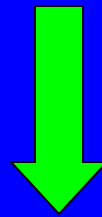


Taxonomy and Semantics



European Media Wrapper Round Table II

Madi Solomon

Madi Weland Solomon: the taxonomy

Professional Experience

Independent Consultant

Walt Disney Company,
Corporate Nomenclature
Taxonomist

Getty Research Institute
Project Associate,
Standards &
Vocabularies

Broad Art Foundation,
Curatorial Assistant

Professional Associations

Chair, AMIA Metadata
Joint Subcommittee

LOC Moving Image
Collections (MIC):
Cataloging and
Metadata Working
Group & Speakers
Bureau

SMPTE W25 Metadata
and Wrappers working
group

Contributor to ISAN
*Best Practices for Audio
Visual Works*

Personal info

Family

husband
daughter
dog

Interests

Visual Art
contemporary art
Playing Musical instruments
Piano
Guitar
Banjo
Drums
Literature
Reading
Writing

definitions

A **Taxonomy** is a list of terms that enable classification of information

- Typically expresses hierarchical relationships (parent/child)
- Emphasizes context

A **Thesaurus** is a specialized taxonomy

- Emphasizing relationships (synonyms, related terms)
- Preferred terms, variant terms, AKAs

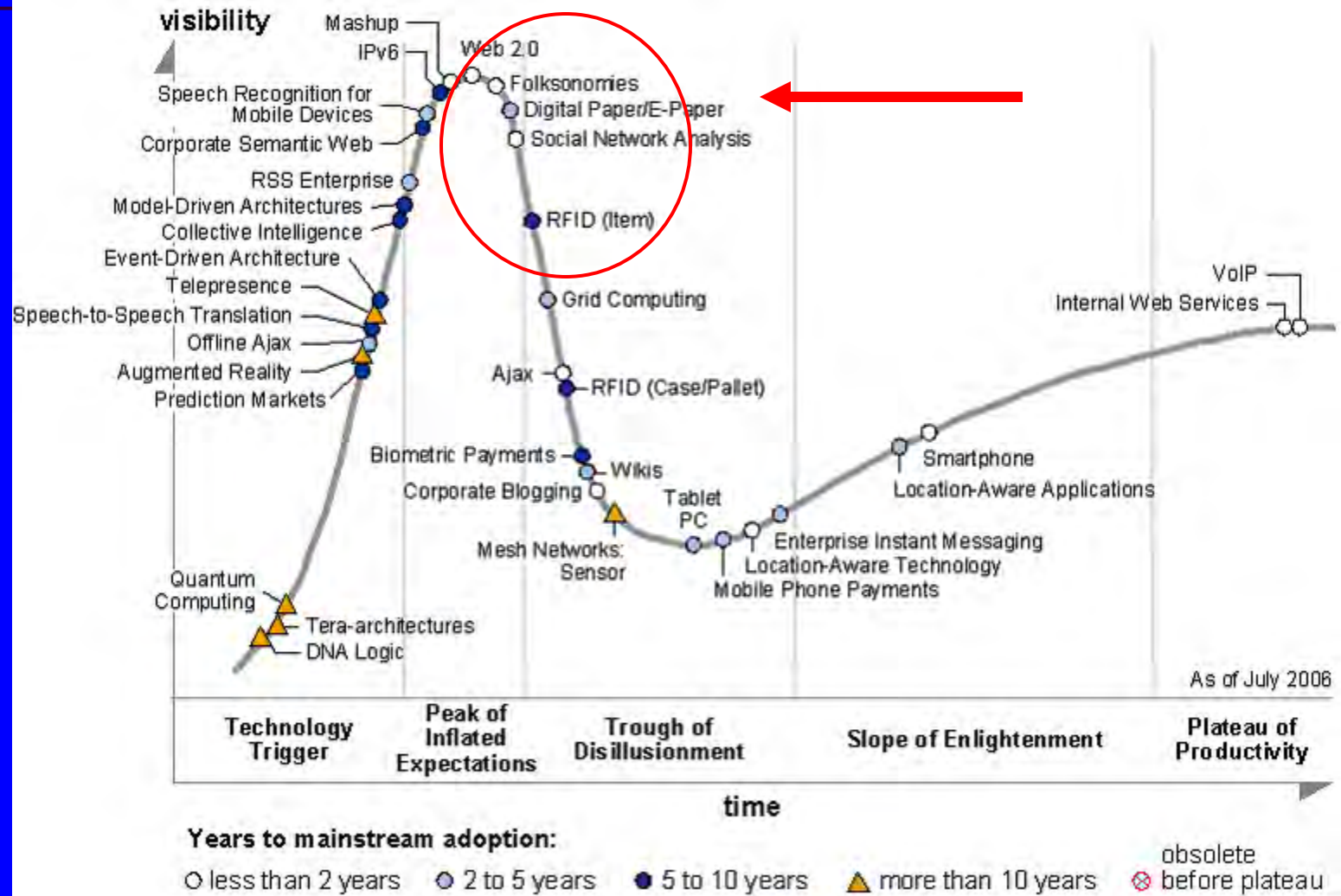
A **Faceted Classification** is a lateral taxonomy which allow for multiple classifications to an object

