Building an Enterprise Metadata Strategy

Presentation to the International Society for Knowledge Organization, Singapore Chapter

Michael Andrews | StoryNeedle.com | 5 April 2019
What we will cover in this session

1. How metadata and content structuring can help overcome knowledge silos
2. The value of graphs or knowledge graphs
3. How to build a shared model of knowledge
4. Strategic benefits of better metadata
5. Reusing information in different contexts
Overcoming silos

Silos are a byproduct of specialisation
Knowledge becomes proprietary
Compartmentalisation doesn’t lead to a common language (shared vocabularies)

Expertise and accountability drive formation of specialized groups that create boundaries around knowledge.

Each group has its own mental models for thinking about things.
Transcend Single-Purpose Learning and Knowledge

Images: https://www.news.ucsb.edu/2015/015831/something-crow-about
Do you have the ability to connect different information created by different parties?
Standards make information easier to share, because there’s a common language.

Source: Open Data Institute
Data integration promises a lot...

**Step 1**
YOU choose data
- private data
- CRM data
- crime data
- social network data
- supplier data
- product data
- geo data
- open data
- weather forecast data
- website data
- payments data

**Step 2**
WE link data
- CRM data
- crime data
- weather forecast data
- website data
- payments data

**Step 3**
YOU ask questions
- CRM data
- crime data
- weather forecast data
- website data
- payments data
“Computers do not automatically remove silos from our lives. Far from it.”

“Data does not reorganise itself, or break down silos by itself.”

Gillian Tett of the Financial Times, in The Silo Effect
“Data ages like wine. Software ages like fish.”
Transcending silos is not entirely a technical issue

New approaches to managing information, especially graphs, are promising.

But first need to get alignment within the organisation about goals.

All kinds of stakeholders need to know what available, see what’s available, and use what’s available.
What’s important to know about a neighbourhood?

Airbnb example of a knowledge graph
Graphs express relationships

How important are noodles?

How are they connected to other food concepts?

How are they connected to restaurants and orders?

Source: Uber/UberEats
Knowledge management is facilitated when considered in terms of entities:

- Resources, Physical items, Capital
- People, Organisations, IT systems
- Buildings, Cities, Rivers
- Durations, Dates
- Eras, Stages, Significant Transitions
- Eligibility, Standards, Processes
Entities as an anchor for knowledge management

How are these topics on Quora related to one another?

https://blogs.bodleian.ox.ac.uk/digital/2018/02/14/some-ways-wikidata-can-improve-search-and-discovery/
Connections in Wikidata show relationships

“Pinterest and Quora are examples of sites in which content is organised by topic. To use topics for discovery, it helps if we can distinguish topics that are closely related, somewhat related and unrelated.”
Knowledge Graph: Airlines with hub at Changi

Source: Wikidata via Metaphactory
Free tagging often results in problems later

Need to reconcile different labels to determine what’s equivalent.

Image: Microsoft

Discover different ways people refer to products, people, locations, and more.
Standardised identifiers are very helpful

PermID from Thomson Reuters
Improving the customer experience

Customers want their content just right
Contextual information on demand

See also [edit]

- DBpedia
- Google Assistant
- Graphiq
- Linked data
- Ontology (information science) – also called knowledge graphs
- Semantic web
- Wikidata

Reference

Increasing the reach of the information

How easily can information be consumed by machines and by people?

Reach depends on the **interoperability** of data

Audiences must be able to access information using their devices and platforms

How to enable information to be available where it is needed?

- Metadata standards
- Modularised information can be repurposed if planned appropriately
The smartphone is the gateway to the Cloud

People have many roles.

They often use one device.

Figure 2 — Example of roles a user can assume in device use scenarios

Source: International Standards Organisation
Delivering information to new channels

Mobile (*mission not yet accomplished*)
- Still difficult to deliver certain information (tables, long docs) to small screens
- Requires ‘refactoring’ content (e.g., cards)

Augmented reality
- Overlaying information on real world entities

Bots
- Entity based interactions

Internet of Things
- Information from entities in the physical world

Needs go beyond delivering articles as webpages.
GS1: Metadata as enabler

- Traceability
- Sustainability
- Automation
- Empowerment

Transparency

“Provide structured metadata”

Just because information is online doesn’t mean it is accessible

Source: US General Accountability Office
Better analytics

How often is information being used?

In what context is it being used?

As information gets more widely connected, our assumptions about who is using the information and why it is being used may be wrong.

Can now track content according to topic and actions associated with it.
Improving publishing operations

Everyone is a stakeholder
Make it easy for colleagues or partners to add info to the graph

Source: Airbnb

A tool to categorize experiences manually
When everyone contributes, everyone is responsible for quality.

Source: Airbnb
Develop a shared schema to organise relationships
ESCO: European Skills, Competencies and Occupations

Figure 4: The link between the three pillar
manage content metadata

Description
Apply content management methods and procedures to define and use metadata concepts, such as the data of creation, in order to describe, organise and archive content such as documents, video and audio files, applications and images.

Alternative label
manage metadata
manage content

Skill type
skill

Skill reusability level
cross-sector skills and competences

Essential skill/competence of
technical communicator
e-learning developer
big data archive librarian
instructional designer
web content manager
ICT documentation manager

Optional skill/competence of
marketing manager
online community manager

Status
released

Concept URL
http://data.europa.eu/esco/skill/9710092c-f174-4e84-95bf-2064d39ad7a0
What’s common? Inheritance

Source: ESCO
Once you know what you have, you can reuse the information more easily
Don’t Repeat Yourself
Wikipedia: One of the most successful examples of large scale content reuse

The concept of ‘transclusion’ allows the same instance of content to appear in many places at the same time.

There is one master copy.

Semantic MediaWiki is used by corporate and government users as a knowledge platform

**Semantic MediaWiki** (SMW) is a free, open-source extension to MediaWiki – the wiki software that powers Wikipedia – that lets you store and query data within the wiki’s pages.

Semantic MediaWiki is also a full-fledged framework, in conjunction with many spinoff extensions, that can turn a wiki into a powerful and flexible knowledge management system. All data created within Semantic MediaWiki can easily be exported or published via the Semantic Web, allowing other systems to use this data seamlessly.

**SMW 3.0.1**
To reuse information, must decide basic procedures

Who gets to re-use the information (role-based reuse)

- Creator or republisher rights

Where can the information be re-used? (context-based reuse)

- Deciding and designing opportunities to re-use information
Information reuse is only partly a technical issue

Need to determine where content can be reused

Reuse can be valuable, but on occasion can be excessive

**Spamming**: promoting the same information too frequently
  - Marketers often guilty of this

Pushed information may not be aligned with user needs
  - Reused information can be too generic or broad to answer questions
The process starts with shared goals within the enterprise

➔ Highlight current pain points to build awareness of missed opportunities
➔ Adopt external terminology when categorising your knowledge
➔ Use open standards when implementing technical solutions
➔ Connect to external resources (data and code libraries) to bootstrap
Questions