The News Agency of today and the vision for tomorrow

Current practices in the news agencies; tools and issues to be addressed by new technologies

Stéphane Guérillot – AFP
IPTC – Chairman of the Board
stephane.guerillot@afp.com
chair@iptc.org

December 2006

AFP

IPTC – few words…

AFP – Introduction to the business

Challenges

Constraints

Existing solutions

Projects under evaluation

Conclusions

samt
2006

AFP

A few words about IPTC …

- The International Press Telecommunications Council (IPTC) was founded in 1965 to safeguard the **telecommunication interests** of the world's press.

- IPTC develops and maintains technical **standards** to improve the **free exchange of news,**
adopted by every major news provider world wide.

- IPTC standards are opened but they are published under copyright.

  - NITF – IIM – NewsML™ - SportsML – News Codes …

  - **NewsML is a media independent standard for describing news. NewsML defines an XML based language for expressing the structure of content items, associated metadata, and relationships between content items, throughout their lifecycle.**

  - **News Codes is a set of topics to be assigned as metadata values to news objects. This allows for a consistent coding of news metadata over the course of time.**

- a **News Agency** is a **Content Provider**
- Could be:
  - For **General** News
  - **Specialised** (in Sports or Business or People…)
  - With a **National** perspective
  - With an **International** coverage
  - In **one** or **multiple** languages
  - In **one Media** (Text, Photo, Graphics, Video…)
  - With coverage in **more than one Media**
  - With **Clients** in **one country** or worldwide

