

# Národní agregátor na Slovensku

Seminár „Národní agregátor ve světě eCulture

Richard Hrabčák, Richard Bednárík

Akademie věd ČR, Praha, 14.7.2015

[richard.hrabcak@nocka.sk](mailto:richard.hrabcak@nocka.sk)

[richard.bednarik@nocka.sk](mailto:richard.bednarik@nocka.sk)

# Obsah

- Predstavenie projektu CAIR a portálu Slovakiana.sk
- Centrum pre autorské práva
- Služby národného agregátora
- Rozhrania národného agregátora
- Možnosť využitia služieb NA v CAIR pre komunikáciu PFI s Europeana



# **PROJEKT CAIR A PORTÁL SLOVAKIANA.SK**

# Projekt CAIR

- CAIR – Centrálna aplikačná infraštruktúra a registratúra
- Registre – NRKO, NRDO, NRA, NRAP, NRD
- Národný agregátor
- Portál Slovakiana.sk
- Centrum pre autorské práva



# Projekt CAIR (HIS, CDA)

- **CAIR – Centrálna Aplikačná Infraštruktúra a Registratúra**
  - Registre (NRKO, NRA, NRD, NRAP) a agregátor (aj e-born obsah)
  - Portál = prezentácia objektov a elektronické služby pre verejnosť a PFI – [www.slovakiana.sk](http://www.slovakiana.sk)
  - Centrum autorských práv
  - Integrácia s projektom CDA
- **HIS – Harmonizácia Informačných Systémov**
  - Cieľ - dostať zapojené PFI na jednotnú úroveň vybavenosti IT infraštruktúrou pre účely prepojenia s centrálnou infraštruktúrou
  - Výstupy
    - dodanie hardvéru a softvéru a uvedenie infraštruktúry do prevádzky
    - Úprava súčasných IS PFI pre účely ich integrácie, harvestovania
    - Dodanie modulu pre vytváranie a expedíciu SIP
    - Vytvorenie modulu pre riadenie inštitucionálnych vzťahov
      - riadená metodika pre projekty digitalizácie pre riešiteľov národných a dopytových projektov, riadiace a kontrolné orgány
- Projekty **CAIR, HIS a CDA** tvoria **centrálnu infraštruktúru** pre digitalizáciu kultúrneho dedičstva



# Portál Slovakiana.sk

Hlavné potreby najdôležitejších skupín užívateľov z pohľadu ich správania sa, predstáv a cieľov

- **„Známe položky“** - užívatelia vedia čo chcú, ako to vyjadriť, a kde hľadať
- **„Prieskumné“** - užívatelia majú nejakú predstavu o tom čo chcú, ale nie ako to vyjadriť, alebo kde hľadať
- **„Nevieš, čo potrebuješ vedieť“** - užívatelia môžu začať s jedným konkrétnym cieľom, ale je potrebné ho vymeniť za iný, keď sa objaví nejaký kľúčový poznatok
- **„Refinding“** - užívatelia vedia čo chcú, keď to vidia, ale nevedia nevyhnutne, ako to vyjadriť, alebo kde hľadať

