



Associated Press

the **essential** global news network

Information Management at the Associated Press

Joel Summerlin

Deputy Director, Information Standards

May 2009



A faster way to browse the web

[Install Google Chrome](#)

[Advanced Search](#)
[Preferences](#)
[Language Tools](#)



[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

[Make Google my homepage](#)

©2009 - [Privacy](#)

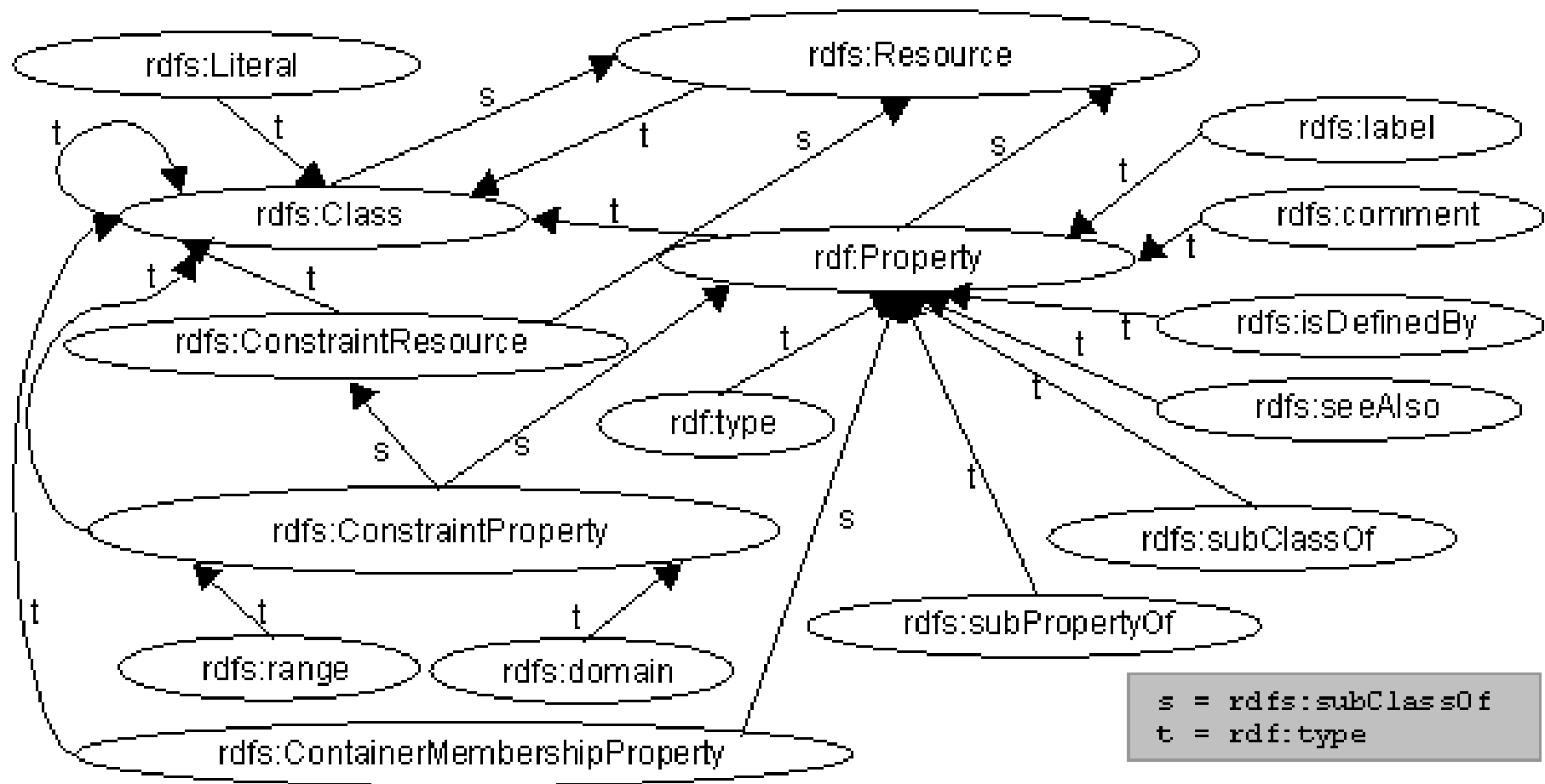
Taxonomies are dead!

