

Architecting the Open Education: The Integrated Metadata Warehouse

Dr. Sergey V. Zykov

& Olufemi Isheyemi

Higher School of Economics

13th Software Engineering Conference in Russia (SECR), Oct. 20-22, 2017, Saint-Petersburg

OER Definition and Research Problem



- **OERs** – freely accessible, open-licensed text/digital assets useful for teaching, learning, and research
- **The Problem**: Huge collections of underutilized learning materials published in the Internet
- **Tasks**:
 - (i) Locate OERs;
 - (ii) Evaluate OER quality;
 - (iii) Relate/connect OERs to the other OERs;
 - (iv) Share OERs with prospective users

Usually these are done manually by the individuals/end users



Research Idea, Aim and Features

- **Idea** - integrated approach for metadata warehousing in open educational resources (OERs)
- **Aim** – develop an architecture that integrates:
 - (i) automatic metadata extraction;
 - (ii) rule-based methodsto better utilize the OERs
- **Architecture features:**
 - (i) Aggregates the metadata into a single target repository;
 - (ii) Synchronizes the metadata with versatile OERs in the Web

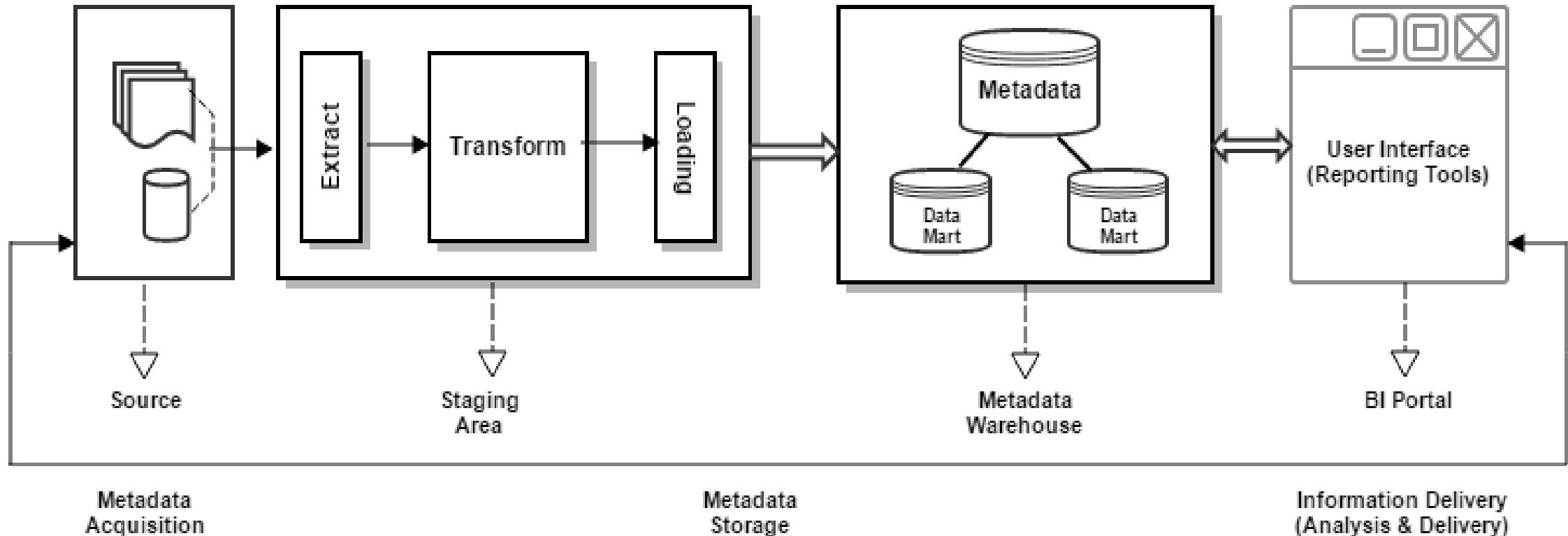
Objectives



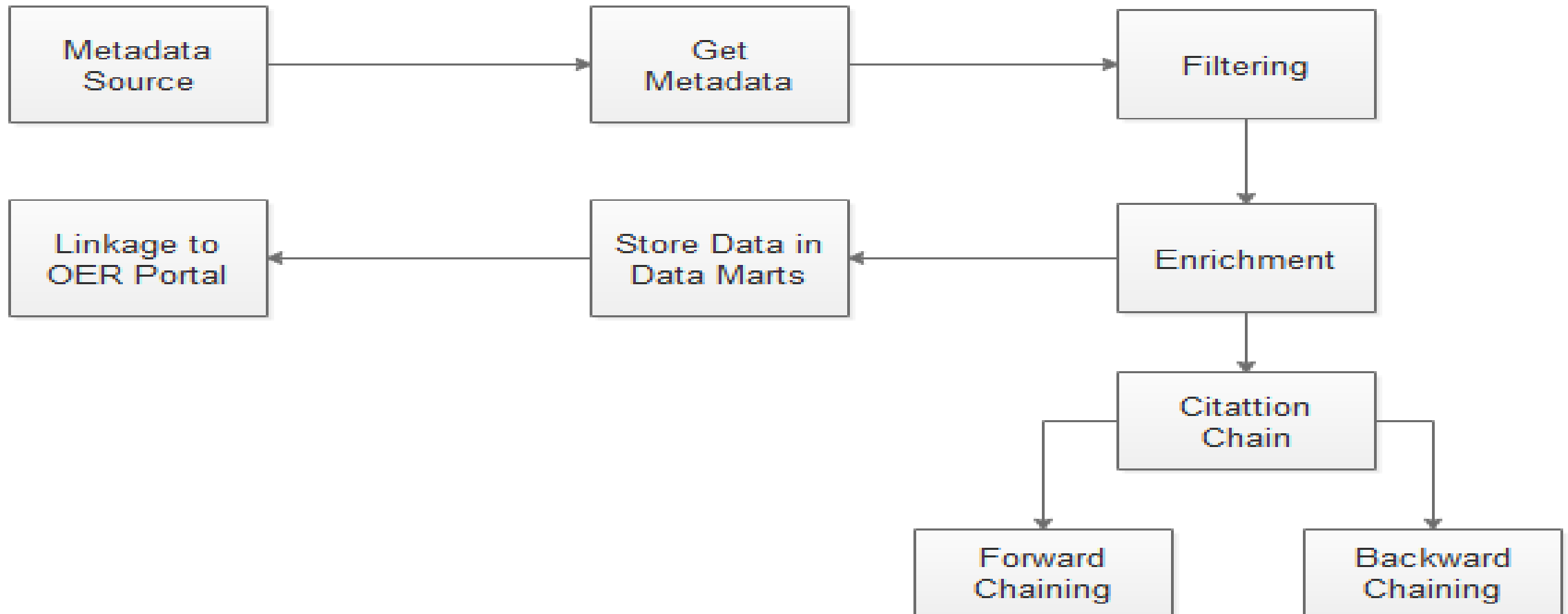
Develop:

- (i) Flexible method to extract OER metadata from major e-learning collections;
- (ii) Efficient technique to classify extracted metadata into groups;
- (iii) Repository to warehouse OER metadata;
- (iv) Web portal (prototype) to make OER metadata available to end users

Proposed Approach (ETL-based)