SLOVAKIANA

KULTÚRNE DEDIČSTVO  
SLOVENSKA












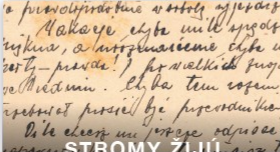


# Ukážky

KULTÚRNE OBJEKTY VIRTUÁLNE VÝSTAVY KATEGÓRIE

**SLOVAKIANA** hľadaj autora, objekt, miesto KULTÚRNE OBJEKTY VIRTUÁLNE VÝSTAVY KATEGÓRIE MENU

**SLOVAKIANA** hľadaj autora, objekt, miesto KULTÚRNE OBJEKTY VIRTUÁLNE VÝSTAVY KATEGÓRIE MENU

1 2 3 4 5 6 ZORADIŤ PODLA > SPÔSOB ZOBRAZENIA >

 <p><b>ŽENA - MÚZA</b> <i>fenomén ženy v sochárskej tvorbe</i></p>	 <p><b>STROMY ŽIJÚ</b> <i>príerez slovenskou krajinou 18. str.</i></p>	 <p><b>VECI a VECIČKY</b> <i>užitkové umenie súčasnosti</i></p>	 <p><b>AMEN</b> <i>sakrálna tvorba 15. str.</i></p>
 <p><b>ŽENA - MÚZA</b> <i>fenomén ženy v sochárskej tvorbe</i></p>	 <p><b>STROMY ŽIJÚ</b> <i>príerez slovenskou krajinou 18. str.</i></p>	 <p><b>VECI a VECIČKY</b> <i>užitkové umenie súčasnosti</i></p>	 <p><b>AMEN</b> <i>sakrálna tvorba 15. str.</i></p>
 <p><b>ŽENA - MÚZA</b></p>	 <p><b>STROMY ŽIJÚ</b></p>	 <p><b>VECI a VECIČKY</b></p>	 <p><b>AMEN</b></p>

KULT  
VIRT  
ZORN  
KULT

© 2014 Ministerstvo kultúry SR Slovákiana Mapa stránok Pomoc Kontakt Prehlásenie o prístupnosti En-

# Hlavné obavy a výhrady PFI vo vzťahu k digitalizácii

- Centrálna infraštruktúra nahradí naše lokálne systémy
- Naše dáta odovzdáme do nejakej inej inštitúcie a stratíme nad nimi kontrolu
- Ak dáme obsah na centrálny portál, nikto sa nepozrie na našu stránku a ľudia prestanú chodiť do „kamenných“ priestorov
- Ak dáme všetky dáta na internet, už nebude nikto potrebovať služby našich odborníkov





# Základné odkazy o projektoch CI pre PFI

- Vytvárame **centrálnu infraštruktúru s funkcionalitou pre normálne fungovanie digitalizácie a pripravenú aj pre e-born objekty**, ktorá bude plne dostupná **PO** skončení projektov, t.z. od cca júna 2015 – nie je to jednorázový systém počas behu projektov PO2
- Jeden z hlavných ukazovateľov celej prioritnej osi je **počet zverejnených/sprístupnených objektov** – je ukazovateľ projektu CAIR – **zodpovednosť** zaň je ale **spoločná** – nie sme nezávislé projekty
  - Ukazovateľ je nastavený na **minimálne 30%** zverejnených/sprístupnených objektov **za všetky**, nie len za niektoré **sektory**
- Vytvárame centrálny **register VŠETKÝCH kultúrnych objektov** - nie len vstupujúcich do digitalizácie
- Pre PFI sú **CAIR** a **HIS** spolu s **CDA služby CENTRÁLNEJ INFRAŠTRUKTÚRY** kam ukladajú príslušný obsah a využívajú ich služby
- **Centrálna infraštruktúra je SLUŽBA pre PFI** - nie náhrada ich systémov
- Nechceme len brať (informácie o objektoch a prezentačné deriváty), ale budeme aj dávať (služby)



Richard Bednárík

# **CENTRUM PRE AUTORSKÉ PRÁVA**

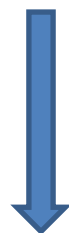
# Aktivity počas projektu

1. Vypracovanie metodík
2. Analýza aktuálneho stavu v PFI
3. Problematika vysporiadania práv
4. Problematika sprístupňovania verejnosti
5. Možnosti využitia verejných licencií
6. Ako získať súhlas autorov
7. Vypracovanie vzorových licenčných zmlúv a iných dokumentov