samt
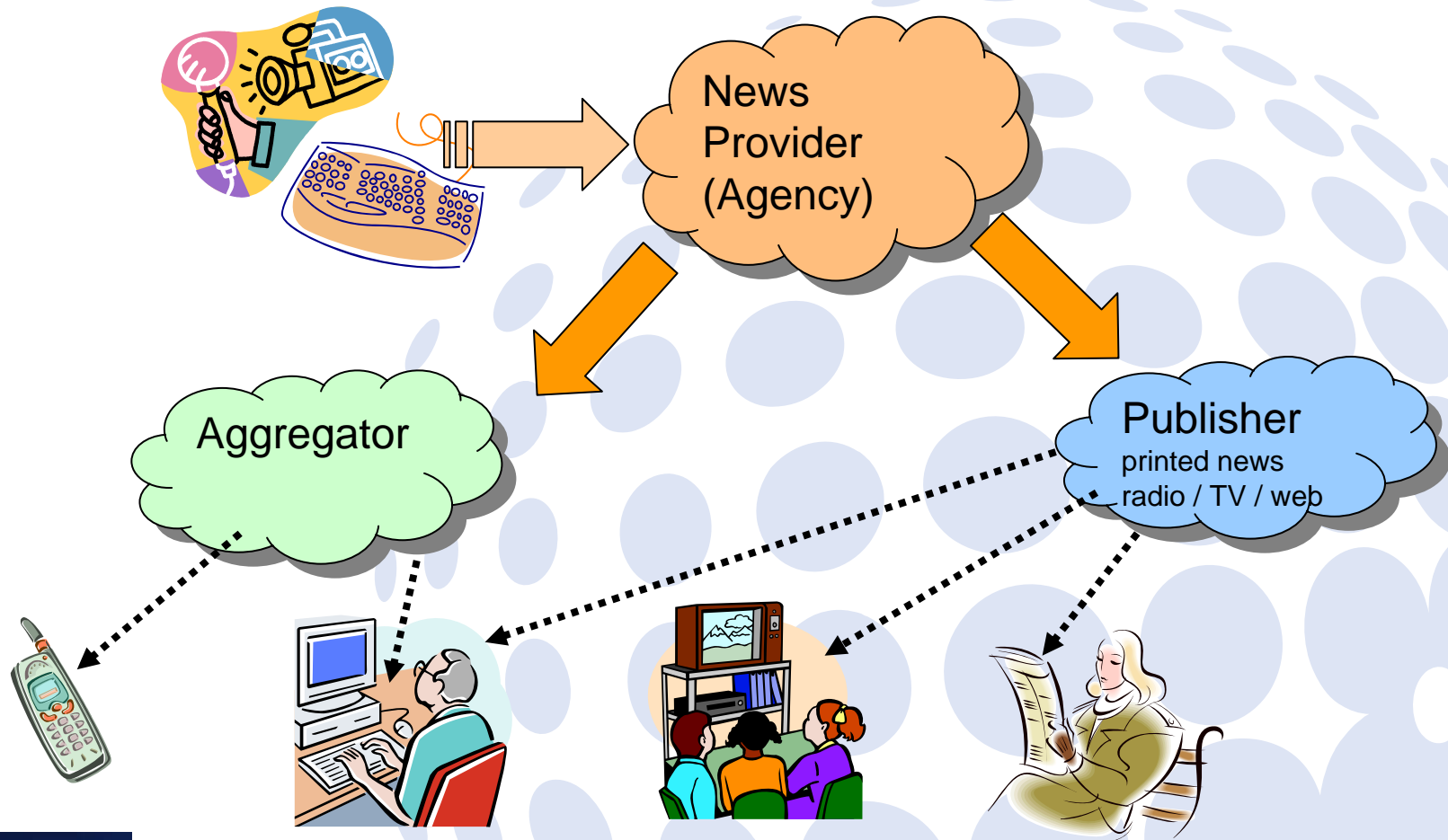2006

AFP

Agence France Presse

worldwide multilingual and multimedia news coverage, 24 hours/day, 7/7

- Founded in 1835, AFP is one of the 3 worldwide news agencies.

- 1.200 journalists, 200 photo-reporters and 2.000 stringers based in 165 countries.

- AFP provides a 24/7 worldwide multilingual news coverage and produces daily more than 2 millions words,1500 photos, 50 news graphics, as well as video, audio and multimedia content.

- Main languages used are French, English, German, Spanish, Portuguese and Arabic.

- Its news, pictures and multimedia services are edited at regional news centres in Paris, Nicosia, Washington, Montevideo and Hong Kong.

samt
2006

AFP

"The Agency shall be independent of all political, economic, philosophical or religious influence" (Article 2 of the Agence France-Presse statute, 1957)

- in a news agency, **speed** and **reliability** are the key words

- **Accuracy** is the guideline

- We aim to provide a news service which gets to the heart of the issues and provides analysis of world events.

- As well as general news from around the world, we offer economic and financial news, sports coverage, human interest stories, celebrity news, science, culture, new technology, lifestyle and offbeat items.

samt
2006

AFP

A standard news Item

Britain-media-BBC-Internet, INTERVIEW

BBC's global future is on the Internet: corporation's chief

by Robin Millard

=(PICTURE)=

LONDON, June 11, 2006 (AFP) - BBC director-general Mark Thompson wants to take advantage of the latest technology to turn one of the world's foremost broadcasters into a truly global media brand.

