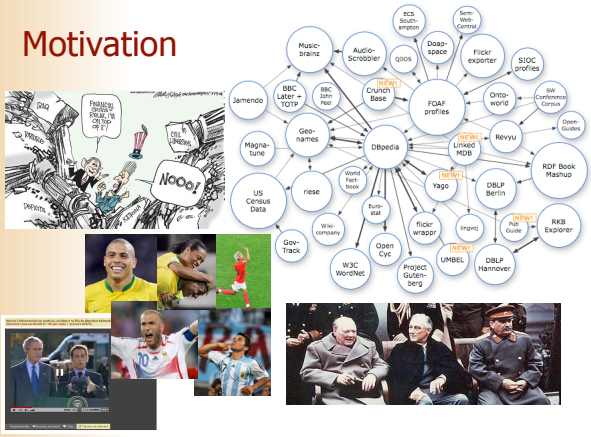
Semantic Media Web: Create, Annotate, Present and Share your Media

Lynda Hardman, Raphaël Troncy (Eurecom)
<Lynda.Hardman | Raphael.Troncy@cwi.nl>

CWI, Interactive Information Access
UvA, Institute for Informatics

Motivation



Learning Objectives

- Understand multimedia applications workflow
  - Canonical processes of media production model
- Understand design rationale of COMM, a Core Ontology for Multimedia

Creating a semantic media web

- Allocate URI to media asset
- Attach metadata using RDF

But…

- want to attach metadata to parts of a media asset
- want to combine multiple (parts) of assets into new assets/presentations
- each media type/format needs a different player (whereas everyone can display text)

We don't even care about media!

- We want to enable
  - the processing of information-bearing content
  - of one or more media types
  - that can be interpreted by end users
- End-users are primarily interested in
  - the meaning conveyed by a combination of media assets
  - interacting further with the media
    - as part of complex search task
    - passing it on to someone else in media "chain"

We can use the (semantic) web

- *Web* enables the identification and delivery of units of information of different data types
- *Semantic web* enables the association of metadata with each identified unit/fragment
- To find and use them, we need mechanisms:
  - for identifying (part of) an individual media asset
  - for associating metadata with an identified fragment
  - that enable larger meaningful structures to be composed, identified and annotated

Really need…

- to be aware of the human aspects of multimedia
- multimedia assets are not created "*in vacuo*", but by someone for a specific purpose
- more than the information "expressed" by the media asset itself, e.g.
  - the creator and the intended purpose
  - provenance also important on the semantic web, where a multimedia presentation may be composed of assets published by many different sources
  - the reason for organising assets in a specific way

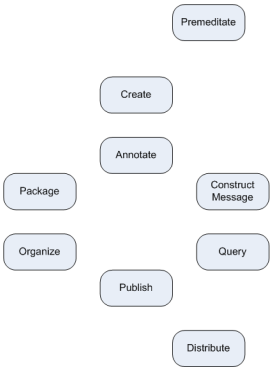
7

Understanding Multimedia Applications Workflow

- Identify and define a number of canonical processes of media production
- Community effort
  - 2005: Dagstuhl seminar
  - 2005: ACM MM Workshop on *Multimedia for Human Communication*
  - 2008: Multimedia Systems Journal Special Issue (core model and companion system papers)
    editors: Frank Nack, Zeljko Obrenovic and Lynda Hardman

8

Overview of Canonical Processes



9

Example 1: CeWe Color PhotoBook

- Application for authoring digital photo books
- Automatic selection, sorting and ordering of photos
  - Context analysis methods: timestamp, annotation, etc.
  - Content analysis methods: color histograms, edge detection, etc.
- Customized layout and background
- Print by the European leader photo finisher company

http://www.cewe-photobook.com

10

CeWe Color PhotoBook Processes

- My winter ski holidays with my friends

Premeditate

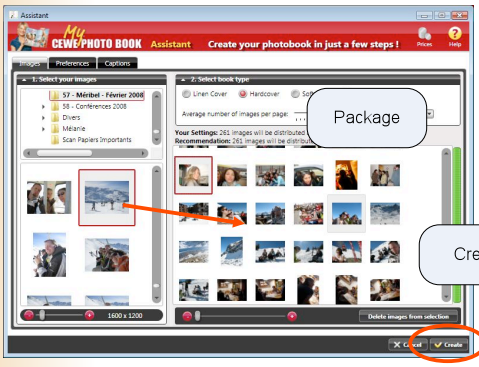
-

Construct Message

11

CeWe Color PhotoBook Processes

-

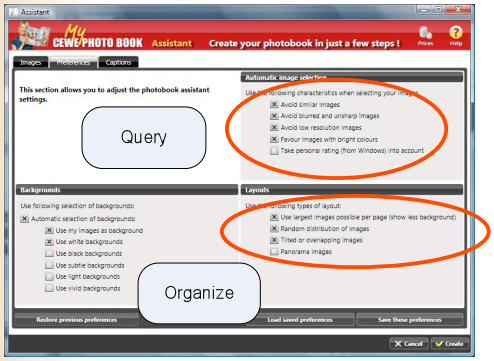


12

CeWe Color PhotoBook Processes

13

CeWe Color PhotoBook Processes

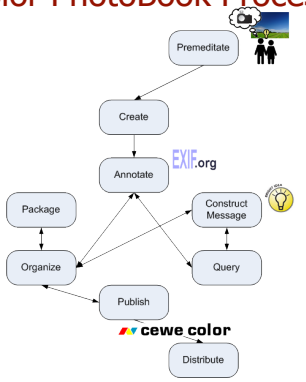


14

CeWe Color PhotoBook Processes

15

CeWe Color PhotoBook Processes



16

Canonical Processes 101

- Canonical: reduced to the simplest and most significant form possible without loss of generality
- Each process
  - short description
  - illustrated with use cases
  - input(s), actor(s) & output(s)
- Formalization of processes in UML diagrams in paper (see literature list)