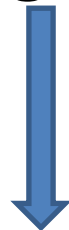


# Národný register autorských práv

Zdrojové metadáta inštitúcií



Prevodník kategorizácie – práva



Automatizované spracovanie a vyhodnotenie



# Europeana - Rights statement

Potrebné označenie práv

**Dôvod**

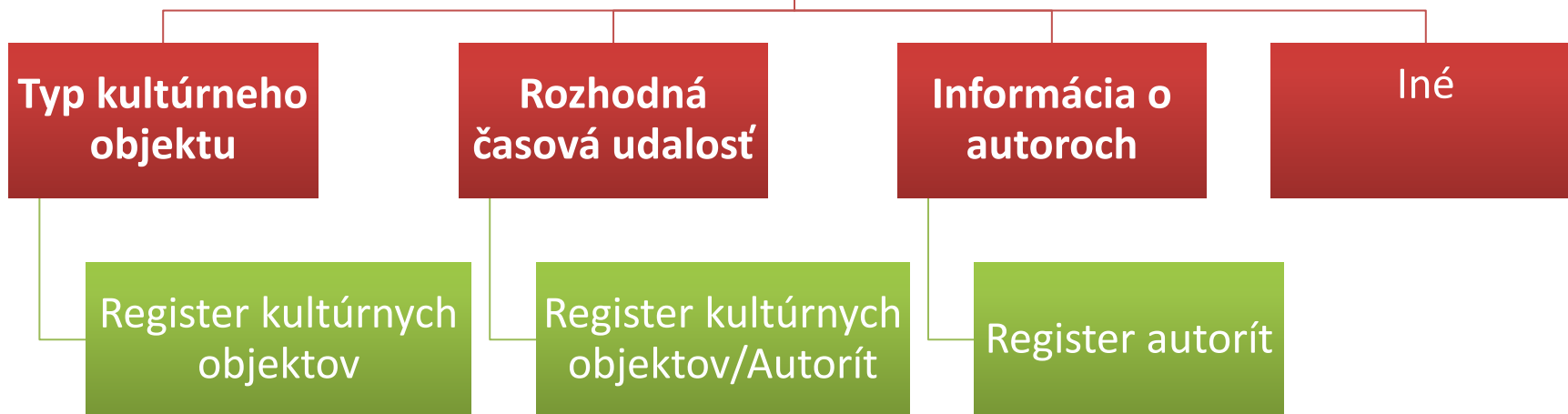


**Možnosť ďalšieho prístupnenia  
Spracovanie obsahu**



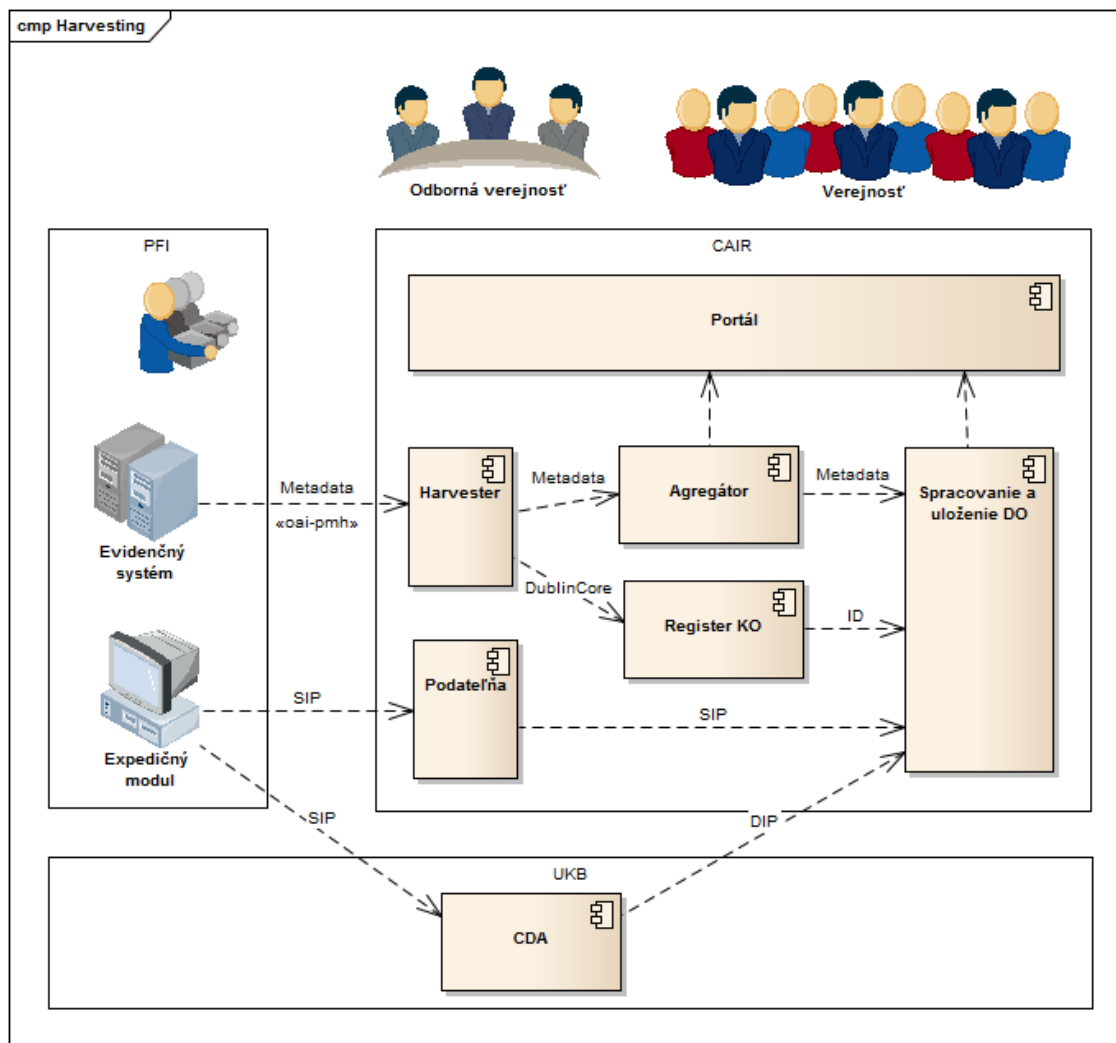
# Obsah NRAP

## ZÁZNAM O KULTÚRNOM OBJEKTE



# **SLUŽBY A ROZHRAŇIA NÁRODNÉHO AGREGÁTORA**

# Národný agregátor v kontexte IS CAIR





# Služby NA

- zber metadát - heterogénne modely aj formáty
  - validácia, transformácia a normalizácia metadát
  - spoločný dátový model pre KO - u nás sa volá CDM (Canonical Data Model) - založený na sémantickom modeli (CIDOC-CRM), pričom naše CDM (2.0) je aplikáciou CIDOC-CRM
- umožňuje vyhľadávanie a analýzu (podľa atribútu, aj fulltextovo)
- sprístupnenie - dodáva dáta pre Portál, aj cez OAI-PMH (rôzne datasety - podľa PFI, podľa typu KO, rôzne formáty - vrátane CDM, ESE3/EDM = Europeana)