He insists that the Internet is the future for the 84-year-old British Broadcasting Corporation but warns that it has to be globally relevant to ensure it does not get left behind.

The BBC, funded mainly by a television licence fee levied on British homes, already reaches more than 250 million people around the world.

"I think we are one of the very few, perhaps the only European media brand which could become a global consumer media brand," Thompson told AFP in an interview.

"We were set up in the 1920s as a set of domestic radio services but also there was this idea right from the start -- our motto is 'Nation shall speak peace unto nation' -- that we would have global potential.

"We now have the technological means of doing that.

….

"We're there in people's portfolio of news and there are all sorts of parts of the world where we are effectively the only source of news."

rjm/dmh/bm

samt 2006

METADATA:

News Provider: AFP
Date: 11 June 2006
News item revision: 1
Location: Country: United Kingdom / City: London
Slug line: Britain-media-BBC-Internet, INTERVIEW

Britain-media-BBC-Internet, INTERVIEW

BBC's global future is on the Internet: corporation's chief   ← **HEADLINE**

by Robin Millard   ← **BYLINE**

=(PICTURE)=   ← **REF TO AN EXISTING PHOTO**

LONDON, June 11, 2006 (AFP) - BBC director-general Mark Thompson wants to take advantage of the latest technology to turn one of the world's foremost broadcasters into a truly global media brand.