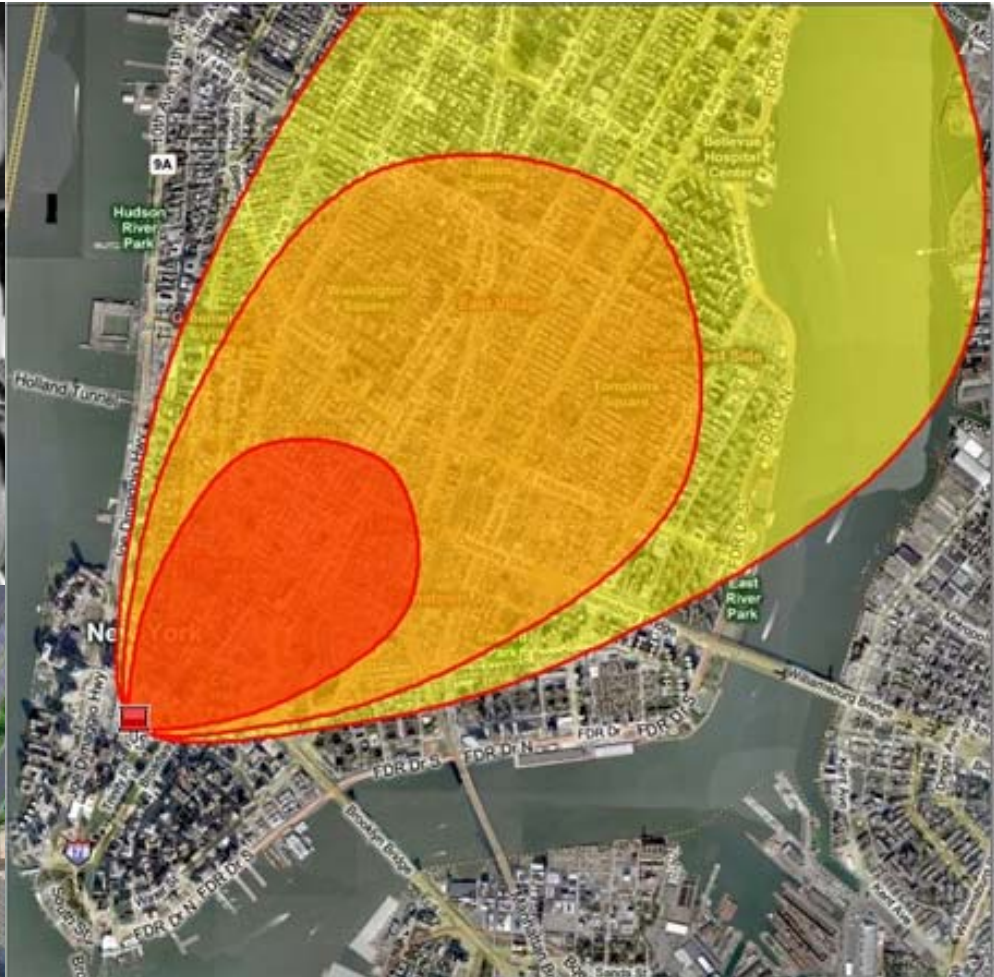
Long live taxonomies!

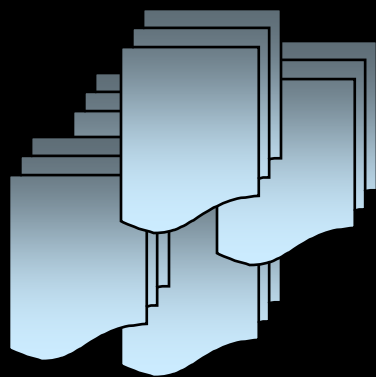
PRIZE	all	filter	lock	GENDER	all	filter	lock	COUNTRY	all	filter	lock	DECADE	all	filter	lock	YEAR	all	filter	lock
medicine				male				United States of America				2000s				2004			
			182				701												
physics								United Kingdom				1990s				2003			
			167																
chemistry				female			34	France				1980s				2002			
			139																
peace								Germany				1970s				2001			
			111																
literature								Federal Republic of Gern				1960s				2000			
			101																
economics								Switzerland				1950s				1999			
			55																
								Sweden				1940s				1998			
								the Netherlands				1930s				1997			
								USSR				1920s				1996			
								Italy				1910s				1995			
								Denmark				1900s				1994			
								Austria								1993			
								Belgium								1992			
																1991			
																1990			

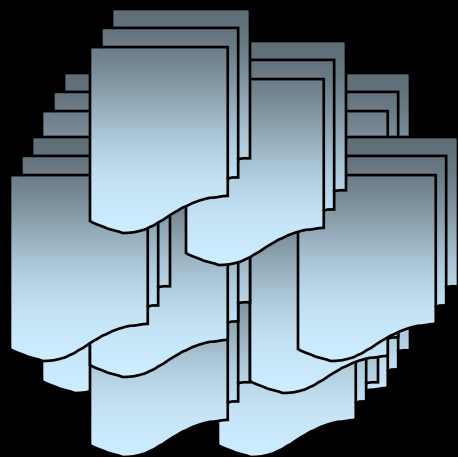
PLEASE MAKE A SELECTION ABOVE.

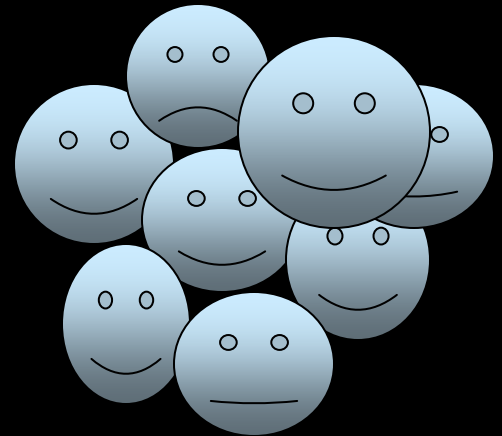
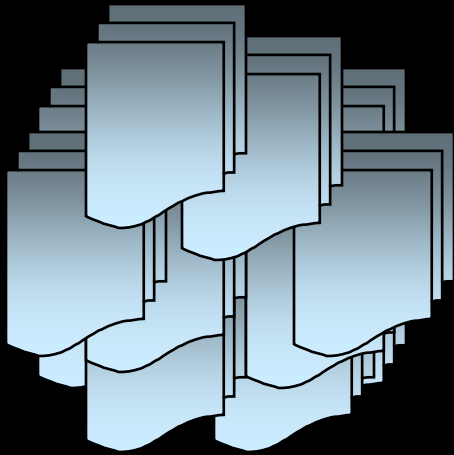
addition admin applications being blog ch **change**
city delicious design different done facebook
fact feature flickr functionality greenear hurricane
information interesting interface itp
katrina level needed **networking** nice
orleans page people perhaps possible **posted** projects
providing quite simply **site social** software
something surprised tabs **tags** today **ui**
users web york

