



WIDaT
2019

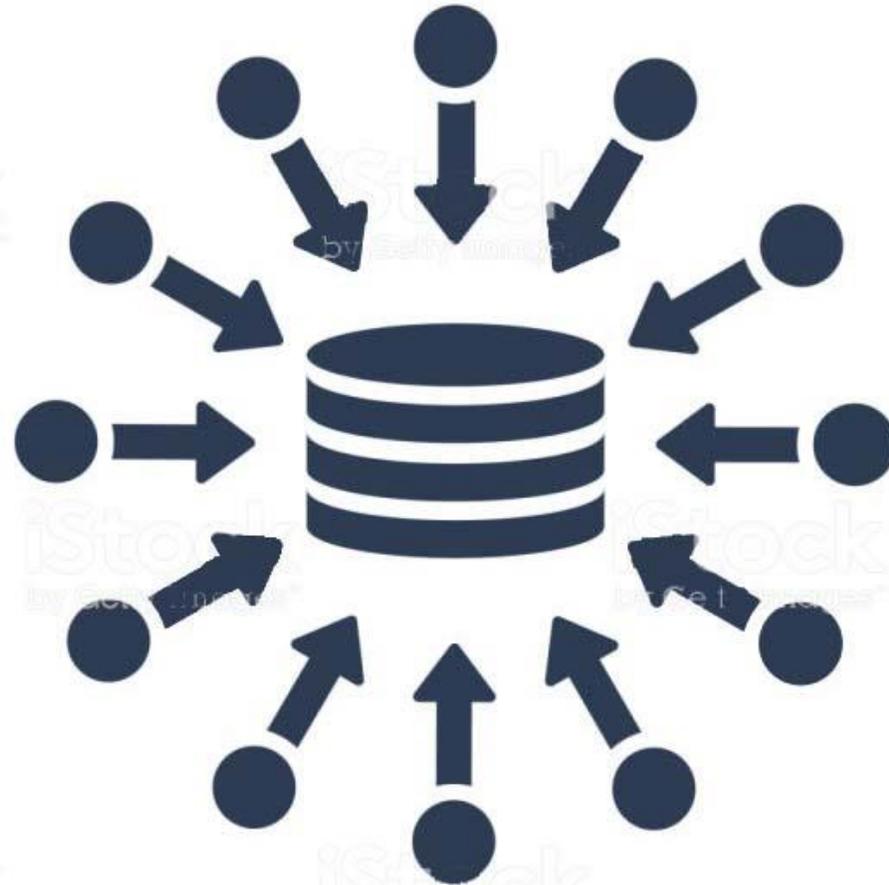
WORKFLOW DE AGREGAÇÃO DE DADOS: Processos para Criação de uma Interface de Busca Integrada do Patrimônio Cultural

Joyce Siqueira

Dalton Lopes Martins



AGREGAÇÃO DE DADOS



OBJETIVOS

Localizar e analisar workflows de agregação de dados, desenvolvidos por instituições culturais, a fim de identificar as principais etapas propostas.

METODOLOGIA

Pesquisa descritiva e bibliográfica, de natureza qualitativa, realizada em bases de dados acadêmicas e na literatura cinzenta.

As buscas foram realizadas no Google, Google Acadêmico, EBSCOhost e BRAPCI, utilizando os termos: *“pipeline”, “workflow”, “architecture”, “aggregation”, “metadata ingest”, “metadata aggregation”, “europeana”, “mexicana”, “dpla”, “digital public library of america”, “trove”, “digitalnz”, “aggregative data infrastructures”*.

