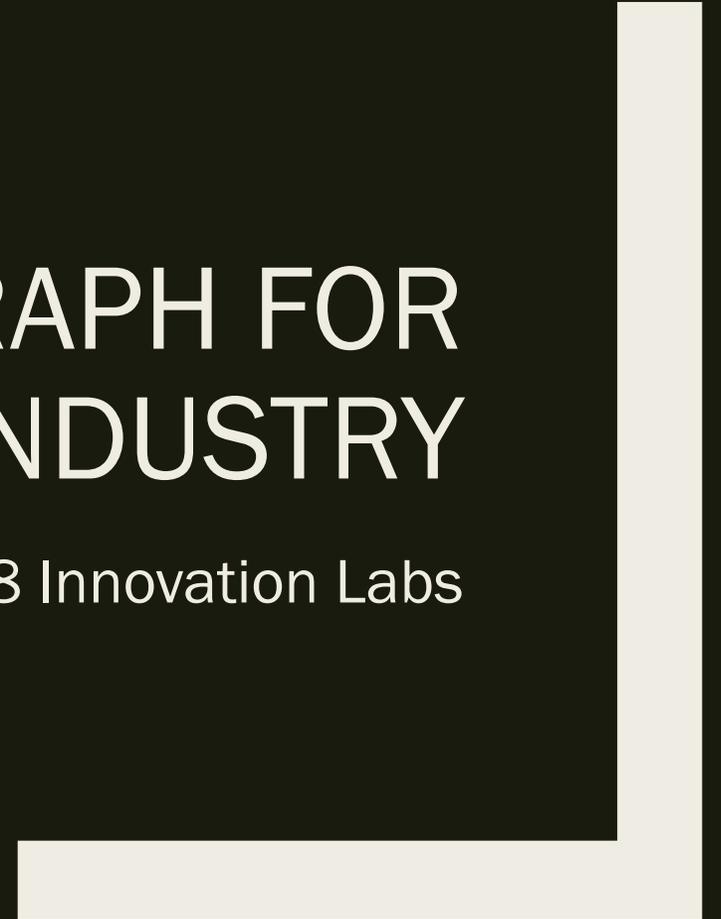# CHAPTER 8
SUCCESS STORIES

Martin Liahagen

Agenda

- A Knowledge Graph for Innovation in the Media Industry

- Applying Knowledge Graphs in Cultural Heritage

- Applying Knowledge Graphs in Healthcare

A KNOWLEDGE GRAPH FOR INNOVATION IN THE MEDIA INDUSTRY

HAVAS 18 Innovation Labs

The Business Problem

- The way consumers and brands communicate is changing dramatically
  - *Personalisation*
  - *New communication channels*
  - *New technology*

- HAVAS 18 Innovation Labs
  - *A "better together" approach*
  - *Wants to connect innovators, universities, start-ups and technology trends*

The HAVAS 18 Knowledge Graph I

- ■ "Enable search, discovery and understanding of information about start-ups in their first 18 months"

Data sources:

- ■ Generalist and specialist web sites for core data
  - – *AngelList, CrunchBase*
- ■ Social networks for relationship data
  - – *Facebook, LinkedIn, Twitter*
- ■ Newsfeeds for extended media coverage

The HAVAS 18 Knowledge Graph II

- Structured and integrated in RDF
- Schema built on top of
  - *Schema.org*
  - *FOAF*
  - *SKOS*
  - *rNews*
- RESTful API with JSON
  - *API services include CRUD methods*

- Rapporteurs
  - *Can add and modify entities and relationships*
 - *Can also be curators*

Value Proposition

■ Gathers insight and information automatically

– *Meaning business strategists have something to start with*

■ Can be used to evaluate trends and start-ups

■ Provides knowledge about the best-in-class talent for new technology

Challenges

- Entity resolution and disambiguation
  - *Evidence model, with key classes*
    - E.g. start-ups could have Founder, Client and Technology
  - *Knowledge tagger extracts entities, which are matched with evidence models*
- Version management
- Resilience against changes in data sources, like web APIs
- Monitoring potential decay