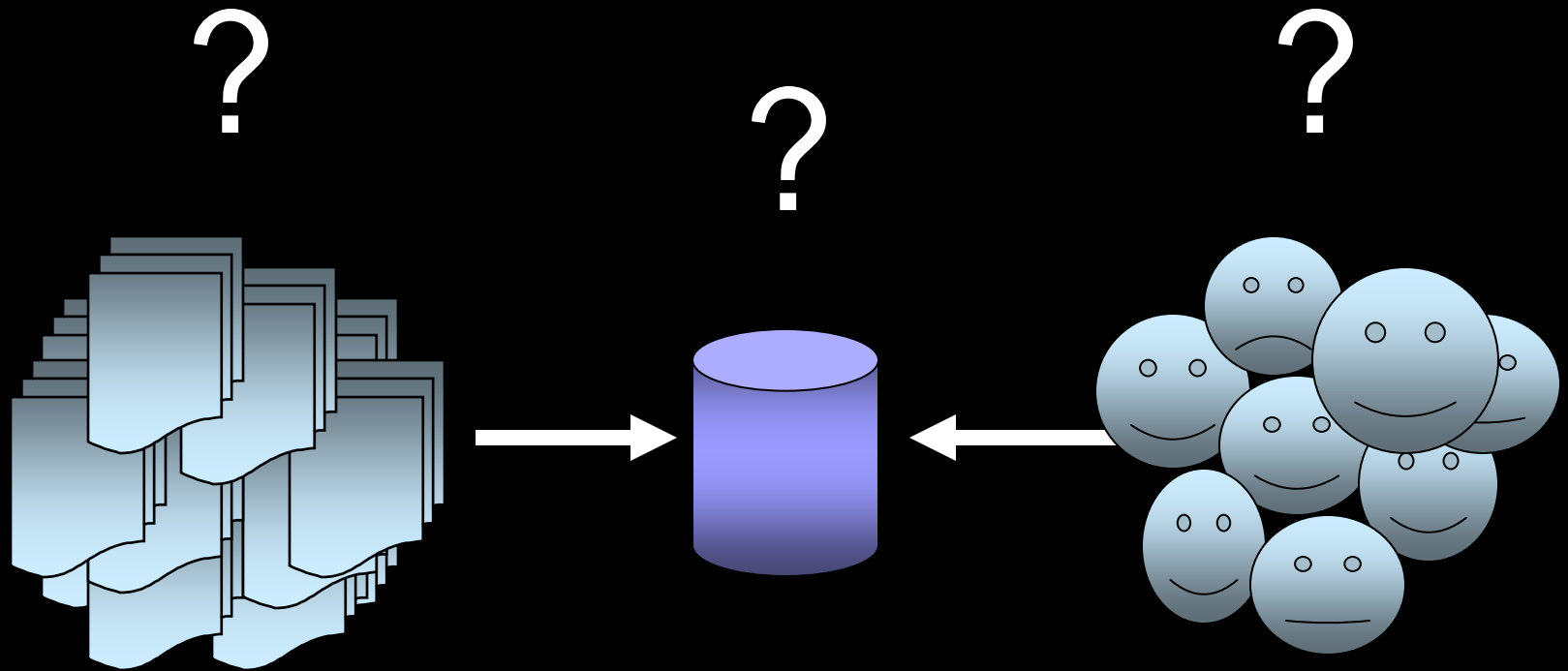






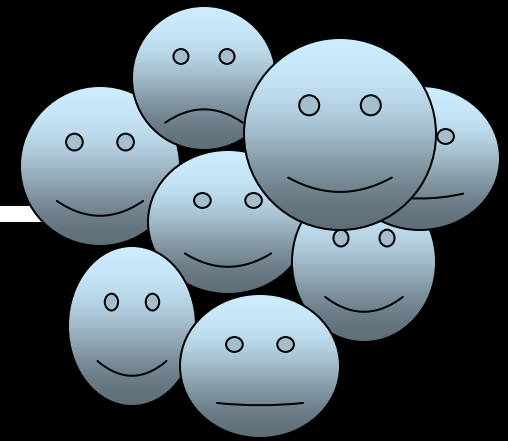
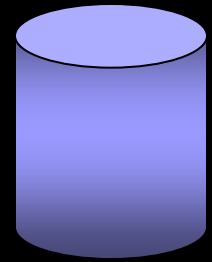
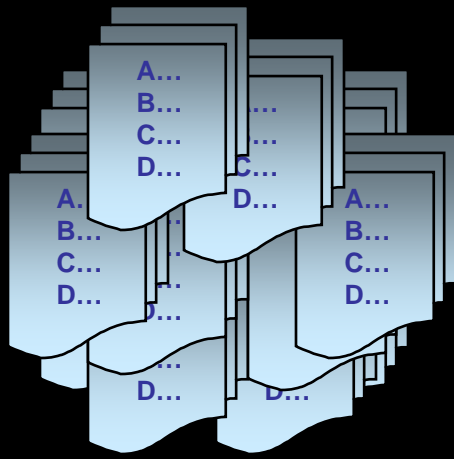