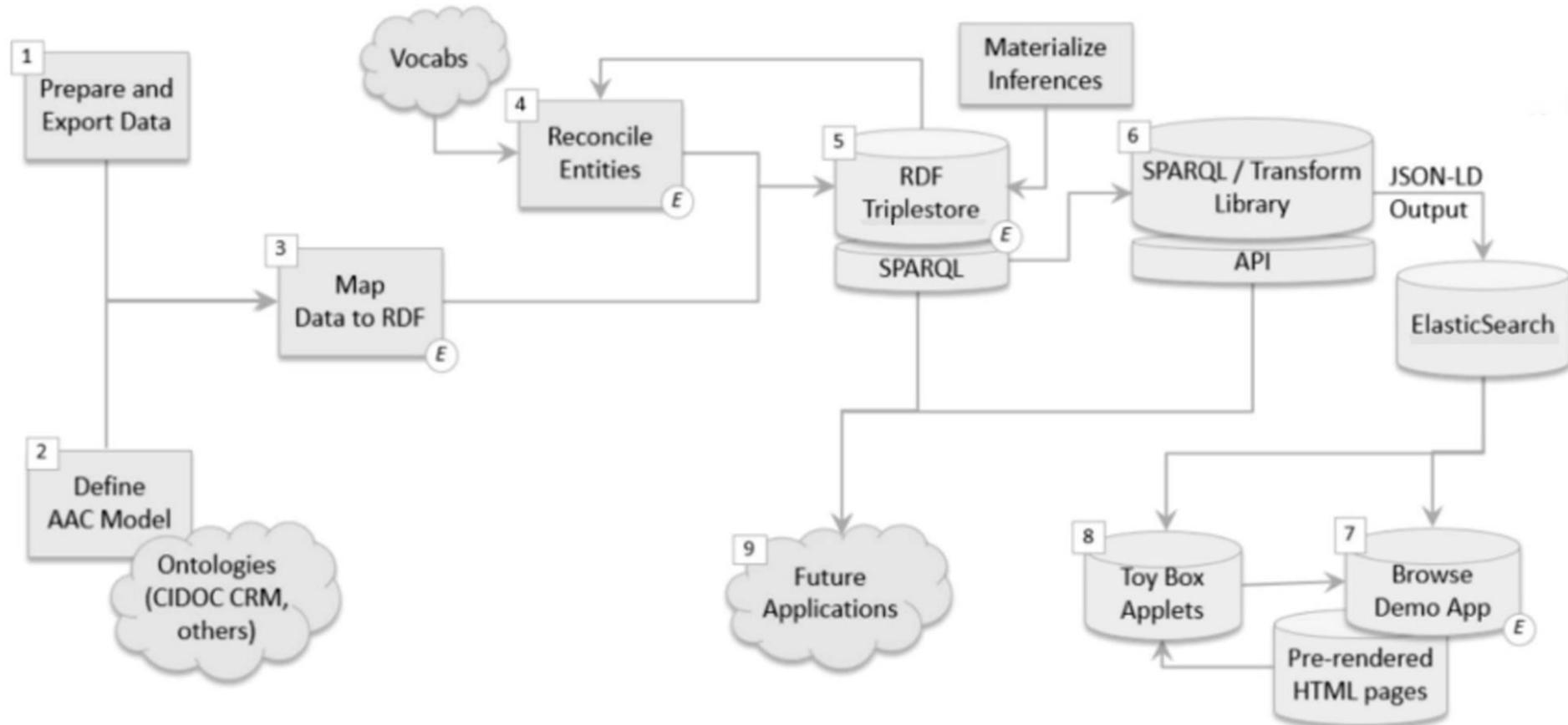
RESULTADOS

Foram encontrados sete workflows:

- 1. *American Art Collaborative***
- 2. Secretaria de Cultura do México - Repositório Mexicana**
- 3. Universidade de Nevada - *UNLV's Linked Data Project***
- 4. Europeana**
- 5. Biblioteca Nacional da Austrália - Trove**
- 6. Istituto di Scienza e Tecnologie dell'Informazione – DNET**
- 7. Biblioteca Nacional da Nova Zelândia - DigitalNZ**