AFP

Yahoo!   My Yahoo!   Mail   Make Yahoo! your home page

YAHOO! NEWS   Sign In
New User? Sign Up

News Home − Help

| Home | U.S. | Business | **World** | Entertainment | Sports | Tech | Politics | Science | Health | Travel | Most Popular |

| Middle East | Europe | Latin America | Africa | Asia | Canada | Australia/Antarctica | Kevin Sites |

Search: [   ] [All News & Blogs ∨] [Search] Advanced

World - AFP

AFP

Riots, gunfire after US troops shoot dead four

1 hour, 10 minutes ago

KABUL (AFP) - The Afghan capital erupted in gunfire and riots after US troops shot dead at least four people following a traffic accident, with angry crowds shouting "Death to America."

View: Headlines Only | Include Summaries | **Include Photos**

Indonesia struggles to cope as quake toll nears 5,000

Mon May 29, 4:15 AM ET

BANTUL, Indonesia (AFP) - Indonesia struggled to cope with the scale of the earthquake disaster, but help was slowly starting to arrive for thousands of homeless survivors and injured who have overwhelmed hospitals.

East Timor president appeals for end to

MORE IN YAHOO! NEWS

· AP
· Reuters
· **AFP**
· CSMonitor.com
· OneWorld.net
· NPR

- We all are under a huge flow of information
- More difficult for professionals
- A jungle for the layman
  - France, per day:
    - > 1200 news items
    - from 600 to 1000 pict.
  - Archives:
    - Text: 7 M news stories per managed language
    - IF: # 5 M photos

- AFP Platforms :
  - Our own Editorial Systems
  - Our open systems for clients:
    - **At their premises**
    - **On our Web sites such as Image Forum**

- At Content Aggregators or Syndication

- In news rooms on various editorial systems:
  - Wedia (ex Datox), Eidos Media, Protec, MVS, Ailink, Unisys, Trias, Sinedita, Fotoware, ATC …

- On the WEB

A News Agency's business today has changed:

To

Adding value to content of various media types

and

Selling and delivering customised content to various types of customers

… with a direct impact in our internal organisation

- news gathering
- news distribution
- building/providing a news repository

- Answer users' needs:
  - Search, Alerts, Browse
  - Clusters text-photo for a specific event
  - Real time or archives
  - In-house (AFP) users or for our clients
- How
  - Enrich our Content by use of METADATA and Concepts
  - Improve our WEB platforms
  - Allow for our Clients to improve their own platforms
    - **In distributing enriched Contents**
    - **Give access to management processes for those contents**
    - **Implement and provide APIs for an easier access to our platforms.**

- Slugs or Slug line:
  - Semantic-and-Digital-Media-Technology (text)
  - ATHENS-SAMT2006-ALAN-SMEATON (photo)
- Categorisation in a semi automatic mode
- Extraction of organisation names to add ISIN numbers (Business section)
- Categorisation by the editor or photographer (photo dept.)
- Persons Names additions (photo archives)
- Topic Filtering (topic engine at the distribution end)

■ **We use some IPTC latest standards for our feeds such as the News Codes**

BUT

■ **Relying on the willingness of the journalists to enrich their documents, which means increasing their load of work…. Is not satisfactory**

This is not enough!