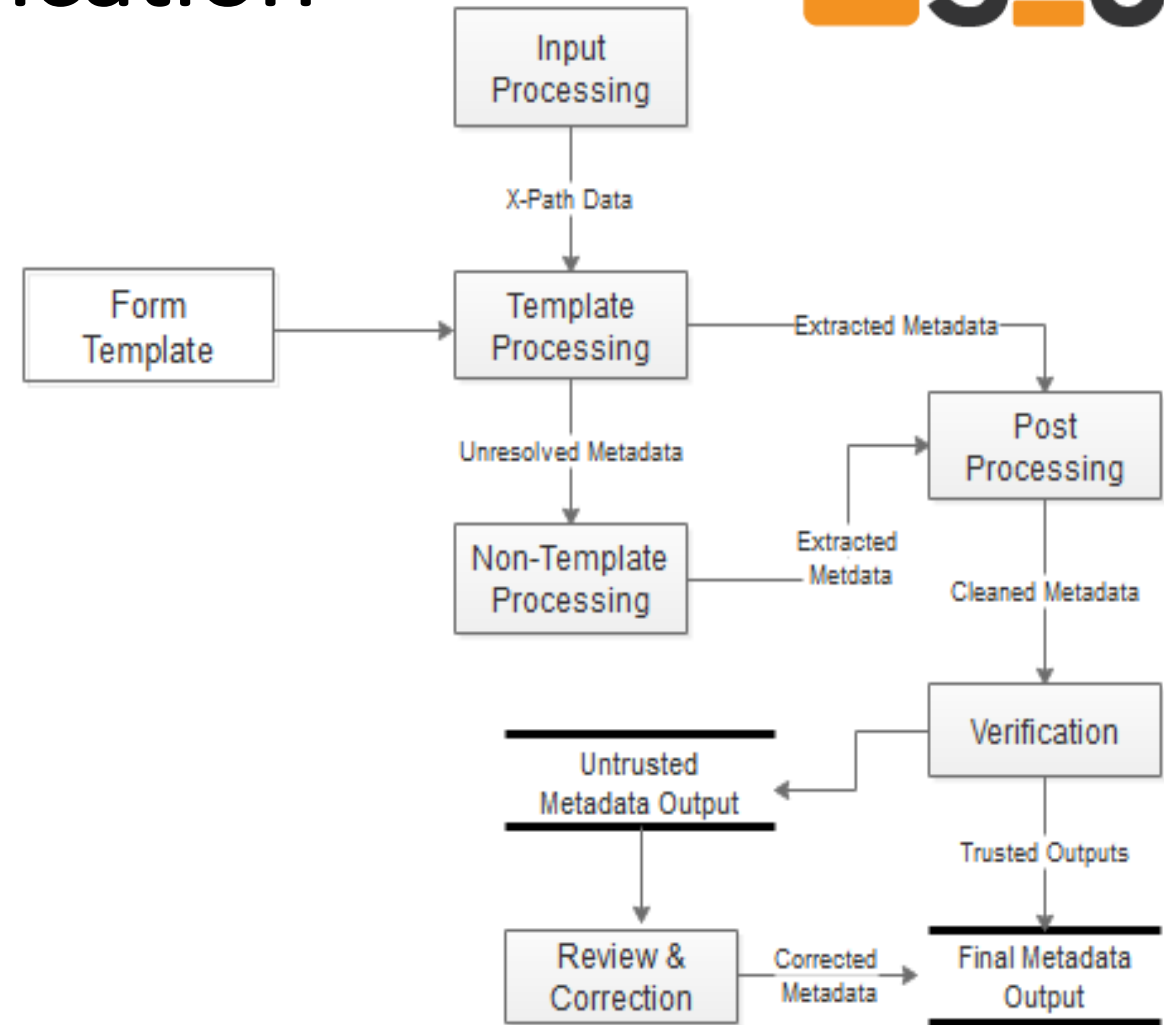


Process Flow (for the OER)



Metadata Quality Verification

- To verify efficiency of OER metadata architecture, we chose the quality metrics (Király & Bruce et al.):
 - (i) Completeness;
 - (ii) Accuracy;
 - (iii) Accessibility
- We assessed these metrics by *Europeana Search API* tool



Metadata Assessment

- The OER metadata quality metrics address the modification changes
- Quality assessed for OER platforms: (i) edX; (ii) Coursera; (iii) OER Commons

OER Platforms	Metrics		
	Completeness $Q_{comp} = \sum_{k=1}^N \frac{P(i)}{N}$	Accessibility $Q_{link} = \frac{links(instance_k)}{Max_{i=1}^N ((links(instance_i)))}$	Provenance $Q_{prov} = \frac{\sum_{i=1}^N Q_{avg_i}}{N}$
<u>edX</u>	4.10	3.90	4.15
OER Commons	4.20	4.13	4.32
Coursera	4.08	4.18	3.85

Results:



- Architecture designed for OER metadata warehouse
- Architecture aimed at discoverability of Web-based learning content by combining proven metadata management methods
- Experimentation proved suitability for OER metadata warehousing



Future Work

- Move beyond capturing metadata
- Explore discovering good quality learning materials
- Apply methods for:
 - (i) elastic search,
 - (ii) citation chain,
 - (iii) automatic indexing
- Synergy should create a network of useful learning resources



References

1. Alexandru Adrian Tole (2015): The Importance of Data Warehouses in the Development of Computerized Decision Support Solutions. A Comparison between Data Warehouses and Data Marts. Database Systems Journal
2. Péter Király (2005): A Metadata Quality Assurance Framework
3. Ralph Kimball and Joe Caserta (2004): The Data Warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Wiley Publishing, Inc.
4. Thomas R. Bruce and Diane I. Hillmann (2004): The Continuum of Metadata Quality: Defining, Expressing, Exploiting
5. Tian Belawati (2014): Open Education, Open Education Resources, and Massive Open Online Courses
6. Sergey Zykov (2016): The Online Evolution: from Early Repositories to State-of-the-Art MOOCs. In: Central and Eastern European Software Engineering Conference in Russia (CEE-SECR '16), Moscow, Russian Federation — October 28 - 29, 2016

13th Software Engineering Conference in Russia (SECR), Oct. 20-22, 2017, Saint-Petersburg



Thank you!