

# Integrating research indicators for use in the repositories infrastructure

Drahomira Herrmannova and Petr Knoth  
CORE

Knowledge Media institute, The Open University  
United Kingdom

## In this talk

- Research indicators: what (and what not) for and why
- Challenges in integrating indicators in the repositories infrastructure

## In this talk

- **Research indicators: what (and what not) for and why**
- Challenges in integrating indicators in the repositories infrastructure

# Problem

- Biblio-, webo-, alt-metrics – controversial in research evaluation
- But, they can be applied with measurable success in information retrieval and research analytics
- Most repositories (and aggregators) do not make an effective use of these metrics yet

# Freely available collections

## Citation data

- Microsoft Academic Graph - free alternative to Scopus and WoS
- Initiative for Open Citations (I4OC)

## Usage data

- Altmetrics API
- Mendeley API
- IRUS
- Others

# Where can research indicators be applied

## 1. Enhanced information retrieval

- Search
- Recommender systems

## 2. Research analytics

- Analysis of research trends
- Identify areas of strength within institutions
- Expert search
- Analysis of research collaboration networks
- Analysis of research argumentation

# Result of not using indicators

Repository and cross-repository information retrieval systems have poor performance (and no one wants to use them when in combination with metadata only indexing)

Little or poor research analytics available

## In this talk

- Research indicators: what (and what not) for and why
- **Challenges in integrating indicators in the repositories infrastructure**



## Challenges in integrating these datasets with the repositories infrastructure?

- Not a complete overlap to merge
  - DOI
  - Combination of fields
- Size of the datasets
- The process can be resource intensive/complex:
  - Beyond the ability of a typical repository
  - Role for aggregators

## Challenges in integrating these datasets with the repositories infrastructure?

- **Indicators are changing all the time, but metadata and resources are not**
  - Merge
  - Integrate (index)

# Approaches to integration

- **Batch:** Merge data and index once in a while
- **Continuous (streaming the changes):** Integrate immediately as indicators change

# Indexing

No updates in indexing!

- Update means delete and reinsert
- Many reinserts cause the index to grow in size and become less optimal for retrieval => rebalance the index

# Consequences

	Batch	Continuous
Always up to date	✗	✓
Make efficient use of metrics in IR systems	✓	✗

# The conflict

## Staying up to date vs System response

# Is there a middle path?

- Elasticsearch parent-child relationship
  - Children are stored as separate documents
  - Lightweight structure, can be recreated quickly
- But ... you still pay a price in performance

# Conclusions

- Research indicators currently widely (and wrongly) used in research evaluation have a significant potential in academic information retrieval and research analytics.
- Technical challenges in integrating them: staying up to date vs system response
- Parent-child indexing approaches offer a solution somewhere in the middle