17

Premeditate

- Establish initial ideas about media production
  - *Design a photo book of my last holidays for my family*
  - *Create argument-based sequences of videos of interviews after September 11*
- Inputs: ideas, inspirations from human experience
- Actors:
  - camera owner
  - group of friends
- Outputs:
  - decision to take camera onto ski-slope
  - structured set of questions and locations for interviews

18

Create Media Asset

- Media assets are captured, generated or transformed
  - Photos taken at unspecified moments at holiday locations
  - Synchronized audio video of interviewees responding to fixed questions at many locations
- Inputs:
  - decision to take camera onto ski-slope;
  - structured set of questions and locations for interviews
- Actors:
  - (video) camera, editing suite
- Outputs:
  - images, videos

19

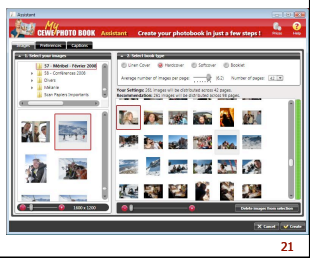
Annotate

- Annotation is associated with asset
- Inputs:
  - photo, video, existing annotation
  - optional thesaurus of terms
- Actors:
  - human, feature analysis program
- Outputs:
  - Complex structure associating annotations with images, videos

20

Package

- Process artifacts are packed logically or physically
- Useful for storing collections of media after capturing…
- … before selecting subset for further stages



21

Query

- User retrieves a set of process artifacts based on a user-specified query
- Inputs:
  - user query, in terms of annotations or by example
  - collection(s) of assets
- Actors:
  - human
- Output:
  - subset of assets plus annotations (in no order)

22

Construct Message

- Author specifies the message they wish to convey
  - *Our holiday was sporty, great weather and fun*
  - *Create clash about whether war is a good thing*
- Inputs: ideas, decisions, available assets
- Actors:
  - author
- Outputs:
  - the message that should be conveyed by the assets

23

Organize

- Process where process artifacts are organized according to the message
  - *Organize a number of 2-page layouts in photobook*
  - *Use semantic graph to select related video clips to form linear presentation of parts of argument structure*
- Inputs: set of assets and annotations (e.g. output from query process)
- Actors: human or machine
- Outputs: document structure with recommended groupings and orderings for assets

24

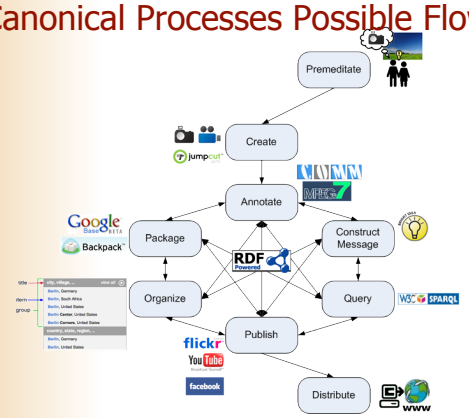
Publish

- Presentation is created
  - associated annotations may be removed
  - *create proprietary format of photobook for upload*
  - *create SMIL file containing videos and timing information*
- Inputs: set of assets and annotations (e.g. output from organize process)
- Actors: human or machine
- Outputs:
  - final presentation in specific document format, such as html, smil or pdf

Distribute

- Presentation is transported to end user, end-user can view and interact with it
  - *photobook uploaded to printer, printed then posted to user*
  - *SMIL file is downloaded to client and played*
- Inputs: published document (output from publish process)
- Actors: distribution hardware and software
- Outputs: media assets presented on user's device

Canonical Processes Possible Flow



Summary

- Community agreement
- Large proportion of the functionality provided by multimedia applications can be described in terms of this model
- Initial step towards the definition of open web-based data structures for describing and sharing semantically annotated media assets

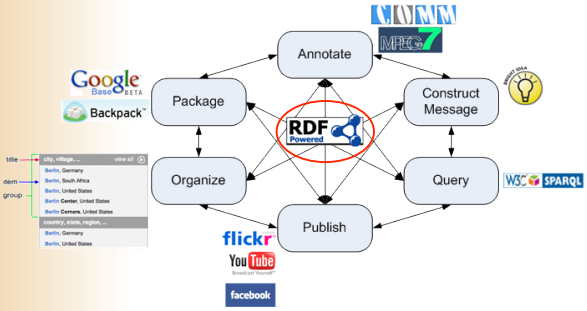
Frequently asked questions

- Complex processes
- Interaction
- Complex artifacts and annotations can be annotated