APPLYING KNOWLEDGE GRAPHS IN CULTURAL HERITAGE

The CURIOS Project
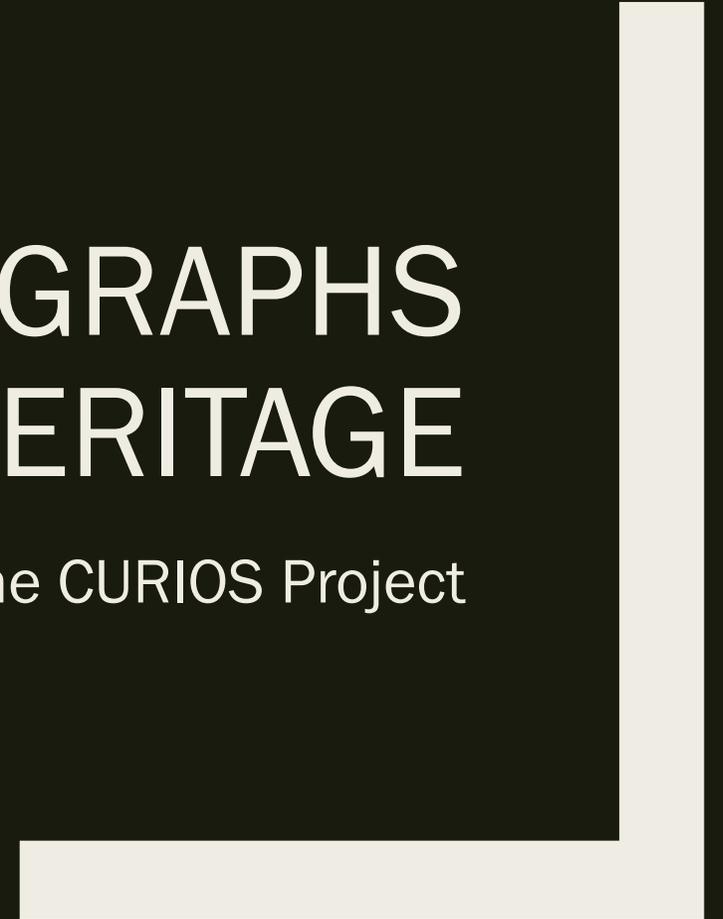
Digital Cultural Heritage

■ Projects
  – *UK Culture Grid*
  – *Continuous Access to Cultural Heritage (CATCH)*

■ Ontologies
  – *Categories for the Description of Works of Art*
  – *International Committee for Documentation Conceptual Reference Model (CIDOC CRM)*

■ Technologies
  – *XML*
  – *Distributed databases*
  – *RDF/OWL*

Digital Cultural Heritage and Linked Data

- OpenART
  - *The London Art World 1660 – 1735*
  - *Linked Open Data format*
- CultureSampo
  - *Cultural heritage archive for Finland*
- datos.bne.es
  - *Library data from The National Library of Spain*

Challenges

- Data heterogeneity
  - *Data formats are not consistent*
  - *Makes it hard to integrate different archives*

- Computer literacy of the community
  - *Ranges from novice to experienced*
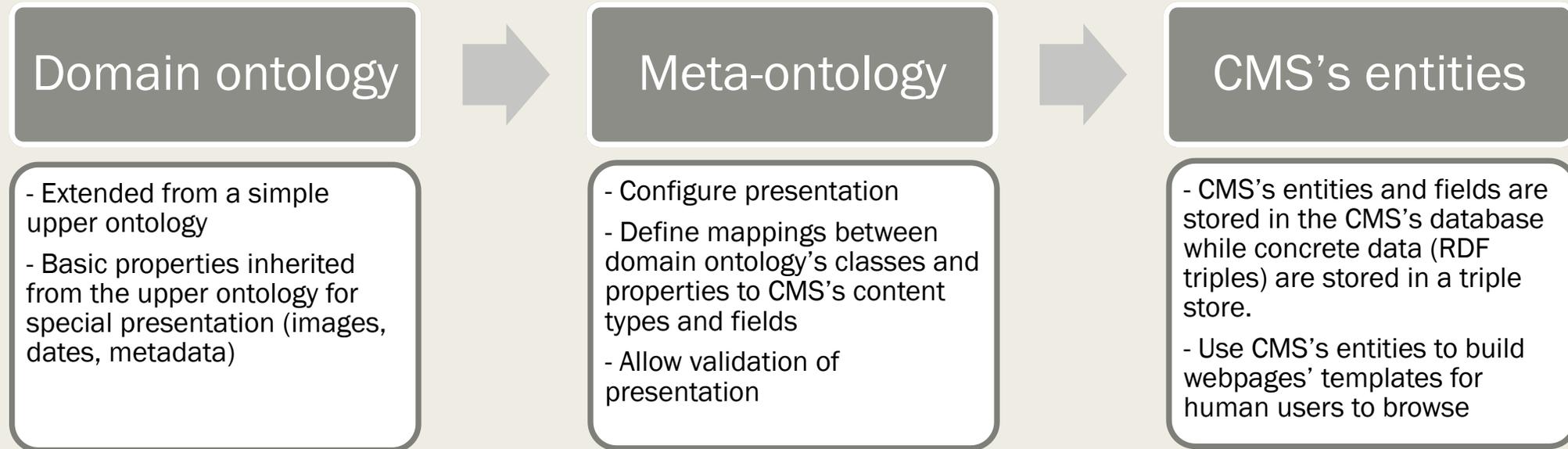 - *Current tools require in-depth technical knowledge*