# Služby NA

- kontextové služby - agregátor metadát o KO agreguje s ďalšími objektami - DO, Autoritami, autorskými právami - ktoré sa v NA navzájom linkujú (pomocou ID v rámci sémantického modelu)
- obohacovanie - nazbierané info o KO sa obohacia o
  - odvodené informácie (napr. geo koordináty pre slovné uvedené miesta)
  - linkovanie na externé zdroje (napr. wikipedia, ak existuje, a pod)
  - prezentačné informácie (napr. podobizeň autora, obálka knihy)
  - rozpoznanie referencie na authority (negarantovane väzby - automatické dohľadávanie podľa kontextu a napr. záhlavia)
  - dáta z obohatenia sa ukladajú zvlášť, pri prezentácii sú odlišené, aby sa vedelo, čo je z akého zdroja)



# Rozhrania

- rozhranie OAI-PMH 2.0
- formát metadát (okrem iných, napr. DublinCore a ESE3) aj RDF
  - metadata o KO v CIDOC-CRM scheme
  - <http://purl.org/NET/cidoc-crm/core#>
  - obsahuje urn:nbn ako identifikatory entity (KO, DO, Autorita)
  - obsahuje url na pôvodný kontext
  - číselníky v SKOS reprezentácii
  - <http://www.w3.org/2004/02/skos/core#>
- dataseť sú organizované podľa typu aj podľa zdroja
- V kontexte Open Data je z tohto pohľadu CAIR Linked Data ready