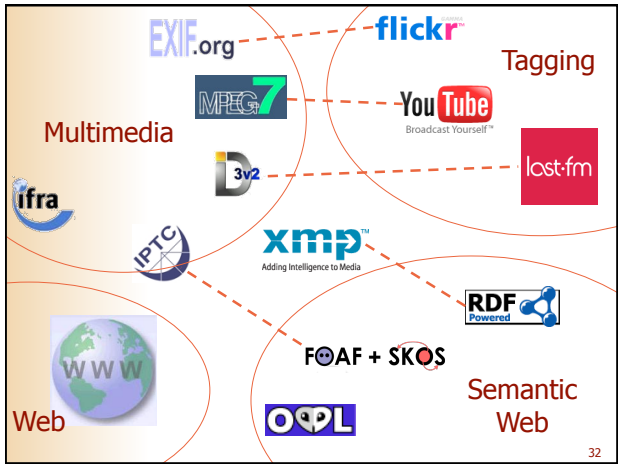
Literature

- Special Issue on Canonical Processes of Media Production http://www.springerlink.com/content/j0l4g337581652t1/ http://www.cwi.nl/~media/projects/canonical/
- Lynda Hardman, Zeljko Obrenovic, Frank Nack, Brigitte Kerhervé and Kurt Piersol: *Canonical Processes of Semantically Annotated Media Production*. In Multimedia Systems Journal, 2008
- Philipp Sandhaus, Sabine Thieme and Susanne Boll: *Canonical Processes in Photo Book Production*. In Multimedia Systems Journal, 2008
- Stefano Bocconi, Frank Nack and Lynda Hardman: *Automatic generation of matter-of-opinion video documentaries*. In Journal of Web Semantics, 6(2), p139-150, 2008.
- Lynda Hardman: *Canonical Processes of Media Production*. In Proceedings of the ACM Workshop on Multimedia for Human Communication - From Capture to Convey (MHC 05), November 2005.

Annotations: the link with COMM



31



32

Image example

Reg1

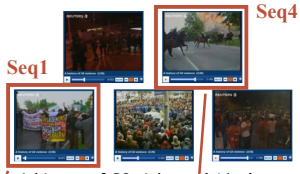
The "Big Three" at the Yalta Conference (Wikipedia)

- Localize a region
  - Draw a bounding box, a circle around a shape
- Annotate the content
  - Interpret the content
  - Link to knowledge on the Web

```
:Reg1 foaf:depicts dbpedia:WinstonChurchill
dbpedia:Churchill rdfs:label "Winston Churchill"
dbpedia:Churchill rdf:type foaf:Person
```

33

Video example



A history of G8 violence (video) (© Reuters)

- Localize a region
- Annotate the content
  - Tag: G8 Summit, Heiligendamn, 2007
  - Link to knowledge on the Web

EU Summit, Gothenburg, 2001

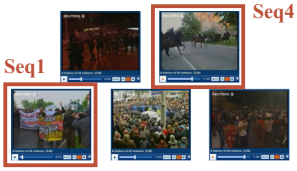
```
:Seq1 foaf:depicts dbpedia:34th_G8_Summit
:Seq4 foaf:depicts dbpedia:EU_Summit
geo:Heilegendamn skos:broader geo:Germany
```

34

Problem

Reg1

Seq1 Seq4



The "Big Three" at the Yalta Conference (Wikipedia)

A history of G8 violence (video) (© Reuters)

- Multimedia objects are complex
  - Compound information objects, fragment identification
- Semantic annotation
  - Subjective interpretation, context dependent
- Linked data principle
  - Open to reuse existing knowledge

⇨ MPEG-7

⇨ D&S | OIO

⇨ RDF

35

COMM: Design Rationale

- Approach:
  - *NOT* 1-to-1 translation from MPEG-7 to OWL/RDF
  - Need for patterns: use DOLCE, a well designed foundational ontology as a modeling basis
- Design patterns:
  - Ontology of Information Objects (OIO)
    - Formalization of information exchange
    - Multimedia = complex compound information objects
  - Descriptions and Situations (D&S)
    - Formalization of context
    - Multimedia = contextual interpretation (situation)
- Define **multimedia patterns** that translate MPEG-7 in the DOLCE vocabulary

36

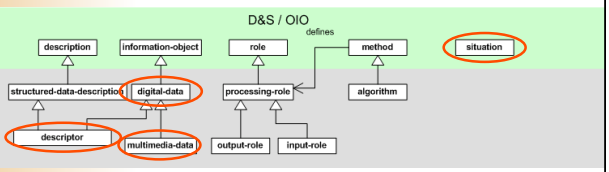
COMM: Core Functionalities

- Most important MPEG-7 functionalities:
  - **Decomposition** of multimedia content into segments
  - **Annotation** of segments with metadata
    - Administrative metadata: creation & production
    - Content-based metadata: audio/visual descriptors
    - Semantic metadata: interface with domain specific ontologies

⇨ Note that all are subjective and context dependent situations

37

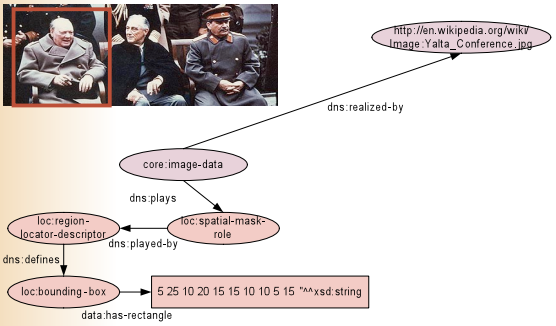
COMM: D&S / OIO Patterns



- Definition of design patterns for **decomposition** and **annotation** based on D&S and OIO
  - MPEG-7 describes digital data (*multimedia information objects*) with digital data (*annotation*)
  - *Digital data* entities are information objects
  - Decompositions and annotations are *situations* that satisfy the rules of a method or algorithm