The CURIOS Project I

- CUltural Repositories and InfOrmation System

- Solves both challenges

- Data heterogeneity
  - *Uses Linked Open Data as standard*

- Computer Literacy
  - *Uses Drupal, a Content Management System (CMS)*
  - *Does not require in-depth knowledge*

The CURIOS Project II

- Loosely coupled
  - *Data is stored in RDF*
  - *Data presentation configurations stored in Drupal's SQL database*
- Supports semantic searching via SPARQL, configurable presentation and visualisation services.
- Used for
  - *preserving cultural heritage*
  - *improving tourist experiences and exhibitions*
  - *storing research data*

Constructing the Knowledge Graph

Domain ontology	→	Meta-ontology	→	CMS's entities

Domain ontology
- Extended from a simple upper ontology
- Basic properties inherited from the upper ontology for special presentation (images, dates, metadata)

Meta-ontology
- Configure presentation
- Define mappings between domain ontology's classes and properties to CMS's content types and fields
- Allow validation of presentation

CMS's entities
- CMS's entities and fields are stored in the CMS's database while concrete data (RDF triples) are stored in a triple store.
- Use CMS's entities to build webpages' templates for human users to browse

Mapping from a domain ontology to CMS's entities

Presentation of Instances

4 Caverstay

Croft 4 was first occupied by Donald Mackinnon and then by his son Roderick.

The croft was particularly congested in the early years of the 20th century, supporting four large families of Mackinnons. The situation forced many to leave the village for Stornoway, the mainland or abroad.

Back to listing

Title:	4 Caverstay
Record Type:	Crofts and Residences
Gaelic Name:	4 Cabhairstaidh
Type:	Croft
Record Owned By:	CEP
Record Maintained By:	CEP
Subject Id:	7560

Louis Mackinnon & family, Caverstay

Lived Here

Angus Mackinnon
Johanna Mackinnon
Donald Mackinnon
Mary Bell Macleod
Ann Mackinnon
Catherine Maclennan
Roderick Mackinnon
Donald Mackinnon
Louis Mackinnon
John Mackinnon
Mary Ann Mackinnon
Louis Mackinnon
Mary Macdonald
Roderick Mackinnon
Christina Mackinnon

Associated With

Euphemia Maciver
Ruaraidh Rob Mackinnon I: Memories...
Ruaraidh Rob Mackinnon II: Off to...
John Murdo Macdonald
Catherine Macdonald

Located At

Caverstay

Details of a croft

Dealing with Vagueness

■ Inexact dates and times

■ Fixed by using ranges
- *E.g. 1780 becomes [1780-01-01, 1780-12-31]*

| General Class	Pattern	Example	Frequency	Subtotal	Covered
Exact to the day	y-m-d	1780-06-13	12949		
	d-m-y	10/6/45	725		
	d-M-y	12 MAY 1780	272		
	M-d-y	May 12 1780	8		
				13954	yes
Exact to the month	y-m	1780-12	274		
	M-y	Aug 1780	443		
	m-y	03/1780	2		
				719	yes
Exact to the year	y	1978	10825	10825	yes
Exact to the decade	dec	IN 1860'S	1415	1415	yes
Exact to a range of years	y-y	1939-45	242		
	beforey	pre 1918	2		
	aftery	AFT 1890	3		
				247	yes
Exact to the century	cent	20th Century	4	4	yes
Vague within less than a month	mend	Aug/Sept 1972	26	26	yes (using a date range)
Vague within more than a month but less than a year	yend	1978/79	7	7	yes (using a date range)
Vague year	cy	C. 1932	566		yes
	moddec	early 1950s	86		(using a
				652	date range)
Vague around a decade	cdec	c 1950s	2		yes
	modcent	LATE 1600S	3		(using a
				5	date range)
Not directly interpretable as a date	unk	D.I.I.	3069	3069	no
GRAND TOTAL				30923	

Analysis of date forms

APPLYING KNOWLEDGE GRAPHS IN HEALTHCARE

IBM

Clinical Practice Guidelines (CPGs)

- *"Collective sets of treatment recommendations that attempt to capture the best medical practices for different pathologies"*

