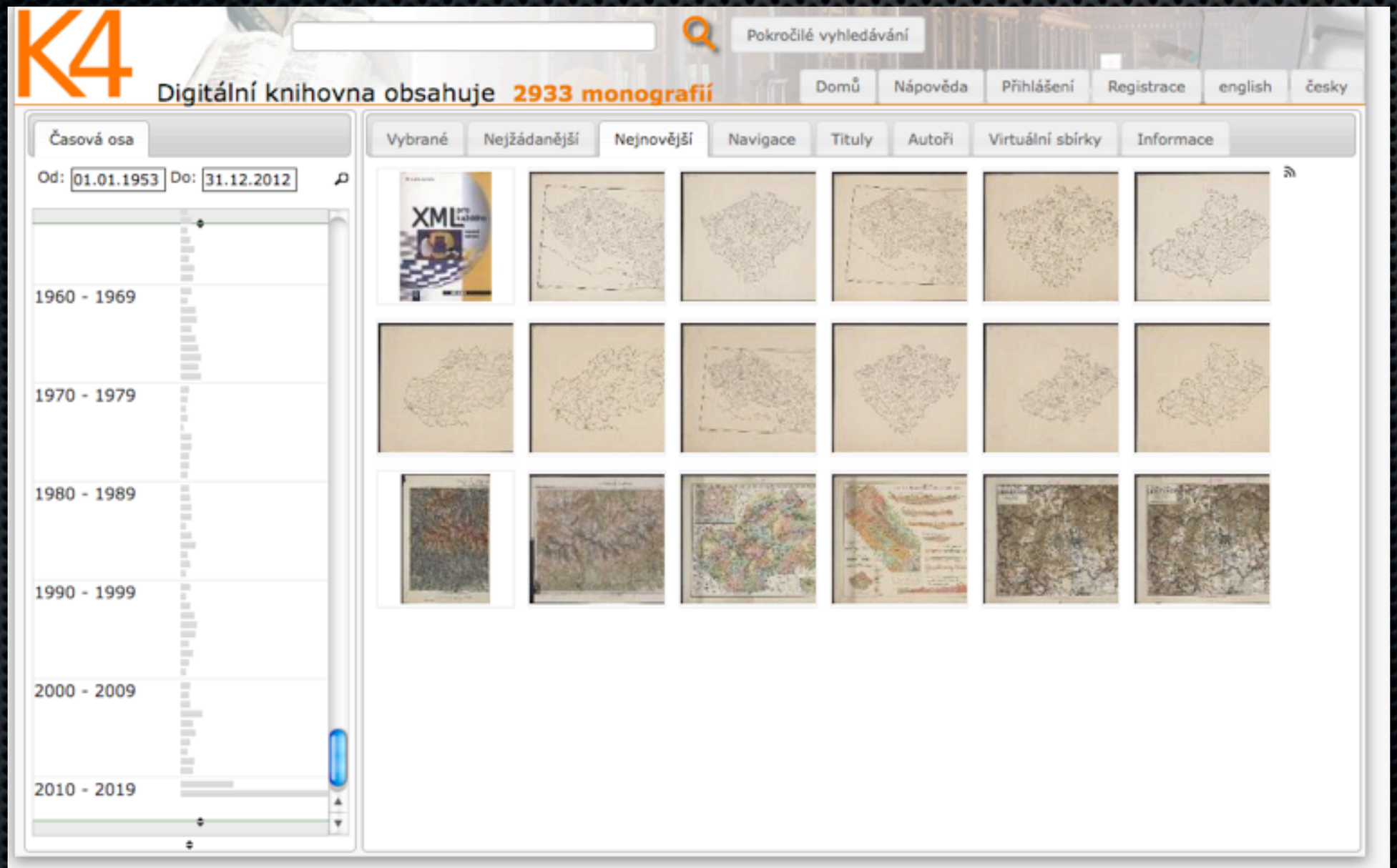
Medit


MORAVSKÁ ZEMSKÁ KNIHOVNA
Vstupní fronta
ID
Jméno
monograph
ANL
testNewOcr
000830944
Quintetto ex G mol
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 :
http://oai.mzk.cz
Prefix identifikátoru :
oai:aleph.mzk.cz:
Báze :
MZK01
SYSNO :
2619211646
Hledej
Najít metadata protokolem Z39.50
Hledat podle :
čárový kód
čárový kód :
2619211646
Hledej
Výsledky hledání
Název
Vydavatel
Datum
Aristotelova kniha o duši ; přel. a vysvětliv. opatř. P. J. Vych
Brno : Papežská knihtiskárna
1885
Detail
Název:
Aristotelova kniha o duši ; přel. a vysvětliv. opatř. P. J. Vychodil
Identifikátor:
http://kramerius.mzk.cz/search/handle/uuid:eca75ae1-0562-11e1-ad36-0050569d679d
Vydavatel:
Brno : Papežská knihtiskárna benediktinů rajhradských,
Datum:
1885
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