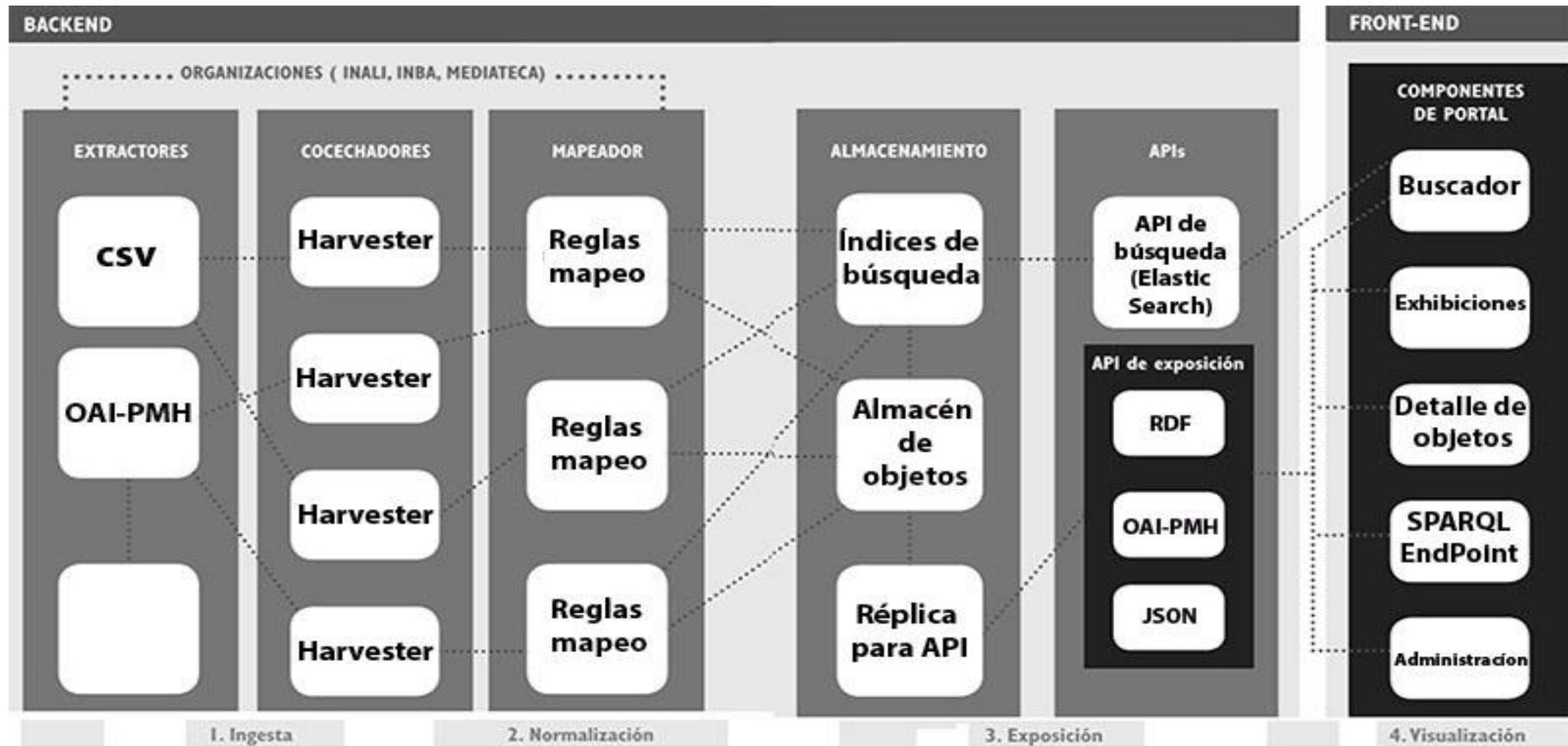
RESULTADOS

A *American Art Collaborative* é um consórcio de 14 instituições de arte, nos Estados Unidos, que visam investigar e começar a construir uma massa crítica de *Linked Open Data* – LOD.