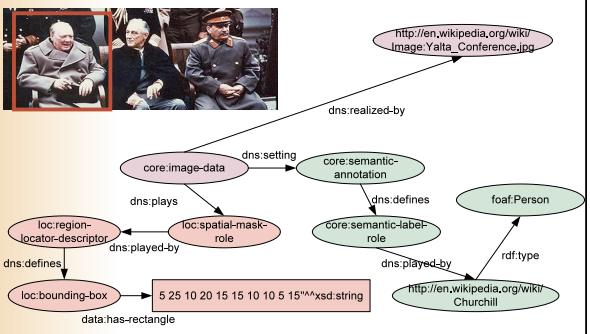
38

Image: Fragment Identification



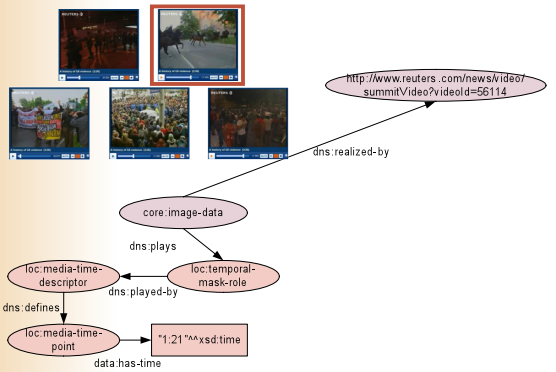
39

Image: Region Annotation



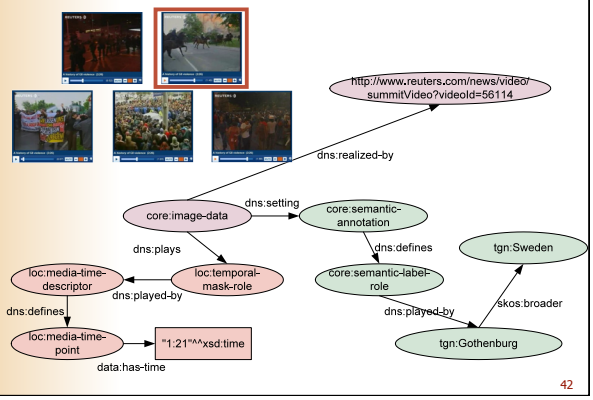
40

Video: Fragment Identification



41

Video: Sequence Annotation



42

43

Implementation

- COMM fully formalized in OWL DL
  - Rich axiomatization, consistency check (Fact++v1.1.5)
  - OWL 2.0: qualified cardinality restrictions for number restrictions of MPEG-7 low-level descriptors
- JAVA API available
  - MPEG-7 class interface for the construction of meta-data at runtime
- http://comm.semanticweb.org/

44

Literature

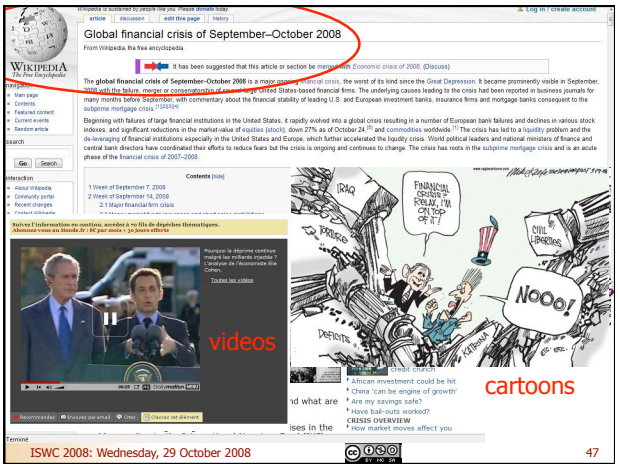
- Michael Hausenblas *et al.*: *Multimedia Vocabularies on the Semantic Web*. W3C Multimedia Semantics Incubator Group Report (XGR), 24 July 2007.
- Raphaël Troncy, Jacco van Ossenbruggen, Jeff Z. Pan and Giorgos Stamou. *Image Annotation on the Semantic Web*. W3C Multimedia Semantics Incubator Group Report (XGR), 14 August 2007.
- Vassilis Tzouvaras, Raphaël Troncy and Jeff Z. Pan. *Multimedia Annotation Interoperability Framework*. W3C Multimedia Semantics Incubator Group Report Editor's Draft, 14 August 2007.
- Richard Arndt, Raphaël Troncy, Steffen Staab, Lynda Hardman and Miroslav Vacura: *COMM: Designing a Well-Founded Multimedia Ontology for the Web*. In 6th International Semantic Web Conference (ISWC'2007), Busan, Korea, November 11-15, 2007.
- Raphaël Troncy, Oscar Celma, Suzanne Little, Roberto Garcia, Chrisa Tsinaraki: *MPEG-7 based Multimedia Ontologies: Interoperability Support or Interoperability Issue?* In 1st Workshop on Multimedia Annotation and Retrieval enabled by Shared Ontologies (MAReSO'2007), Genoa, Italy, December 2007.