# Oblasti, ktoré sme riešili

- Formáty nie sú problém, tým sú rôzne koncepty a referenčné modely pre popis tej istej reality
- rôzne metodiky použité na katalogizáciu aj jedného typu KO v rámci jednej PFI
- množstvo lokálnych číselníkov / thesaurów, ktoré sa vyskytujú v popise KO
- nepoužívanie explicitných číselníkov / thesaurów v iných KO, t.j. použitie textových polí bez jasnej (explicitnej) sémantiky
- osobné dáta, utajované / chránené / citlivé údaje, ktoré sa vyskytujú v metadátach a nie sú explicitne / automaticky rozlíšiteľné

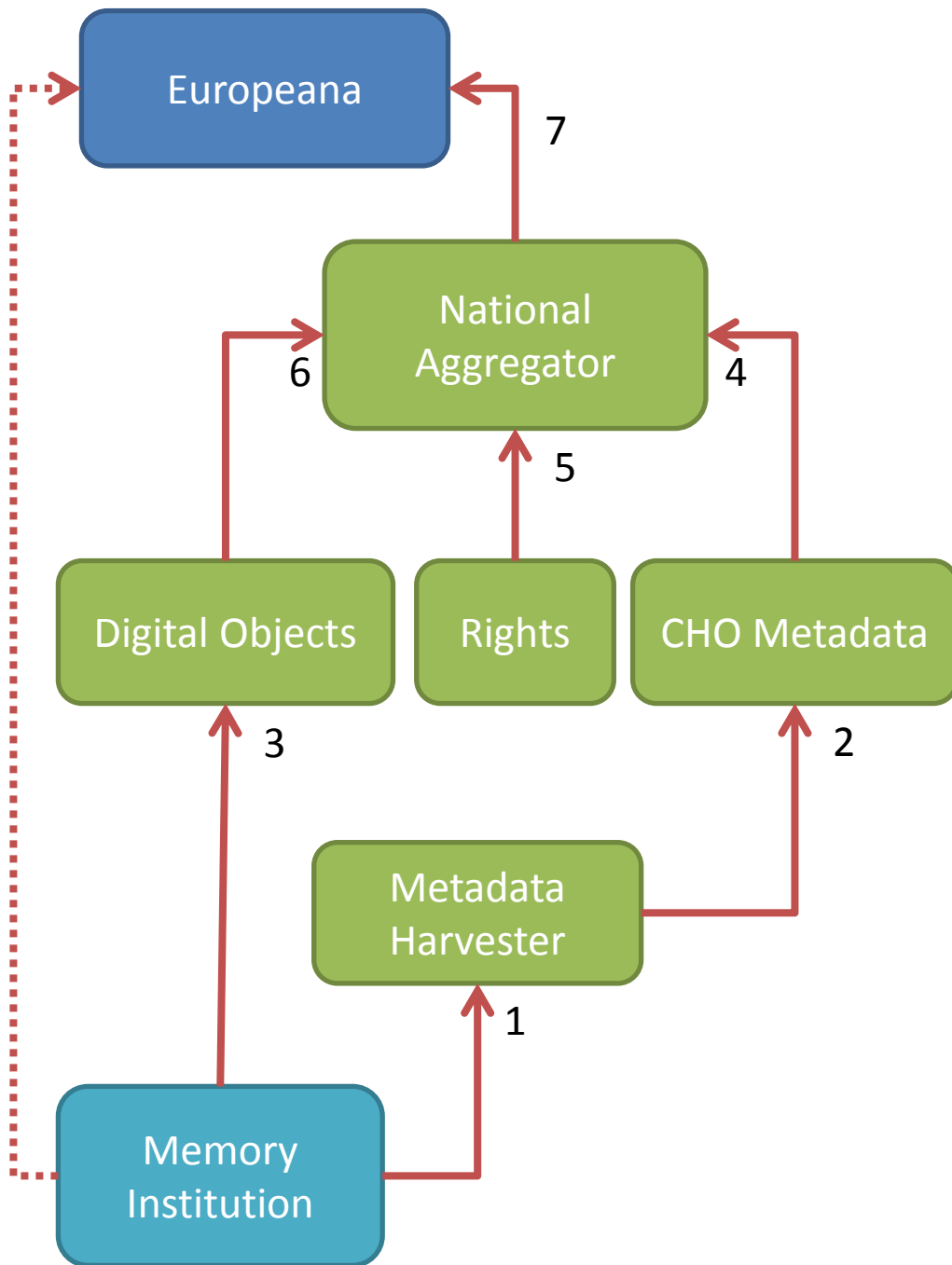


**SLOVAKIANA VS EUROPEANA**

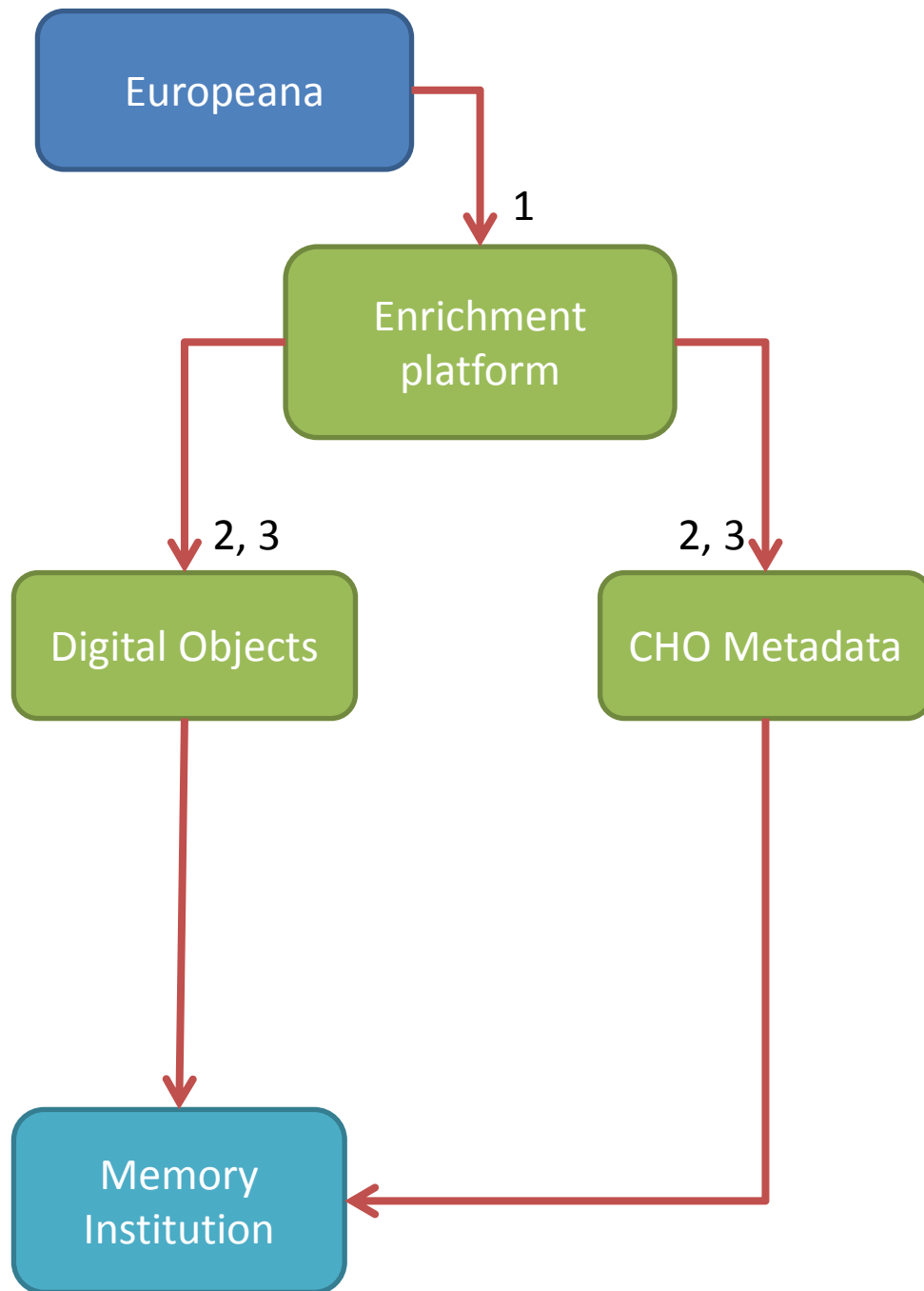
# Slovakiana a Europeana

- Europeana preferuje dodávanie dát na báze národných agregátorov, túto úlohu bude plniť projekt CAIR resp. platforma Slovakiana.sk (Cross Domain NA)
- Jednotlivé inštitúcie zostanú v roli data providers
- Zjednotenie a zjednodušenie komunikácie a nutných technických a organizačných požiadavaviek voči Europeana
- Do Europeana bude CAIR dodávať len kvalitné dáta (rozsah si individuálne určí každá PFI), pri metadátach s CC0 licenciou





1. metadata validation & transformation
2. metadata mapping & normalization