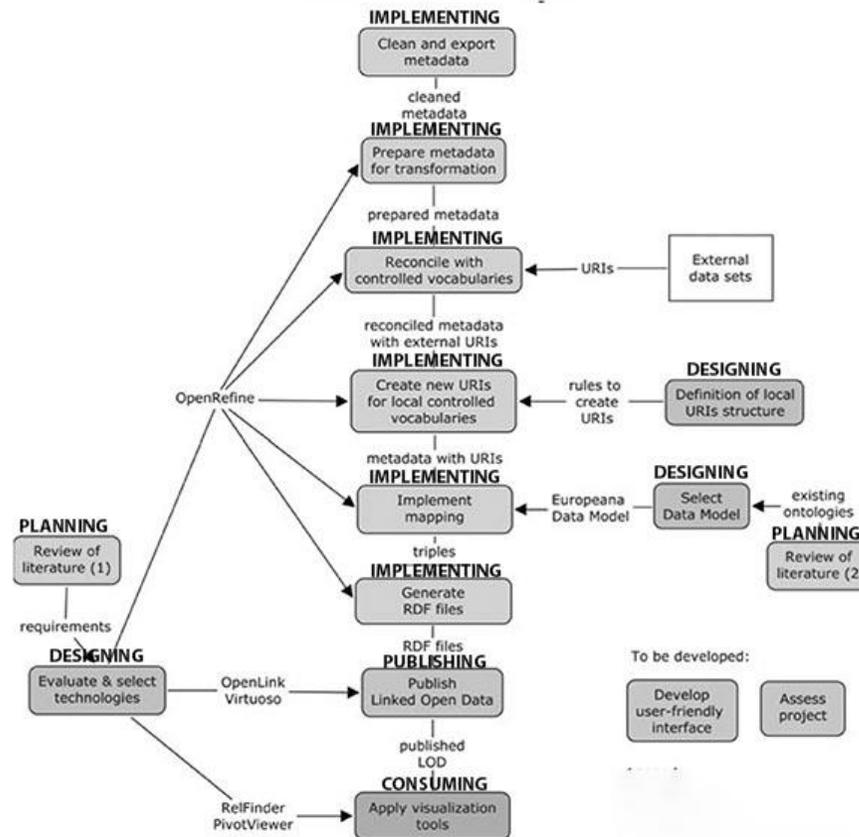
RESULTADOS

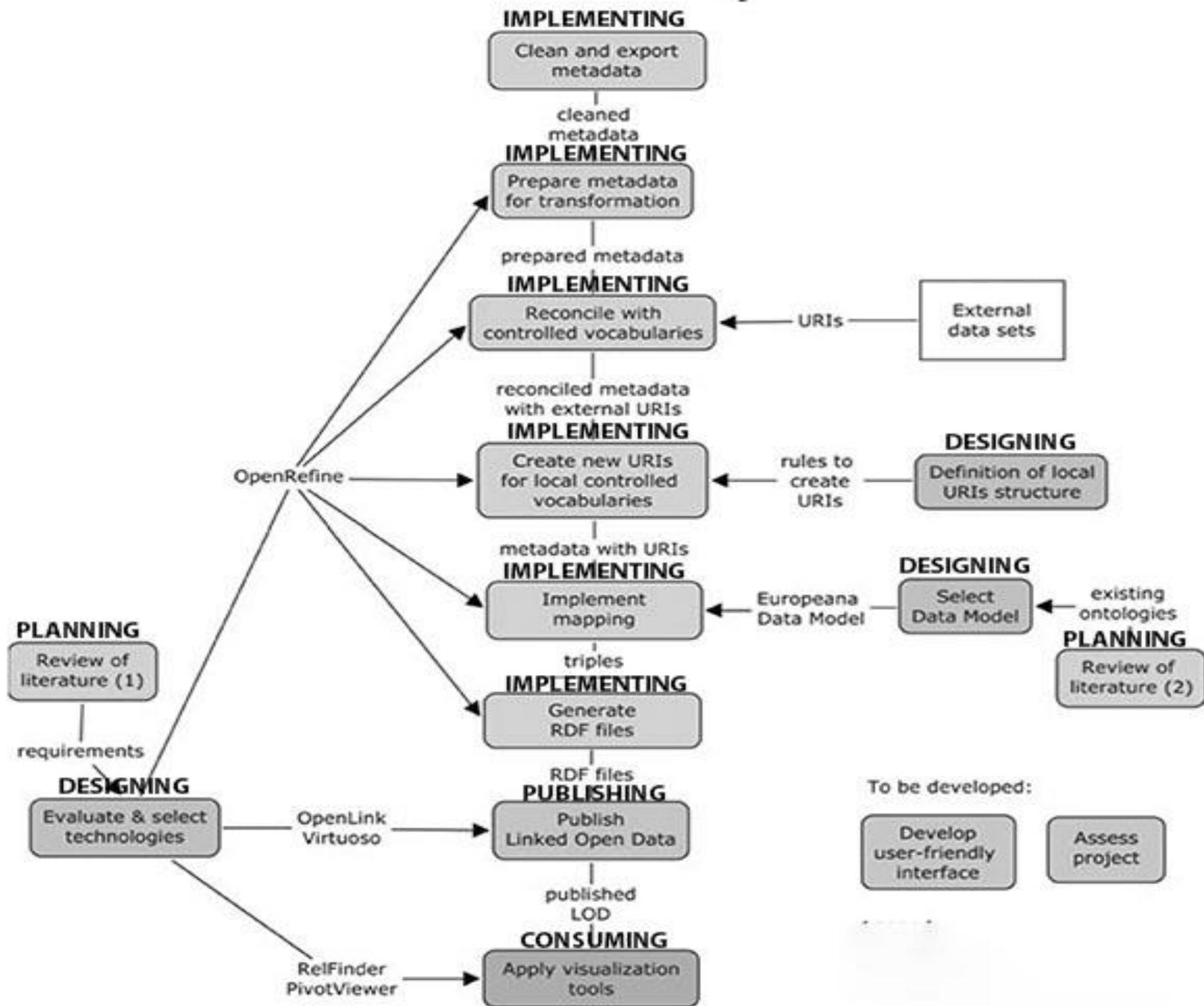
A **Secretaría de Cultura do México** desenvolveu a Mexicana, um Repositório do Patrimônio Cultural do México, livre e aberto, que tem o objetivo principal de difundir e vincular os acervos do patrimônio cultural do México.