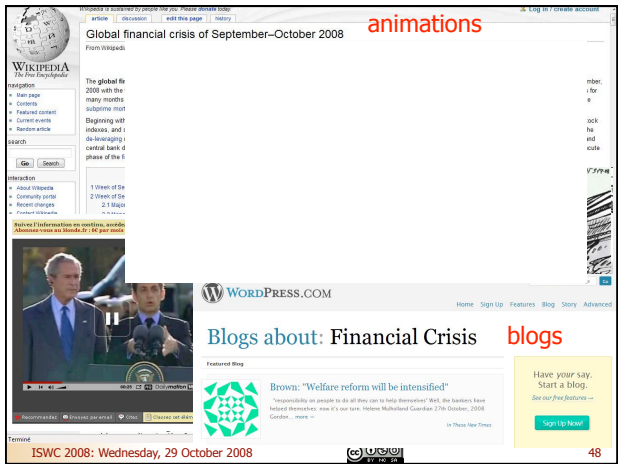
45

Bringing The IPTC News Architecture into the Semantic Web

Raphaël Troncy, <Raphael.Troncy@cwi.nl>

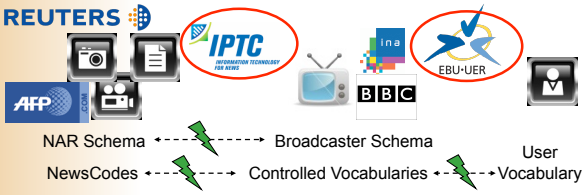
CWI, Semantic Media Interfaces (now Eurecom)





News Workflow Interoperability

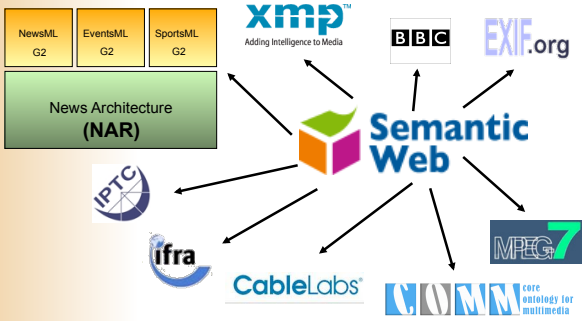
- No integration of media (stories, photo, animation, video)
- Little (or no) context in the news presentation
- Lack of interoperability in the current workflow



Metadata is Key

- (Ultimate) Goal:
  - Provide an environment for ***searching*** and ***browsing contextualized multimedia news*** information
- Required integration:
  - Data: various media, different forms, various sources
  - Metadata: schema integration, semantic models
- Influence and implications of UI:
  - How to **represent** semantic multimedia metadata to facilitate **presenting** information?
  - *in other words* ... What constraints do end-user interfaces put on the modeling of the metadata?

News and Multimedia Formats

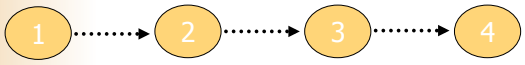


Porting Schemas and Thesauri to the Semantic Web

- Methodologies and tools for building ontologies: ... from scratch
- "SKOSification" of thesauri in the CH domain:
  - preparation, syntactic and semantic conversion, standardization

⇨ Lack of best practices for
modeling ontologies from UML diagrams,
integrating ontologies with various thesauri,
while taking the end-user interface into account

Building a Semantic Web Infrastructure for News



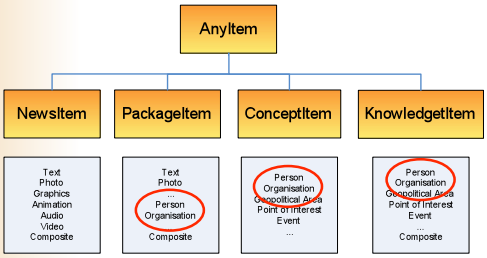
Modeling the NAR ontology

Linking with media ontologies

Building SKOS thesauri

Enriching the metadata

Step 1: Modeling the NAR Ontology



⇨ focus on reuse of XML types leading to multiple repetition resulting in overly complex nested XML structures

Step 1: Modeling the NAR Ontology

- Flattening the XML structure

```
<!--AFP NewsML2 text-photo profile-->
<!--Processed by Xafp1-4ToNewsML2 b11-->
- <newsMessage>
  + <header></header>
  - <itemSet>
    - <newsItem guid="urn:newsml:afp.com:20010101:.DV90996" schema="0.7" version="11" xml:lang="en">
        <catalogRef href="http://iptc.org/std-dev/NAR/1.0/specification/IPTC-TempCatalog-inc_3.xml"/>
      - <itemMeta>
          <contentClass code="ccls:photo"/>
          <provider literal="afp.com"/>
          <itemCreated>2006-07-09T21:20:00Z</itemCreated>
          <modified>2006-07-11T09:14:38Z</modified>
          <fileName>DV90996</fileName>
          <edNote>MOBILE SERVICES OUT</edNote>
        </itemMeta>
      + <contentMeta></contentMeta>
      + <contentSet></contentSet>
      </newsItem>
    </itemSet>
  </newsMessage>
```