3. ingest, data processing
4. metadata quality evaluation
5. rights evaluation
6. data quality evaluation
7. **metadata publishing:**
  1. EDM format
  2. valid rights
  3. metadata quality level > X%
  4. additional validation rules (schematron)



1. metadata transformation & mapping
2. enriched metadata
  1. (Europeana) publication status
  2. links (to Europeana)
3. new objects
  1. (Europeana) object discovery
  2. (Europeana) context discovery



# Hlavné benefity architektúry CAIR / Slovakiana

- Publikovanie voľného aj autorsky chráneného obsahu (zdarma aj za poplatok)
- Plne kompatibilný dátový model s EDM
- PFI môžu publikovať a získavať dáta z Europeana prostredníctvom Slovakiana
- Používatelia môžu vytvárať kolekcie nielen z obsahu v Slovakiana, ale aj z Europeana (alebo iných dátových zdrojov) na jednom mieste
- Automatizácia agregáčného procesu – pre všetok obsah v Slovakiana vieme garantovať definované úrovne kvality, autorských práv, prezentačných formátov a odkazov



**JE DIGITALIZÁCIA NA NIEČO  
DOBRÁ?**

# Čo sme sa naučili

- Dôležití sú ľudia, ak sa postaráte o ich trvalý odborný rast, zlepšíte renomé inštitúcie
- Centralizovaných služieb sa netreba báť, nenahrádzajú, ale slúžia