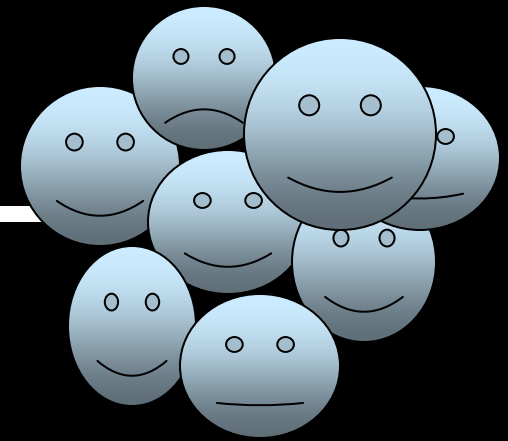
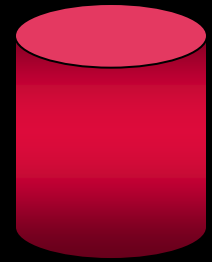
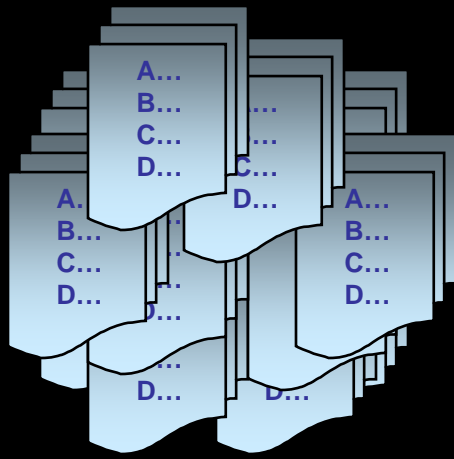


\$\$\$

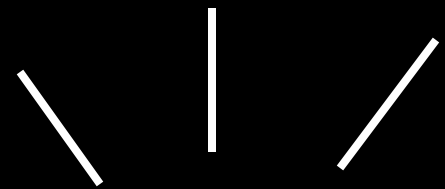
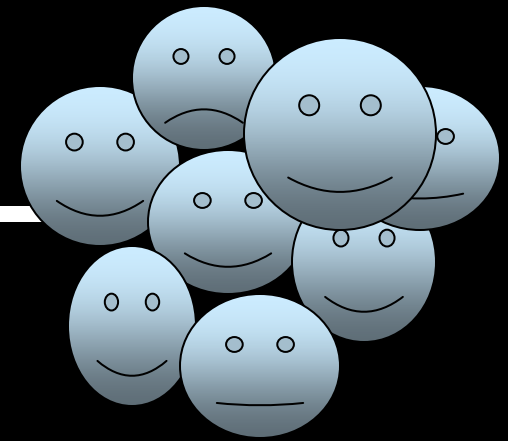
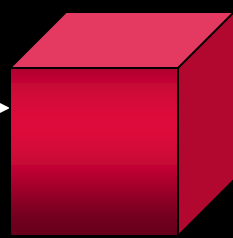
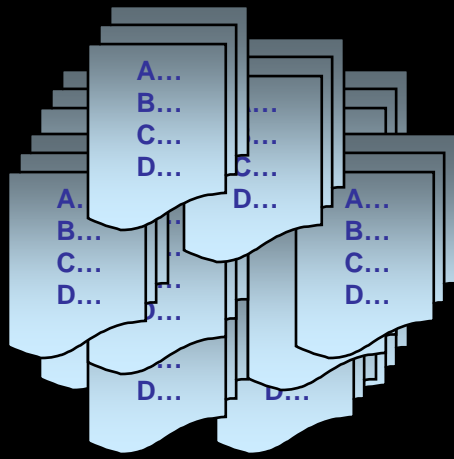
A...
B...
C...
D...



A...
B...
C...
D...

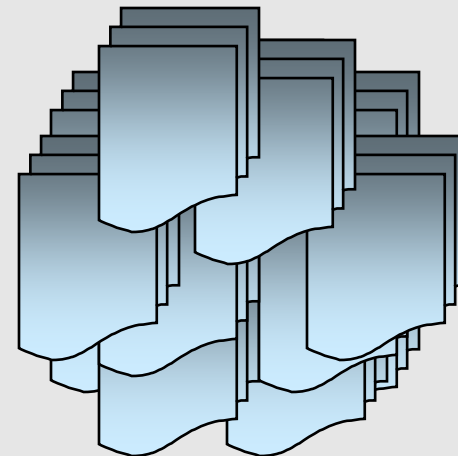


A...
B...
C...
D...



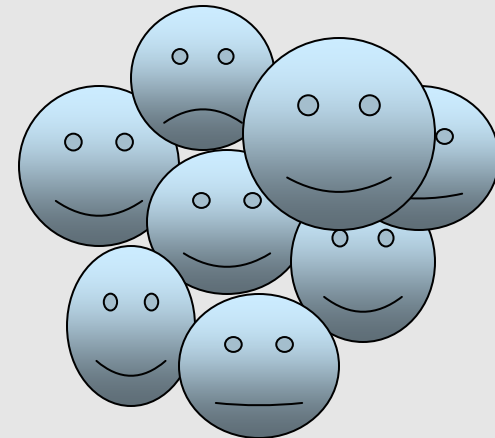
AP CONTENT

- > Multiple sources: *AP, Members, 3rd parties*
- > Many formats: *Text, Photos, Video, etc.*
- > Vast quantity
- > Multiple languages
- > 24/7 news delivery
- > Time-sensitive
- > Usage restrictions