NewsItem

↑

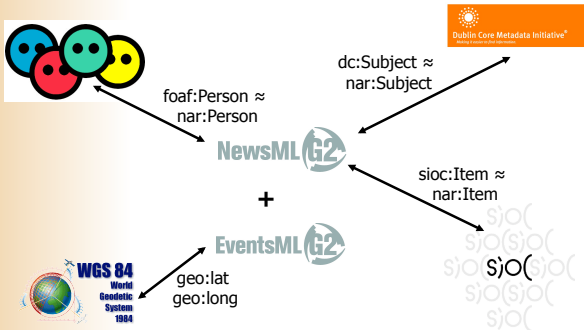
PhotoNewsItem

Step 1: Modeling the NAR Ontology

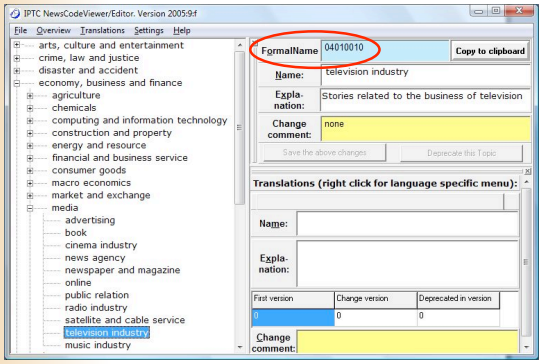
- Modeling unique identifiers
  - Use of dereferencable URIs for any resources (news items + vocabularies)
  - Future: Use of URIs for resource fragments http://www.youtube.com/watch?v=1bibCui3lFM#t=1m45s
- Modeling the provenance of the information
  - Reification
  - Named (and Networked) Graphs

```
{<> nar:subject cat:11002000}
        dc:creator team:md ;
        dc:modified ''2005-11-11T08:00:00Z''.
```

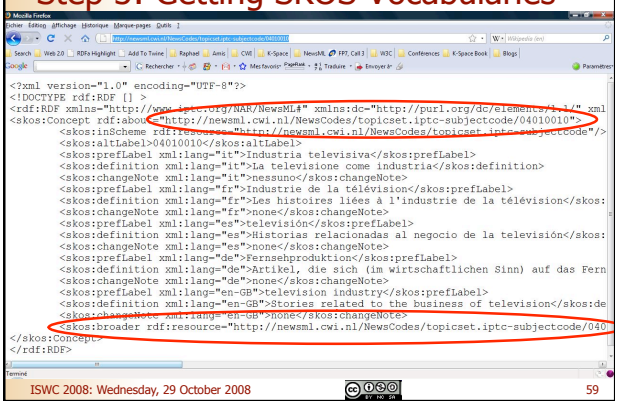

Step 2: Linking with Media Ontologies



Step 3: Getting SKOS Vocabularies

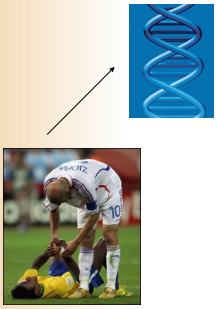


Step 3: Getting SKOS Vocabularies



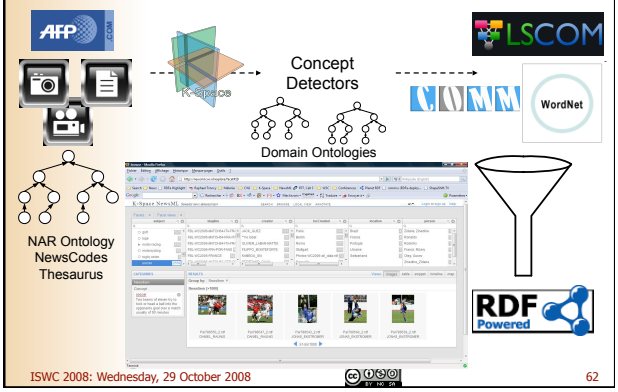
```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rdf:RDF [] >
<rdf:RDF xmlns="http://www.iptc.org/NAR/NewsML#" xmlns:dc="http://purl.org/dc/elements/1.1/" xml
<skos:Concept rdf:about="http://newsml.cwi.nl/NewsCodes/topicset.iptc-subjectcode/04010010">
        <skos:inScheme rdf:resource="http://newsml.cwi.nl/NewsCodes/topicset.iptc-subjectcode"/>
        <skos:altLabel>04010010</skos:altLabel>
        <skos:prefLabel xml:lang="it">Industria televisiva</skos:prefLabel>
        <skos:definition xml:lang="it">La televisione come industria</skos:definition>
        <skos:changeNote xml:lang="it">nessuno</skos:changeNote>
        <skos:prefLabel xml:lang="fr">Industrie de la télévision</skos:prefLabel>
        <skos:definition xml:lang="fr">Les histoires liées à l'industrie de la télévision</skos:
        <skos:changeNote xml:lang="fr">none</skos:changeNote>
        <skos:prefLabel xml:lang="es">televisión</skos:prefLabel>
        <skos:definition xml:lang="es">Historias relacionadas al negocio de la televisión</skos:
        <skos:changeNote xml:lang="es">none</skos:changeNote>
        <skos:prefLabel xml:lang="de">Fernsehproduktion</skos:prefLabel>
        <skos:definition xml:lang="de">Artikel, die sich (im wirtschaftlichen Sinn) auf das Fern
        <skos:changeNote xml:lang="de">none</skos:changeNote>
        <skos:prefLabel xml:lang="en-GB">television industry</skos:prefLabel>
        <skos:definition xml:lang="en-GB">Stories related to the business of television</skos:de
        <skos:changeNote xml:lang="en-GB">none</skos:changeNote>
        <skos:broader rdf:resource="http://newsml.cwi.nl/NewsCodes/topicset.iptc-subjectcode/040
</skos:Concept>
</rdf:RDF>
```