RESULTADOS

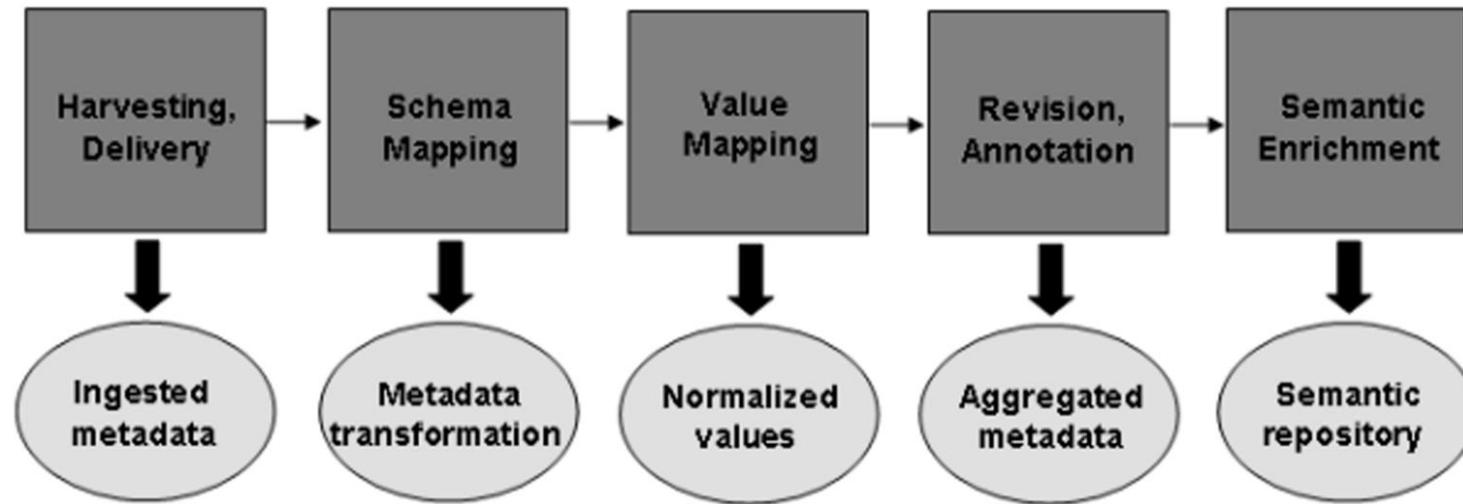
A **Universidade de Nevada**, por meio da equipe do departamento de Coleções Digitais das Bibliotecas da Universidade, desenvolveu o **UNLV's Linked Data Project**.





