

DDI3 Metrics

Claude Gierl

Jon Johnson

Context

The CLOSER Project

Large scale metadata ingest



Software

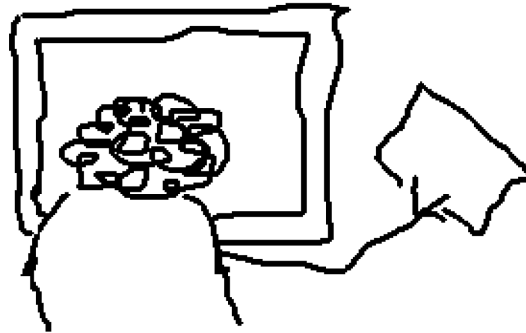
CADDIES (Ruby on Rails)

Colectica Repository



Motivations and Requirements

Monitor the progress of the ingest
Compare metadata sets
Easy to implement



What do we want to measure?

How big?

How good?

One measure not enough



What is 'good'?

Lack of redundancy in the metadata

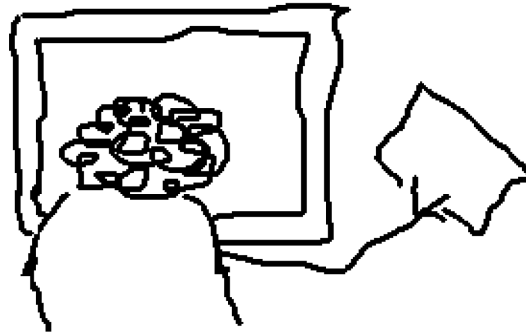
Rich interlinked information



Measure of magnitude: Cell Count

DDI3 is xml schema

Count of xml elements



Cell Count: Pros

Right granularity

Easy to define and reproduce



Cell Count: Cons

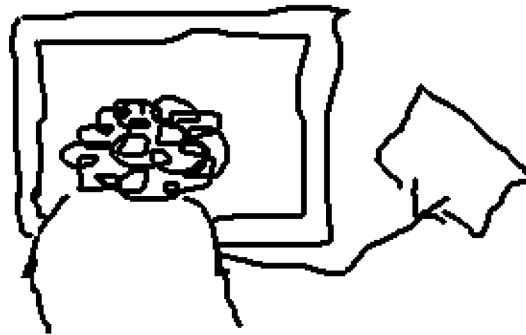
Does not count all information
Sensitive to stylistic variations



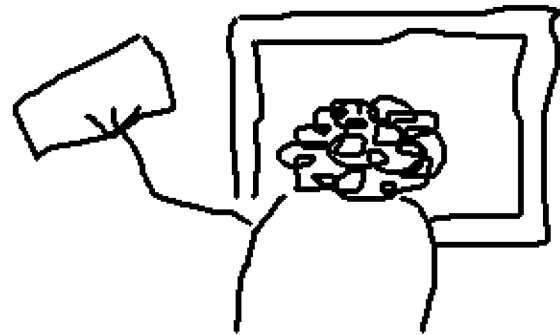
Measure of Goodness: Synaptic Density

Specific to DDI3

Relies on its use of references



Synaptic Density



Synaptic Density: Properties

- Measures what we want
- Increases when references increase
- Increases when cell count decreases



Synaptic Density

Cleaning and Deduplication
Phase:

Cell Count goes down

SD goes up



Synaptic Density

Harmonisation and linking Phase:
Reference count goes up
SD goes up



Synaptic Density: Pros

Useful throughout the lifecycle of the metadata



Synaptic Density: Cons

Sensitive to DDI3 version changes

Discontinuities

For example...



DDI3.1 to DDI3.2

Response Domains now use
references to *Managed
Representations*

SD will go up

Reflects an actual improvement



Conclusion

DDI3 Metrics capture what we want

But not rigid measures

Require analysis and interpretation

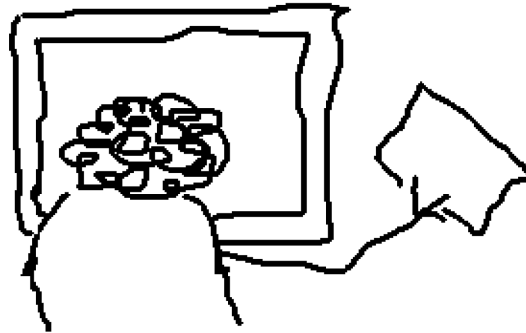


But where are the figures?

Early phase of ingest

Implementation of DDI3.2 nearly done

Next time...



For more info:

www.cls.ioe.ac.uk

www.closer.ac.uk