Step 4: Enriching the News Metadata

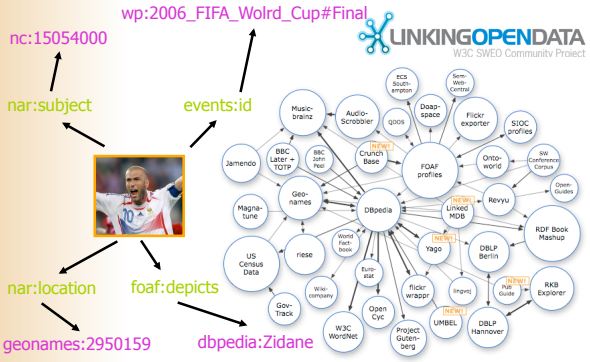


- Concepts/Entities that are subject of news
  - Thematic categories
  - People
  - Organizations
  - Geopolitical Areas
  - Points of Interest
  - Events
  - Products or artefacts

Step 4: Enriching the News Metadata



Web of Data and Linked Data



Presenting News Information

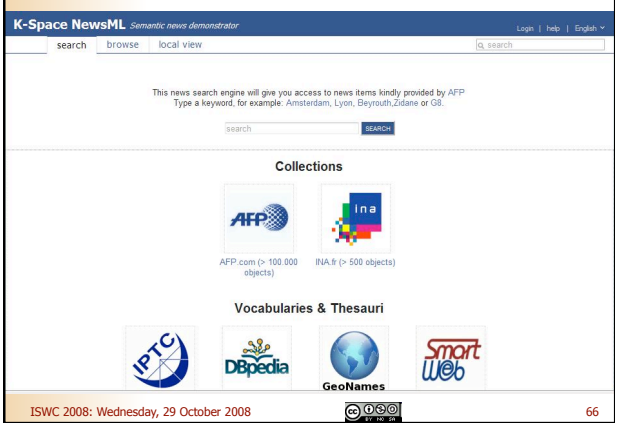
- Dimensions used for searching news items
  - When *time* 10/07/2006
  - Where *location* Paris
  - What *is depicted* J. Chirac, Z. Zidane
  - Why *event* WC 2006
  - Who *photographer* Bertrand Guay, AFP

  } **Metadata**

Semantic Search of Multimedia News

Description	Number of RDF Triples
General Ontologies: NAR, DC, FOAF	7,336
Domain Specific Ontologies: football	104,358
Thesauri: newscodes	34,903
DBpedia, Geonames	53,468
AFP News Feed (June/July 2006)	804,446
AFP Photos (June/July 2006)	61,311
INA Broadcast Video (June/July 2006)	1,932
Total	1,067,754

Powered by ClioPatria 1.0 alpha 3

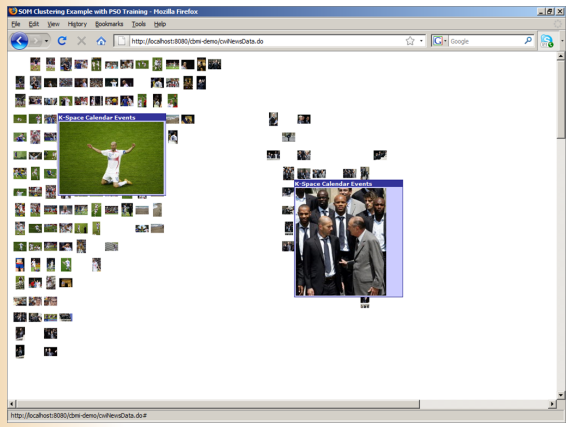


Conclusions

- 4-Steps methodology for building an ontology-based news infrastructure
  - UML-2-OWL: Flatten XML structure, Identify all resources
  - SKOS-ify existing thesauri and use the Web of Data
  - Reuse what is there ... Expose what you make
- Enrich metadata with text and visual analysis
  - Provide new dimensions (facets) for browsing the data
    - Ex: distinguish field images vs stadium and street images with a grass detector for the World Cup dataset