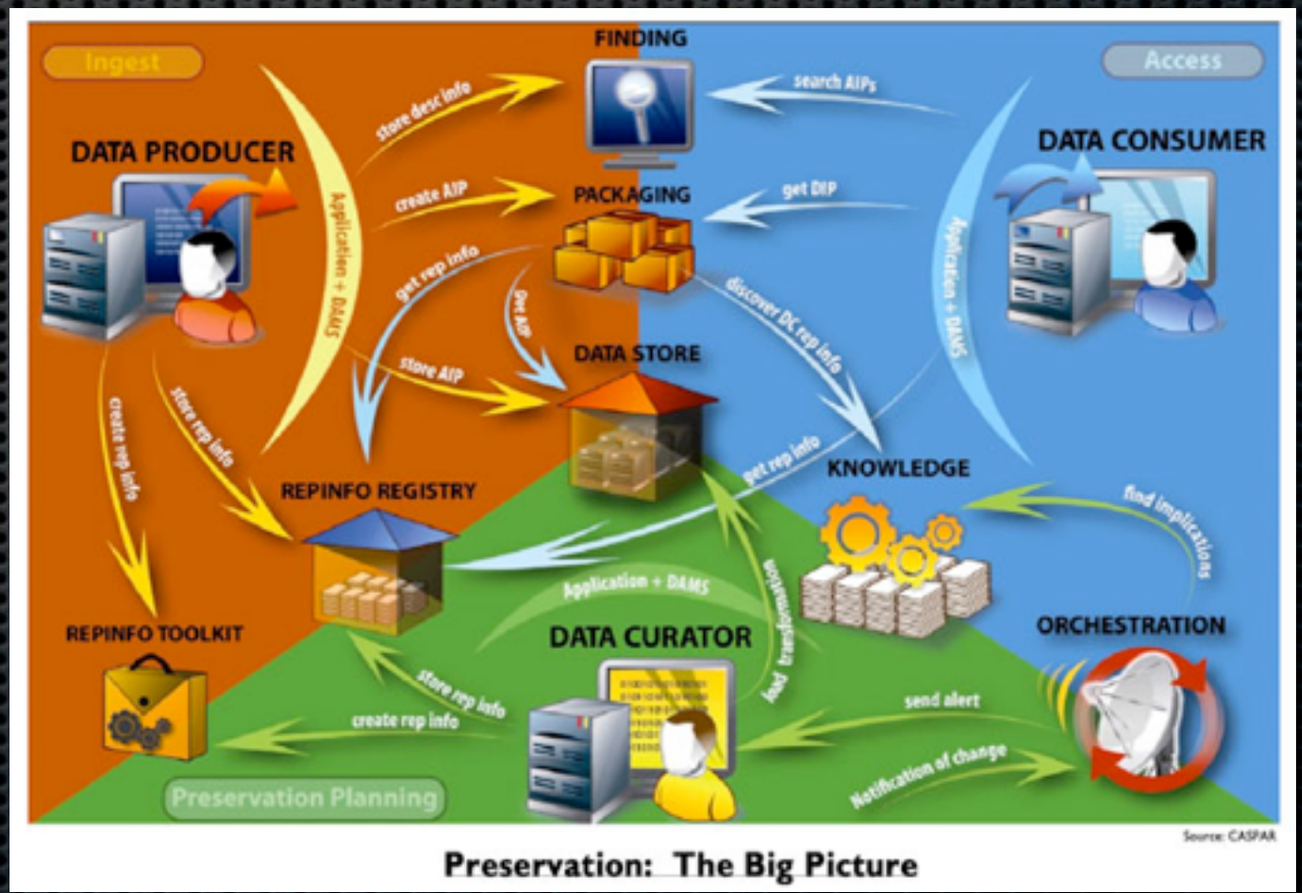
Kramerius 4


K4
Pokročilé vyhledávání
Digitální knihovna obsahuje 2933 monografií
Domů
Nápověda
Přihlášení
Registrace
english
česky
Časová osa
Od: 01.01.1953
Do: 31.12.2012
1960 - 1969
1970 - 1979
1980 - 1989
1990 - 1999
2000 - 2009
2010 - 2019
Vybrané
Nejžádanější
Nejnovější
Navigace
Tituly
Autoři
Virtuální sbírky
Informace