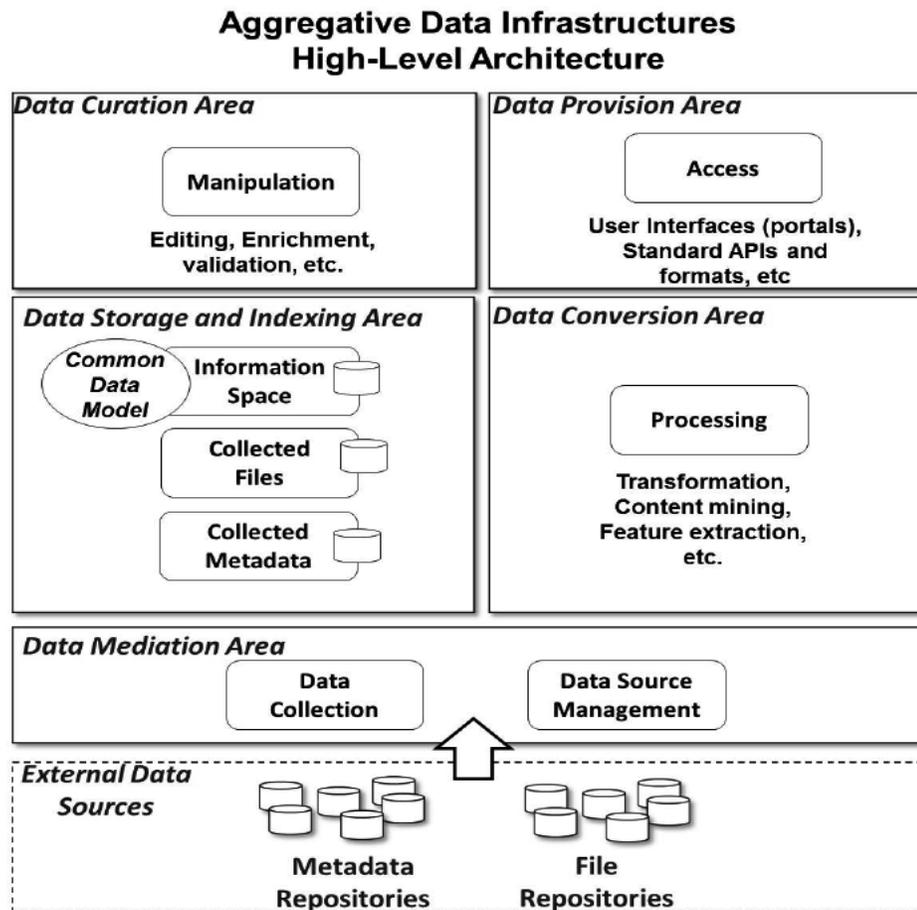
RESULTADOS

A **Fundação Europeia**, desenvolveu a Europeia, que reuniu mais de 55 milhões de objetos digitais das coleções on-line de mais de 3.500 galerias, bibliotecas, museus, coleções audiovisuais e arquivos de toda a Europa.



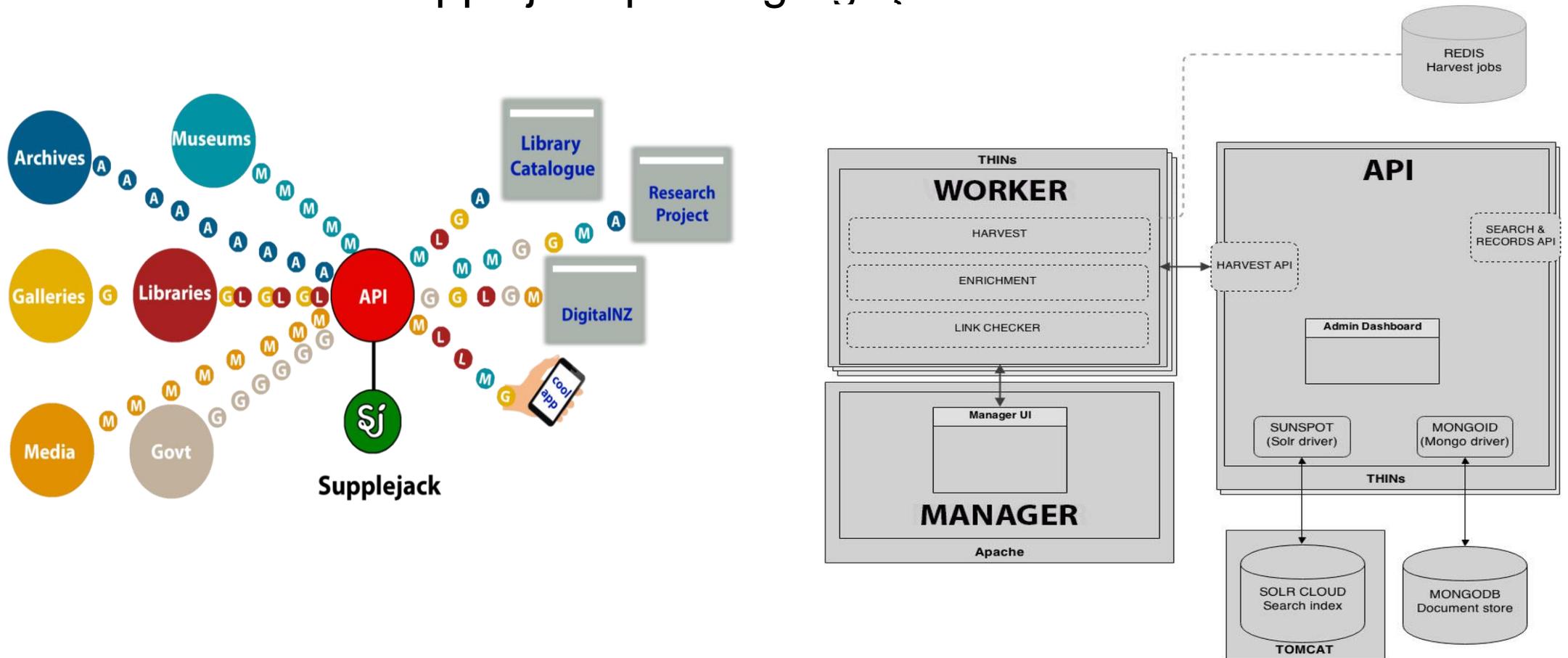
RESULTADOS

O Instituto de **Ciência e Tecnologias da Informação** desenvolveu o **D-NET**, um software que oferece um kit de serviços para a construção de Infraestruturas de dados.