- Dodávaný obsah musí mať svojho pána zodpovedného za kvantitu, ale najmä za kvalitu
- Neodkladajme napĺňanie informácií o objektoch na „potom“
- S autormi a ich zástupcami sa dá dohodnúť na sprístupňovaní obsahu
- Každému partnerovi za dodaný obsah treba na oplátku poskytnúť aj prínosné služby (odborné aj financie prinášajúce)
- Odborníci z rôznych inštitúcií sa kolektívne vedia dohodnúť a spolupracovať
- Na projekty digitalizácie sa treba pozerať ako na program, ktorý musí mať svojho „sponzora“ na centrálnej úrovni, ktorý riadi a aktívne dohliada
- Na prezentáciu obsahu treba použiť všetky dostupné platformy a dodávať na ne obsahovo kvalitný a zaujímavo popísaný obsah
- Digitálny obsah môže prinášať ďalšie finančné zdroje



# Ako na to...

- **Urobiť** aj **menšie** a **postupné kroky** vedúce k využívaniu nových služieb a foriem prezentovania kultúrneho dedičstva
- **Vysvetľovať prínosy** a **možnosti** vyplývajúce z digitalizácie v okruhu svojich známych, kolegov a zákazníkov, vžiť sa do ich pohľadu na digitalizáciu a čo by od nej očakávali
- **Vyskúšať** si **funkcionalitu** a **možnosti** na **menšom množstve** objektov a **ukázať, že sa to dá** (a nie, že sa to nedá, lebo ...)