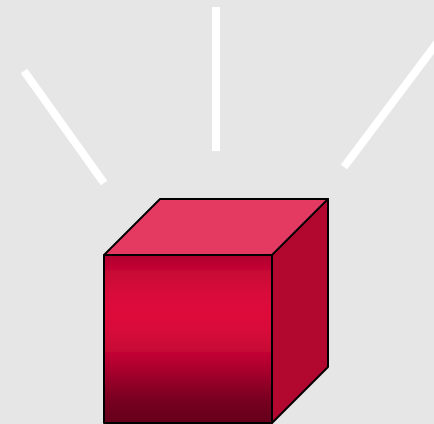
AP CUSTOMERS

- > AP Members (1500 US daily newspapers)
- > Other news providers
 - 8500 international subscribers
 - 5000 radio and TV outlets
 - 850 AP radio network affiliates
- > Web sites and news aggregators
- > Individuals



AP CONTENT DELIVERY

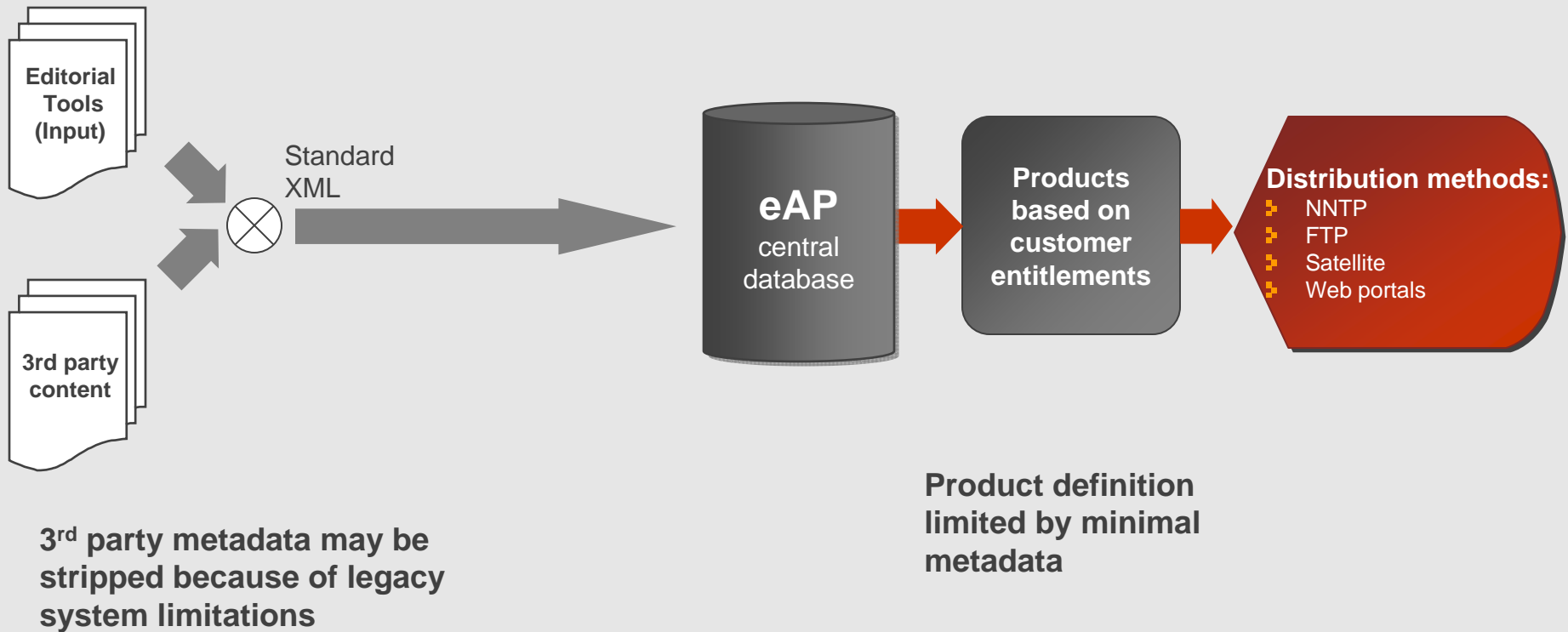
- > Defined products
- > Data feeds
- > Portals
- > Web pages
- > Mobile devices



2005: AP Content Management without Standard Metadata

Metadata is minimal and varies across formats

Topical product definition limited to very high level categorization: National, International, Sports, Entertainment



Metadata Standards: 3 key pieces

- > **Standard schema**: agreement on what data to collect
- > **Standard values**: common language to describe concepts and the relationships between them
- > **Metadata application**: applying the standard values efficiently and consistently

Standard Schema: Types of Metadata

- > Attribution (e.g. byline, source)
- > **Classification** (e.g. subjects, entities)
- > Date (e.g. date created, date published)
- > Description (e.g. dateline, headline)
- > Distribution (e.g. sales classification, product)
- > File Characteristics (e.g. file size, format)
- > Linking (e.g. inline links, related content)
- > Publication (e.g. version, status, priority)
- > Rights (e.g. copyright, restrictions)
- > Legacy Metadata (e.g. ANPA categories)