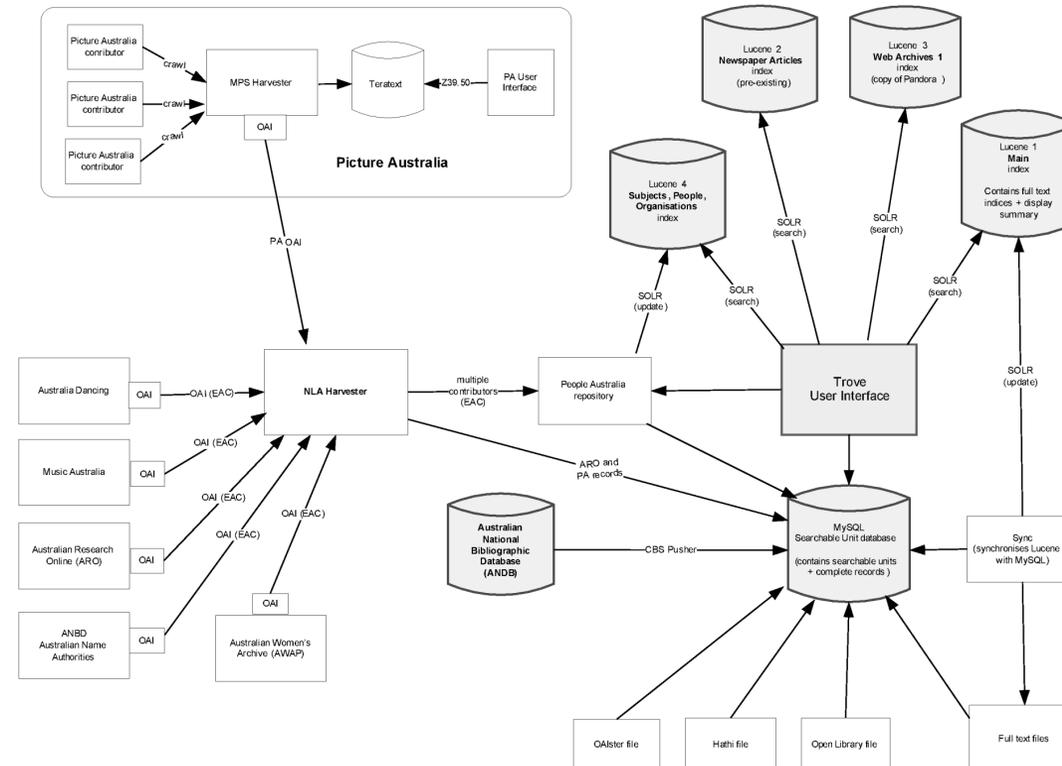
RESULTADOS

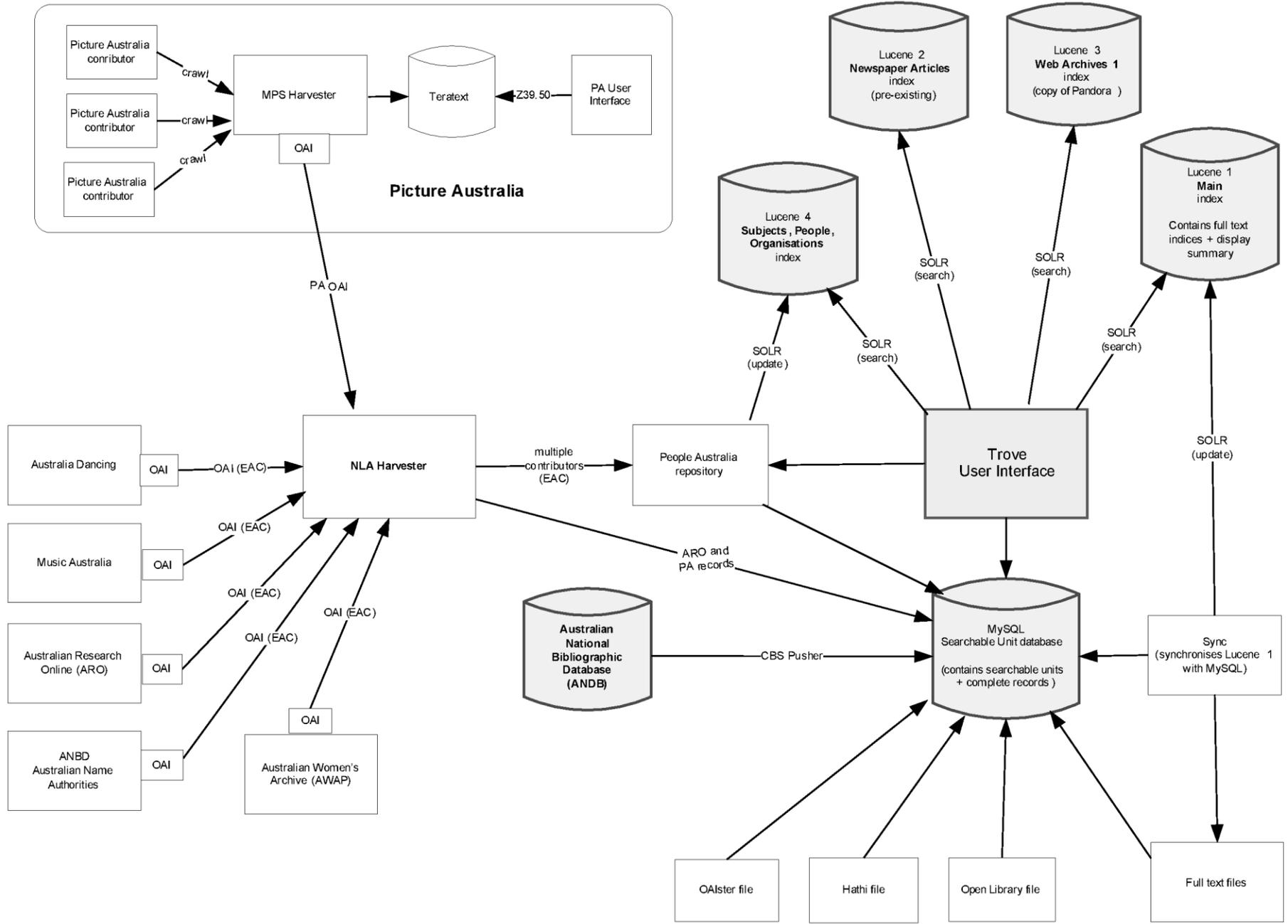
A **Biblioteca Nacional da Nova Zelândia** junto a Rede do povo Aotearoa Kaharoa desenvolveu, no início de 2006, o **DigitalNZ**, que utiliza o software Supplejack para agregação de dados.