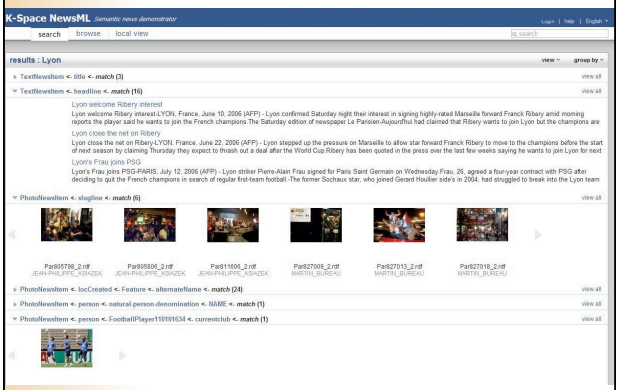
Literature

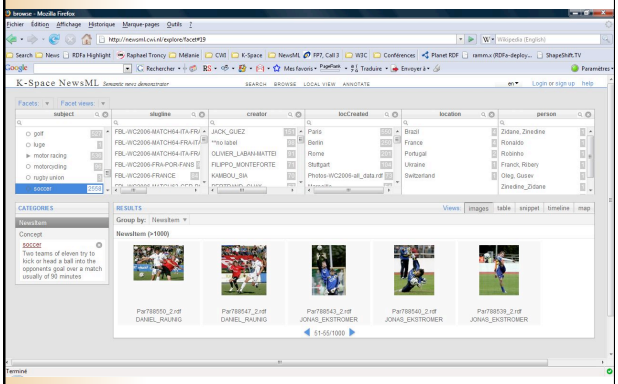
- Michiel Hildebrand, Jacco van Ossenbruggen and Lynda Hardman: */facet: A Browser for Heterogeneous Semantic Web Repositories*. In 5th International Semantic Web Conference (ISWC'2006), pages 272-285, Athens (GA), USA, November 5-9, 2006.
- Jan Wielemaker, Michiel Hildebrand, Jacco van Ossenbruggen and Guus Schreiber: *Infrastructure for thesaurus-based search and annotation: evaluating the standards*. In 7th International Semantic Web Conference (ISWC'2008), Karlsruhe, Germany, October 26-30, 2008.
- Raphaël Troncy: *Bringing the IPTC News Architecture into the Semantic Web*. In 7th International Semantic Web Conference (ISWC'2008), pages 483-498, Karlsruhe, Germany, October 26-30, 2008.
- Raphaël Troncy, Lynda Hardman, Jacco van Ossenbruggen and Michael Hausenblas: *Identifying Spatial and Temporal Media Fragments on the Web*. In W3C Video on the Web Workshop, San Jose (California) and Brussels (Belgium), December 2007.
- W3C Video on the Web Activity, April 2008 http://www.w3.org/2008/01/video-activity.

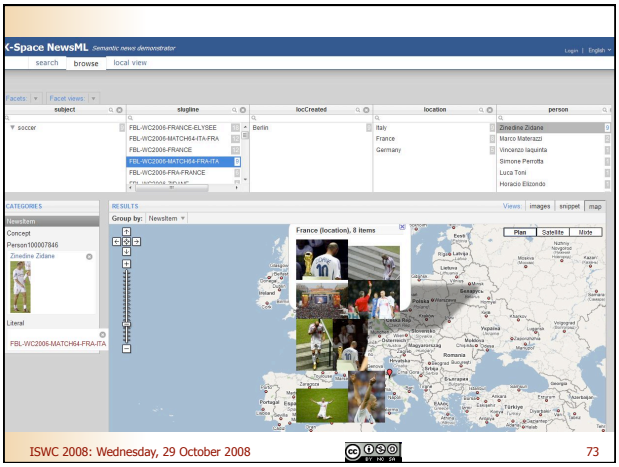


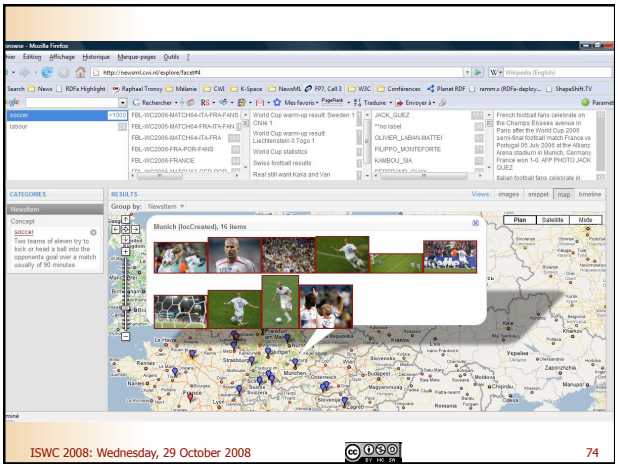
Credits

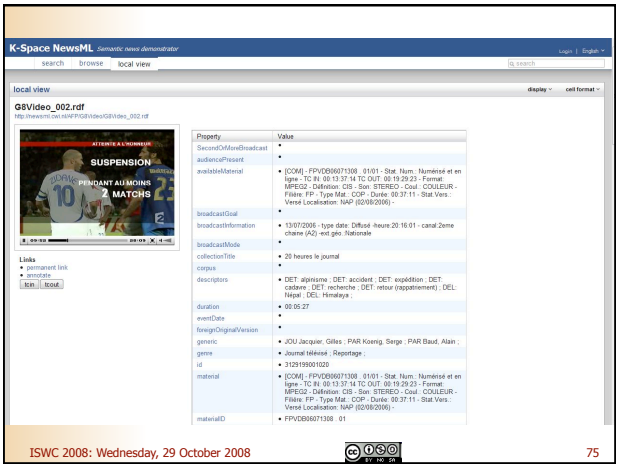
- Datasets:
- People:
- More info: **http://newsml.cwi.nl**










K-Space NewsML
Semantic news demonstrator
search
browse
local view
G8Video_002.rdf
SUSPENSION
PENDANT AU MOINS
2 MATCHS
Property
Value