- They become more complex as new medical treatments are discovered

- CPGs recommend Treatment Programmes (TP)

  - *TP is a sequence of medical procedures*

- Deviations from the CPGs are not wanted, but…

- CPGs cannot cover everything
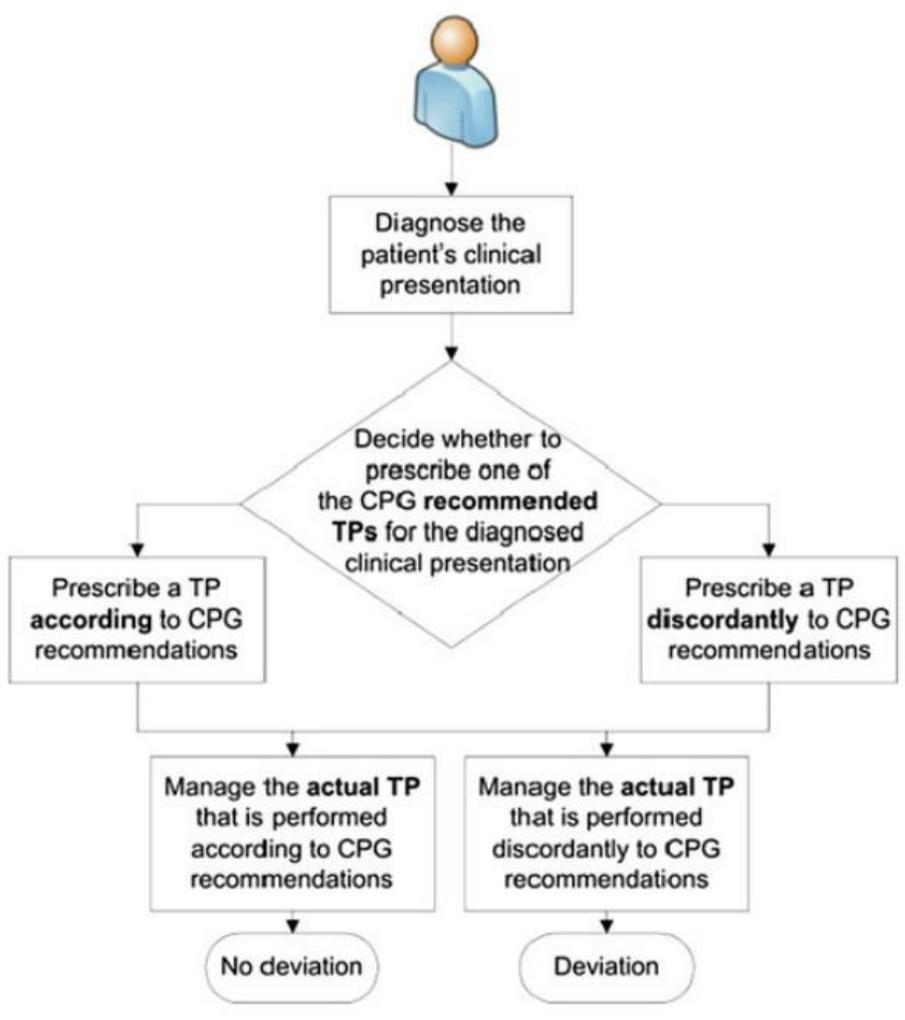
The Project

- **Goal:** Automatically monitor compliance to CPGs and identify potential rationale for deviations

- **How:** Computationally parse Electronic Health Records (EHRs) using
    - *Natural Language Processing*
    - *Data modelling*
    - *Comparison algorithms*

- Data based on EHRs of patients in Italy with Soft-Tissue Sarcoma
    - *Rare cancer with numerous treatment options*

Natural Language Processing Techniques

■ Italian to English machine translation

■ Processed the unstructured information with UIMA framework

    – *Part-of-speech tagging*

    – *Normalisation*

    – *Entity and relationship extraction*

    – *Semantic analysis*

    – *Negation*

    – *Disambiguation reasoning*

Services Based on the Knowledge Graph I

- Treatment programme comparison
  - *Compare the actual and the recommended TP*
  - *Decide how much the actual TP deviates*
- Classifications of deviations
  - *Over- or undertreatment*
  - *Difference in chemotherapy drug*

Services Based on the Knowledge Graph II

■ Extracting justifications

   – *"In light of extension of illness, the patient's age and preliminary activity of molecule in this particular histotype, starting chemotherapy with gemcitabine."*

■ Classification of reasons

   – *Cancer status, previous treatments, patient preference*

Contributions

- Helping doctors understand clinical deviations
- Helping to understand the decision-making process of physicians
- Improving CPGs