■ **We need new tools to automate this enrichment in a new environment by implementing automatic taxonomy, ontology and pre-indexing**

- Read through lists (editors in the daily news papers)
- Search (active mode)
- Alerts (passive mode)
- Browse (interactive mode)
- Basics:
  - Full text
    - **Simple words -> noise**
    - **Sort / rating (Web engines) -> inefficient**
  - Multi-criteria : not mastered
  - Keyword based
  - Semantic (navigation)
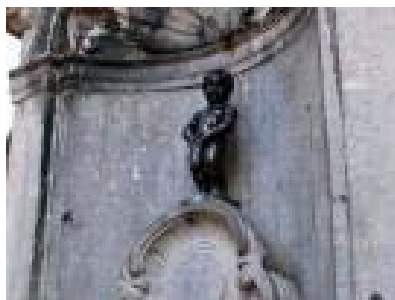
- **Subjects = concepts**
- Themes (IPTC News Codes)
- Named entities
  - Persons
  - Organisations
  - Locations
  - Events
  - Works / Products
- Ex: Uma Thurman, Tiger Woods, Beckham, Zidane, Ronaldo, Le Bourget Air show, Turismo, Mode, Arts, Sweden, Belgique, World Cup, WC, 4th of July …

Multi-criteria Search => better results

Named entities, themes, dates, expressions

Source - Factiva 2.0
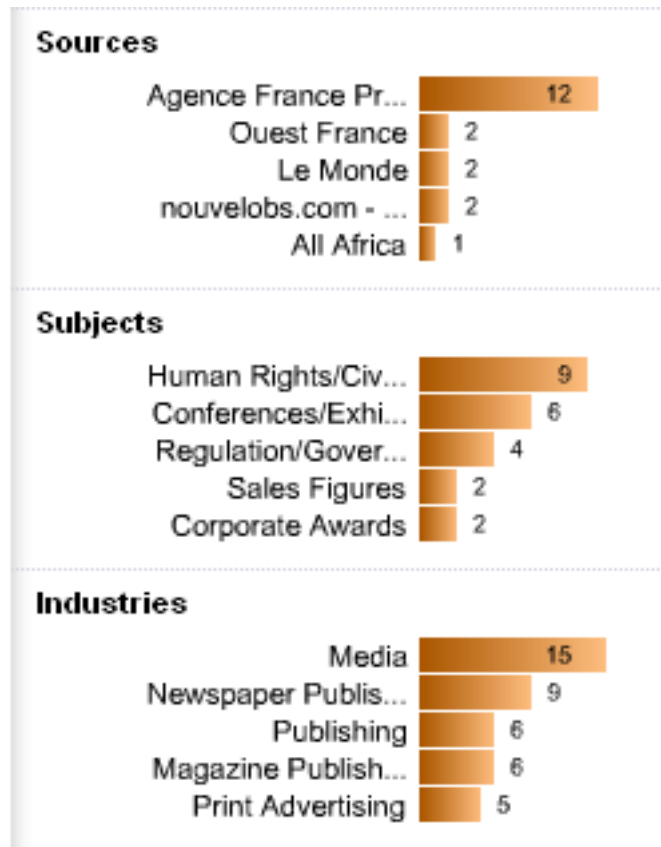
- 1.300 themes (taxonomy of subjects IPTC)
- 5.000 organisations
  (dictionary from the B&F)
- 230.000 persons
  (dictionary from the Photo archive dept.)
- 75.000 locations
  (dictionary from the Photo archive dept.)
- Dictionaries
  - Completes

  But
 - Without any links between concepts

Categorisation

- Phase 1 (Spring 2006) :
  - Selection of 75 "categories" or Subjects
  - Extraction from the archives in the French service and the Business and Finance service (2004)
  - Selection of a learning Corpus: 20 to 50 samples per themes
  - Those samples are taken from our Topic filter engine