Long Term Preservation

- ochrana
- kompatibilita
- použitelnost
- vyhledatelnost

CASPAR

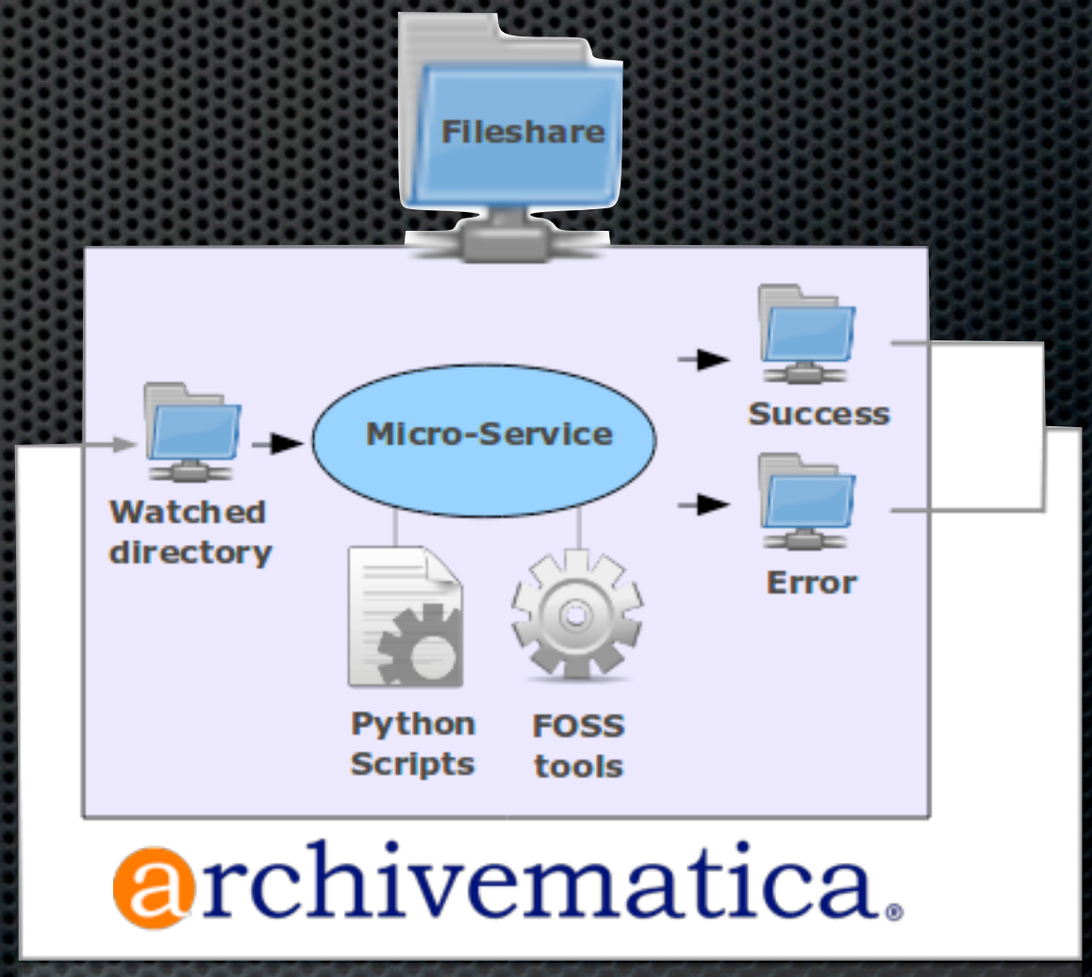


Source: CASPAR

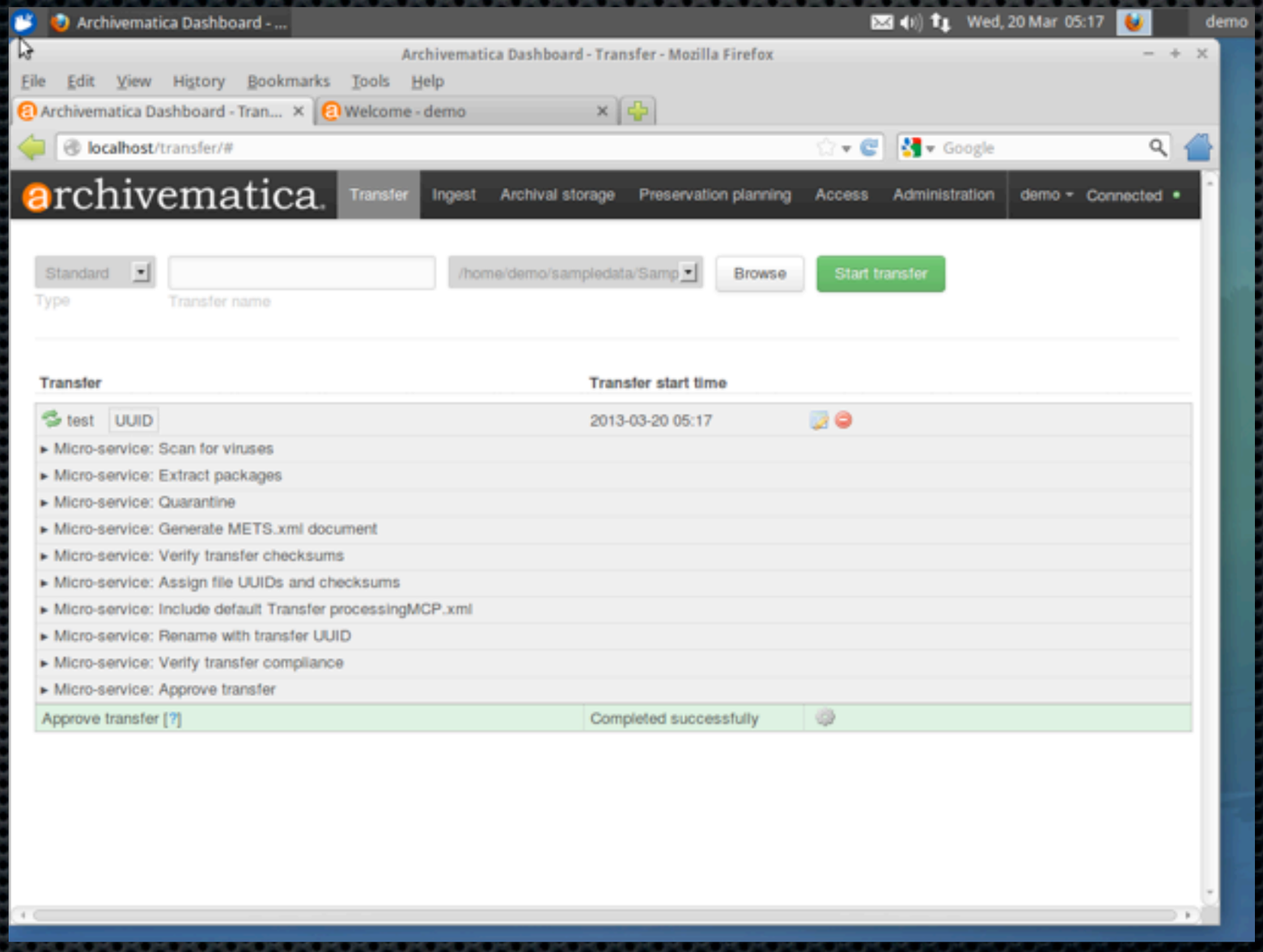
Preservation: The Big Picture

Archivematica

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



Archivematica



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