An **Ontology** is a collection of taxonomies and thesauri

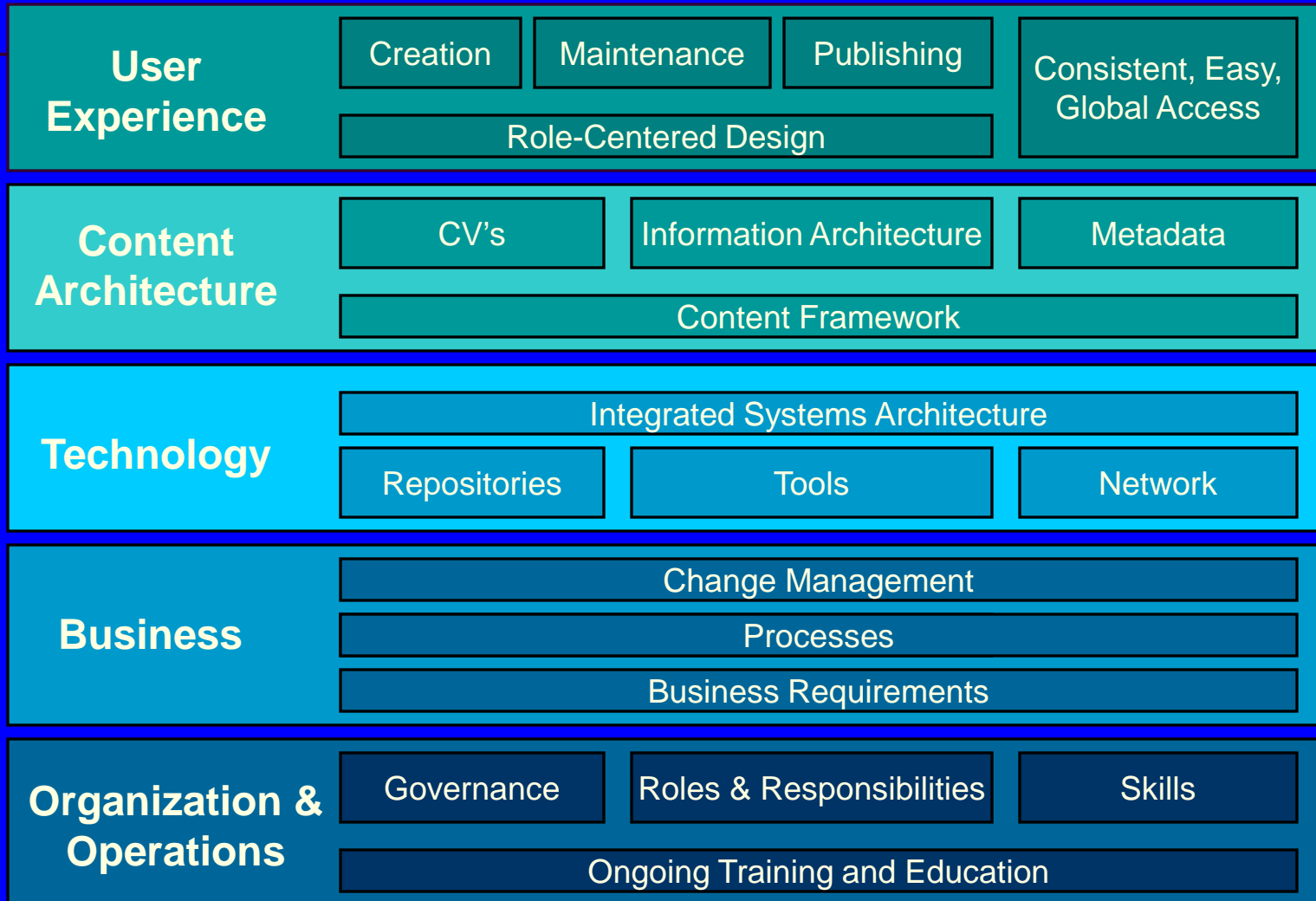
- A body of knowledge is represented by multiple lists of categories
- Categories of various types are conceptually related