Our Current Project

- In a first stage, an entirely automatic Prototype

Desk	→	Enrichissement	→	Distribution

- In a second stage insertion of the Engine in Production cycle
- in a semi-automatic mode – control by the Desk

| Production	→	Enrichissement	→	Desk	→	Distribution
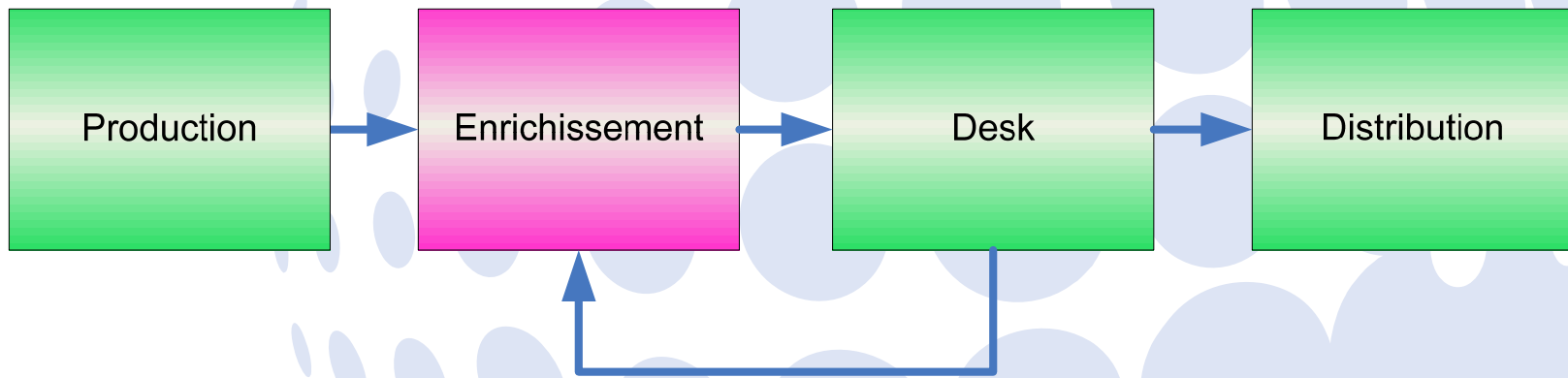
Control of the entities as extracted

```
┌──────────────┐      ┌──────────────┐      ┌──────────────┐
│              │      │              │      │              │
│  Production  │ ───▶ │ Enrichissement│ ───▶ │ Distribution │
│              │      │              │      │              │
└──────────────┘      └──────────────┘      └──────────────┘
                              ▲
                              │
                      ┌──────────────┐
                      │              │
                      │   Base de    │
                      │ connaissance │
                      │              │
                      └──────────────┘
```
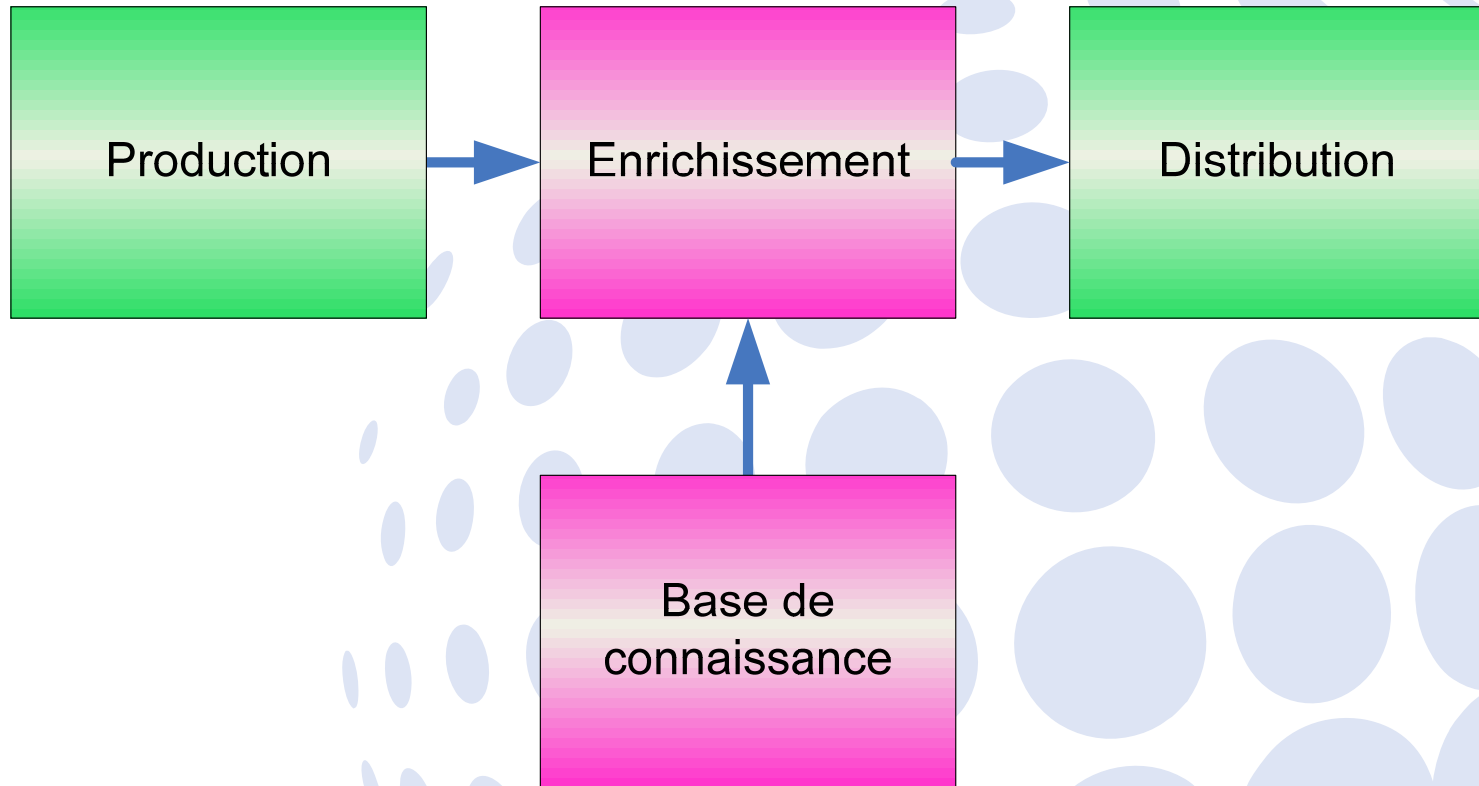
- **Business related**
- **Class definitions and relations between classes**
- **Semantic browsing in a network of concepts**