Standard Schema: APPL

```

- <NewsLines>
  <Title>LG-Dress-for-success-0523</Title>
  <HeadLine>Dress for success</HeadLine>
  <OriginalHeadLine>Dress for success</OriginalHeadLine>
  <CopyrightLine>Copyright: For copyright information, please check with the distributor of this item, McClatchy Tribune (MCT).</CopyrightLine>
  <KeywordLine>LG-Dress-for-success-0523</KeywordLine>
</NewsLines>
- <AdministrativeMetadata>
  <Provider>AP</Provider>
  <Creator>AP</Creator>
  <Source Type="ThirdParty">McClatchyTribune (MCT)</Source>
  <TransmissionSource>EapText</TransmissionSource>
  <TransmissionSource>NotMemberFeed</TransmissionSource>
  <ProductSource>DataFeatures</ProductSource>
  <ItemContentType>Syndicated</ItemContentType>
</AdministrativeMetadata>
<RightsMetadata />
- <DescriptiveMetadata>
- <SubjectClassification System="Editorial" Authority="AP Category Code">
  <Occurrence Id="n" Value="n" />
</SubjectClassification>
- <SubjectClassification SystemVersion="1" AuthorityVersion="825" System="Teragram" Authority="AP Subject">
  <Occurrence Count="43" Id="42a74195df1645dda692dba5db2416d4" Confidence="1.000000" Value="Primary and secondary education"
  ParentId="1af99ec3cb954ff4b349b32d60d0376d" ActualMatch="true" />
</SubjectClassification>
- <SubjectClassification SystemVersion="1" AuthorityVersion="825" System="Teragram" Authority="AP Subject">
  <Occurrence Count="4" Id="60a40e4888f91004804e8503ef5018bd" Confidence="1.000000" Value="Stock markets"
  ParentId="60a31c1888f91004804d8503ef5018bd" ActualMatch="true" />
</SubjectClassification>
- <SubjectClassification SystemVersion="1" AuthorityVersion="825" System="Teragram" Authority="AP Subject">
  .....
```

Standard Values: Controlled Vocabularies

> Subjects

- Business
- Entertainment
- Environment and Nature
- General News
- Government and Politics
- Health
- Lifestyle
- Science
- Social Affairs
- Sports
- Technology

> Entities:

- People
- Companies
- Organizations
- Events
- Geography

> System Vocabularies:

- Content type
- Usage restriction type
- Sales classification

Standard Values: Rich Relationships

- > Standard taxonomic relationships, with explicit types
 - Hierarchical: “is a”; “is a type of”; “is located within”
 - Equivalent: “acronym”; “variant”; “common misspelling”
 - Associative: “plays for”; “is a member of”; “
- > People
 - Athlete → Team, Sport, Hometown, Position, Number
 - Politician → Party, Office, Status
 - Celebrity → Significant other, Family members
- > Companies
 - Ticker symbol, Related Industry, Exchange
- > Geographic place names
 - Latitude, Longitude

Subject Vocabulary Example

- [-] H T Social affairs
 - [+] H T Cultures
 - [+] H T Demographics
 - [+] H T Education
 - [+] H T Family issues
 - [+] H T Philanthropy
 - [-] H T Public opinion
 - [+] H T Religion
 - [+] H T Social groups and organizations
 - [-] H T Social issues
 - [+] H T Animal welfare
 - [+] H T Class conflicts
 - [+] H T Consumerism
 - [-] H T Death penalty controversy
 - [+] H T Education issues
 - [+] H T Gender issues
 - [-] H T Hazing rituals
 - [+] H T Human rights and civil liberties
 - [+] H T Human welfare
 - [-] H T Immigration
 - [-] H T Asylum seekers
 - [-] H T Illegal immigration
 - [-] H T Benefits for illegal immigrants
 - [-] H T Illegal immigrant deportation
 - [-] H T Immigrant smuggling
 - [-] H T Immigration laws and regulations
 - [-] H T Naturalization
 - [-] H T Temporary workers
 - [+] H T Labor issues
 - [+] H T Religious issues
 - [+] H T Sex in society
 - [-] H T Smoking bans
 - [-] H T Stem cell research controversy
 - [+] H T Violence
 - [-] H T Volunteerism

Immigration

Permissions: Administrator

✔ **Approved**

Properties

System Properties

Ext. Properties

Relationships

Localized Values

Name:

Description:

Vocabularies:

Name	Description
Social Affairs	
Government and Politics	

Relationship type:

Entity List Example

- [-] H T PERSON
 - [+] H T ARCHITECT
 - [+] H T ARTIST
 - [+] H T AUTHOR
 - [-] H T ENTERTAINMENT_FIGURE
 - [+] H T CELEBRITY_CHEF
 - [+] H T DANCER
 - [+] H T DIRECTOR
 - [+] H T MISC_ENTERTAINER
 - [+] H T MODEL
 - [+] H T MOVIE_ACTOR
 - [-] H T MUSICIAN
 - [-] H T 50 Cent
 - [+] E T Curtis James Jackson
 - [+] H T A.R. Rahman
 - [+] H T Aaliyah
 - [+] H T Aaron Carter
 - [+] H T Aaron Neville
 - [+] H T Abbey Lincoln
 - [+] H T Ace Frehley
 - [+] H T Adam Clayton
 - [-] H T Adam Duritz
 - [+] E T Adam Fredric Duritz
 - [+] R T (FSO) Winona Ryder
 - [+] H T Adam Levine
 - [+] H T Adele
 - [+] H T Adrienne Bailon
 - [+] H T Ahmet Zappa
 - [+] H T Aimee Mann

50 Cent

Permissions: Administrator

Properties System Properties Ext. Properties Relationships Localized Values

URL

DateBirth (1 occurrence)

DateBirth (1) value: 07/06/1975

07/06/1975

Gender (1 occurrence)

Gender (1)

Male

PlaceOfBirth (1 occurrence)

PlaceOfBirth (1)

South Jamaica, Queens, New York City, New York

Auto-Categorization & Entity Extraction

- > Rules-based classification engine (Teragram)
 - Subjects
 - Geography
 - Events
- > Entity extraction based on mentions in the text
 - People
 - Companies
- > Additional data applied based on relationships
 - Athlete hometowns
 - Company ticker symbols
 - Etc.