RESULTADOS

A **Biblioteca Nacional da Austrália** desenvolveu o **Trove**, que tem o objetivo de fornecer recursos relacionados à Austrália. Além de um mecanismo de busca, reúne conteúdo de bibliotecas, museus, arquivos e outras organizações de pesquisa e fornece um conjunto de serviços.





ETAPAS PARA AGREGAÇÃO

Projeto/ Etapas	Extrair	Utilizar ontologias	Transformar	Reconciliar	Armazenar	Publicar	Expor	Possibilitar novas aplicações
AAC	X	X	X	X	X	X	X	X
Digital NZ	X	-	-	X	X	X	-	-
D-Net	X	-	X	X	-	X	X	-
Europeana	X	X	X	X	X	-	-	-
Mexicana	X	-	X	X	X	X	X	X
TROVE	X	-	-	-	X	X	-	-
UNLV's	X	X	X	X	X	X	X	-

ETAPAS PARA AGREGAÇÃO

- 1. Extrair:** extração dos dados em sua forma bruta, que podem estar, por exemplo, em pdf, em planilhas eletrônicas, documentos de texto, XML, em bancos de dados relacionais, dentre outras opções.
- 2. Utilizar ontologias:** selecionar vocabulários controlados pré-existentes para aplicação nos dados.
- 3. Transformar:** realizar a normalização, limpeza e correção sintática dos dados.
- 4. Reconciliar:** enriquecer os metadados por meio de outros dados existentes na web.

ETAPAS PARA AGREGAÇÃO

- 5. Armazenar:** local onde os dados coletados serão armazenados.
- 6. Publicar:** desenvolver uma interface única para busca integrada.
- 7. Expor:** disponibilizar os dados agregados por meio de API, que exponham os dados em formato RDF, OAI-PMH ou JSON.
- 8. Possibilitar novas aplicações:** a partir dos arquivos disponibilizados na etapa 'Expor' novas aplicações podem ser criadas.

CONCLUSÃO

- Workflows são genéricos e não apresentam o fluxo real de processos necessários, contrariando assim, um dos princípios básicos de um workflow, que é a possibilidade de ser replicado.
- Percebe-se a necessidade de um conhecimento técnico avançado e especializado para compreensão de todas as etapas.
- Percebe-se pouca preocupação com a qualidade dos dados inseridos, ou seja, os dados coletados na etapa de extração.
- Constata-se que não há escalabilidade, ou seja, à medida que o fluxo de dados cresce, o workflow torna-se impraticável.



WIDaT

2019

WORKFLOW DE AGREGAÇÃO DE DADOS: Processos para Criação de uma Interface de Busca Integrada do Patrimônio Cultural

OBRIGADA

Joyce Siqueira | joycitta@gmail.com

Dalton Lopes Martins | daltonmartins@unb.br