<->	Person	Organisation	Location	Event	Work
Person	Parenthood	Affiliation Management	Localisation Management (place)	Organisation Participation	Creation
Organisation		Subdivision	Localisation Management (place)	Organisation Participation	Creation
Location			Subdivision	Localisation	Localisation
Event				Subdivision	Presentation
Work					Subdivision
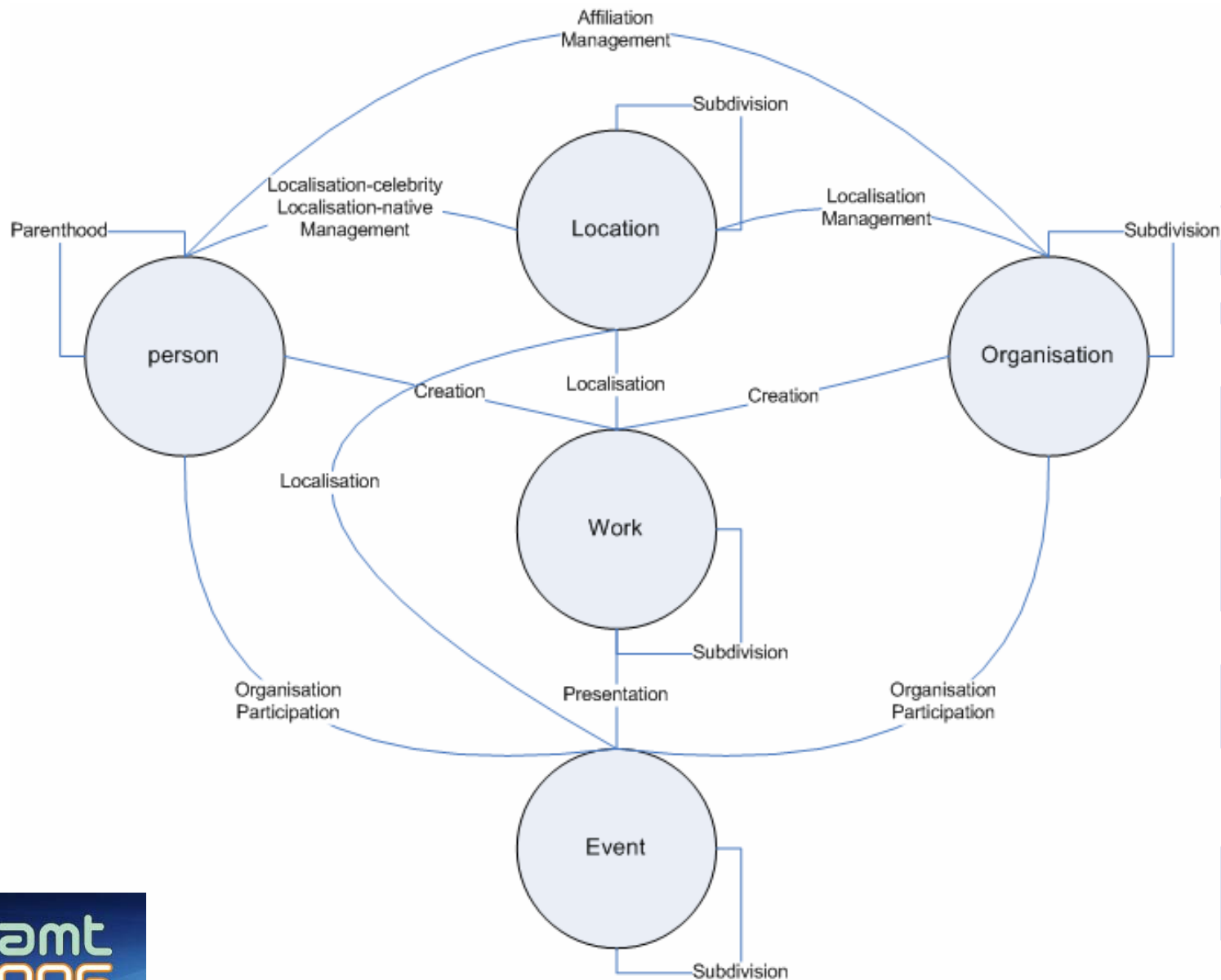
- Phase 2 (spring 2007) :
  - Use of NewsML 2G – draft version
  - 380 categories (draft IPTC News Codes 2G)
  - Extraction archives French services to include years 2000, 2003 et 2005
  - Corpus knowledge database: 200 to 500 samples per theme
  - Selected from our productions by application of filters on the Slug lines and the enhanced slugs including Subject codes from the IPTC lists.
  - Sort and rating given by additional tools
- Implementation at the level of our online photo database

Conclusion (1)

- Traditional "push content" from the news provider to the customers is replaced by interactive push/pull services (Customers select and retrieve what they need)
  - We are no more generating a "Service" to "Professionals!

- We have to implement those features on our different Media but requirements for multi-media news add complexity to news technology

- More metadata is required to describe and categorize content but complexity without mastering is useless
  - news providers are challenged to do this by automated software
  - But we cannot forget the Journalist at the production end !!!

- To get trough the Information maze semantic is one of the keys which will at the end carry the value of the content

We Must not forget that:

- Our life is focussed on real time news coverage and in some cases not the "minute" but the "second" is the reference…

- Journalists have to work to produce these new services

And

- The key will be in the links between NewsItems of the same nature or of different nature leading to Multimedia content management as early as possible in the production chain

Thank you for your attention

See also

- **www.afp.com**
- **www.iptc.org**