Tagging Example: Beijing Olympics

Phelps entered in 9 events at US Olympic trials

OMAHA, Neb. (AP) — Michael Phelps is entered in nine events at next week's U.S. Olympic swimming trials. The six-time Olympic gold medalist plans an ambitious schedule for the eight-day trials, which begin June 29. Phelps is entered in the 100- and 200-meter freestyles, the 200 and 400 individual medleys, the 100 and 200 backstrokes, the 100 and 200 butterflys and the 400 free. Katie Hoff will be nearly as busy. She is entered in seven events in an attempt to make her second Olympic team. The trials' schedule mimics the order of races at the Beijing Games in August. Phelps is the world record holder in four events — the 200 free, 200 IM, 200 fly and 400 IM — all of which he set at last year's world championships in Australia. As in 2004, Phelps' Olympic goal is to break Mark Spitz's 32-year-old record of seven gold medals in a single games. Four years ago, Phelps qualified for the Olympics in six individual events, but dropped the 200 back and swam all three relays. In Athens, he tied the record for medals at one Olympics with six golds and two bronzes. This year's trials will be held in a temporary pool indoors at the Qwest Center.

Using auto-categorization rules, plus relationships in the vocabulary, we can apply the following metadata:

Subject:

Men's swimming, Women's swimming, Swimming, Aquatics, Sports, Men's sports, Women's sports

Beijing 2008 Olympic Games, Summer Olympic games, Olympic games, Olympic trials, Events

Entity:

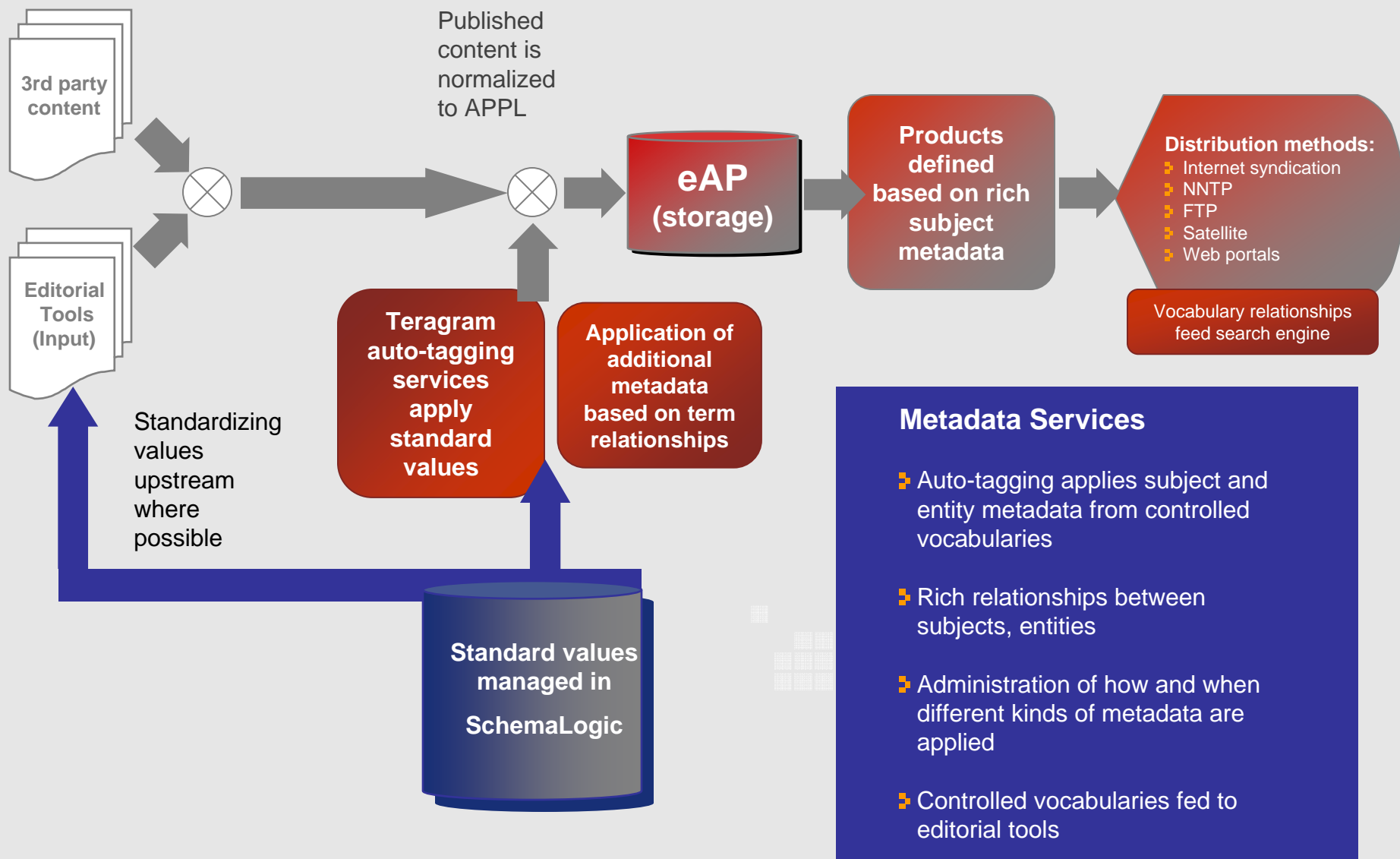
Michael Phelps:

Party type: Olympic athlete, Sports figure,
HometownState: Maryland
Team: United States Olympic Team

Katie Hoff:

Party type: Olympic athlete, Sports figure,
HometownState: Maryland
Team: United States Olympic Team

2009: AP Content Infrastructure



New Capabilities

- > All content combined in a single platform
- > Sharing content
- > Targeted products
- > Better search
- > Linking content

AP Exchange - Windows Internet Explorer

http://www.apexchange.com/pages/portal.aspx

File Edit View Favorites Tools Help

Links

AP AP Exchange

Associated Press Logged in as: jsummerlin | Logout | Contact Us

Home Manage Regular Features Account Settings Help

Search: [Advanced Search](#)

subjectname=libraries

Search within results

AP 360

Recent Searches

Search History

1. [subjectname=libraries](#)

Clear

Bulk Downloads (0)

View

Download

Clear All

Scan News Browse Topics

Show All Narrow Results XML

Business (8)

Entertainment (9)

Environment and nature (1)

General news (3)

Government and politics (7)

Lifestyle (6)

Obituaries (1)

Social affairs (43)

Education (43)

Philanthropy (1)

Religion (1)

Social groups and organization

Social issues (3)

Volunteerism (2)

Technology (2)

All Media Text Photos Graphics Video Interactive Audio Screen: Save This Search

subjectname=libraries

Save This Search

NEW! For quick access to Swine Flu news, click on the new Swine Flu link in the Essential AP section in the left column.

Text : Showing 1 - 10 of 127 Summary Headline Page 1 of 13 more »

<input type="checkbox"/>	9:21 PM	bc-shatford-obit	5/22/2009	//Education Luminary Walter T. Shatford II, 94, Dies// (Los Angeles)			
<input type="checkbox"/>	2:17 PM	BC-RJ-Be-creative-at-Rushville-Public-Library-this-summer-0522	5/22/2009	Be creative at Rushville Public Library this summer			
<input type="checkbox"/>	2:17 PM	BC-RJ-Rushville-Public-Library-building-expansion-may-be-delayed-0522	5/22/2009	Rushville Public Library building expansion may be delayed			
<input type="checkbox"/>	1:26 PM	BC-RN-STATE-BUDGETS-20090521	5/22/2009	Proposed state budgets threaten library funding			
<input type="checkbox"/>	1:05 PM	BC-TS-Be-creative-and-read-this-summer-0522	5/22/2009	Be creative and read this summer			
<input type="checkbox"/>	8:35 AM	BC-HP-Library-school-advocates-want-more-0522	5/22/2009	Library, school advocates want more			
<input type="checkbox"/>	4:45 AM	BC-TB-On-tap-West-suburbs-0522	5/22/2009	On tap: West suburbs			
<input type="checkbox"/>	4:05 AM	BC-PA-Man-arrested-in-Massac-library-break-in-0522	5/22/2009	Man arrested in Massac library break-in			
<input type="checkbox"/>	7:10 PM	BC-DL-Library-board-gets-update-on-summer-programs-0521					

News Bulletins

11:20 AM 5/23/2009
FRANKFURT (AP) — Wolfsburg wins its first Bundesliga title.

11:20 AM 5/23/2009
FRANKFURT (AP) — Wolfsburg wins its first Bundesliga title.

See All

Watch List

View | New

Internet 100%

Standard Metadata in Editorial Tools

Find exact match

Find Previous Next found 0 items

Person	ENTERTAINMENT_FIGURE	MOVIE_ACTOR
Person	ENTERTAINMENT_FIGURE	Lindsay Crouse
Scene	JOURNALIST	Lindsay Lohan
Subject	NEWSMAKER	Lindsay Wagner
	RELATIVE	Lin-Manuel Miranda
	ROYALTY	Lionel Barrymore
		Lisa Bonet
		Lisa Hartman
		Lisa Kudrow
		Lisa Marie
		Lisa Rinna
		Liv Tyler
		Liv Ullmann
		Liza Minnelli
		Lizabeth Scott
		Lloyd Bridges
		Lloyd Hughes
		Lois Chiles
		Lon Chaney

Add

Sibling : Mia Tyler, Taj Tallarico, Chelsea Tallarico
 Parent : Bebe Buell, Steven Tyler
 Child : Milo Langdon
 FormerSignificantOther : Royston Langdon
 description : Liv Tyler is best known for the "Lord of the Rings" film series. Her other notable films include "Stealing Beauty" and "Armageddon."
 Nationality : American

Close

The Future

- > Expanding the categorization services
- > Incorporate descriptive metadata into editorial process
- > Search enhancements
- > Improve distribution systems
- > Sharing vocabularies
- > New revenue opportunities

www.ap.org

Joel Summerlin

jsummerlin@ap.org



Associated Press

the **essential** global news network