Taxonomy and the Hype Cycle

source: Gartner 2006 Hype cycle



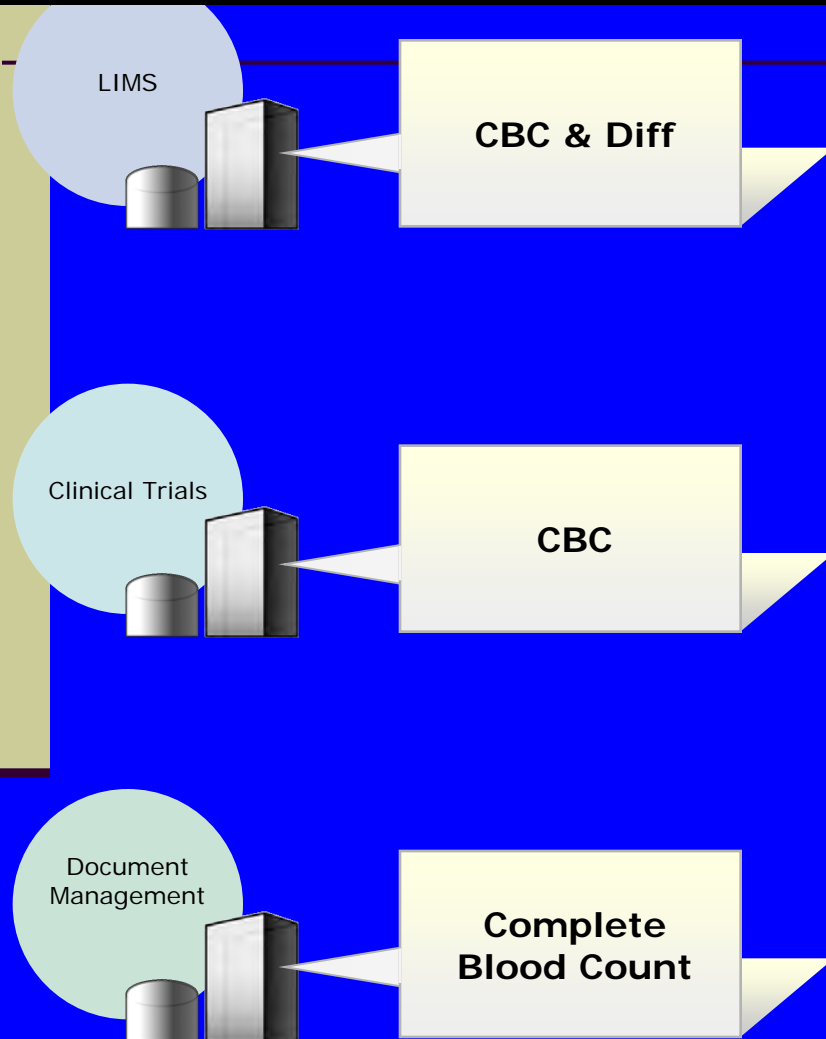
Taxonomy and Content Strategy



Keep in mind: Change is constant

The problem:

Same Term, Different Expressions...



Problems:

- Difficulty finding relevant information
- Federated search configuration is cumbersome
- Inability to view consolidated results
- Limited ability to control shared vocabularies
- Weak governance or demonstrated control
- Costly/cumbersome administrative overhead



MICKEY & FRIENDS
CAST MEMBER

0.00

VALID ONLY

07/07/05

Please mark section below

PINOCCHIO, **CHIP'N DALE**, DAISY
MICKEY, GOOFY, DONALD, MINNIE

SECTION: A B C D E F G H J K

aisle: 1 2 3 4 5 6 7 8 9 10

PLEASE LOCK VEHICLE
256-2365/092216-17:16-07/07/05/045541
See Reverse Side



Chip 'n Dale Treehouse

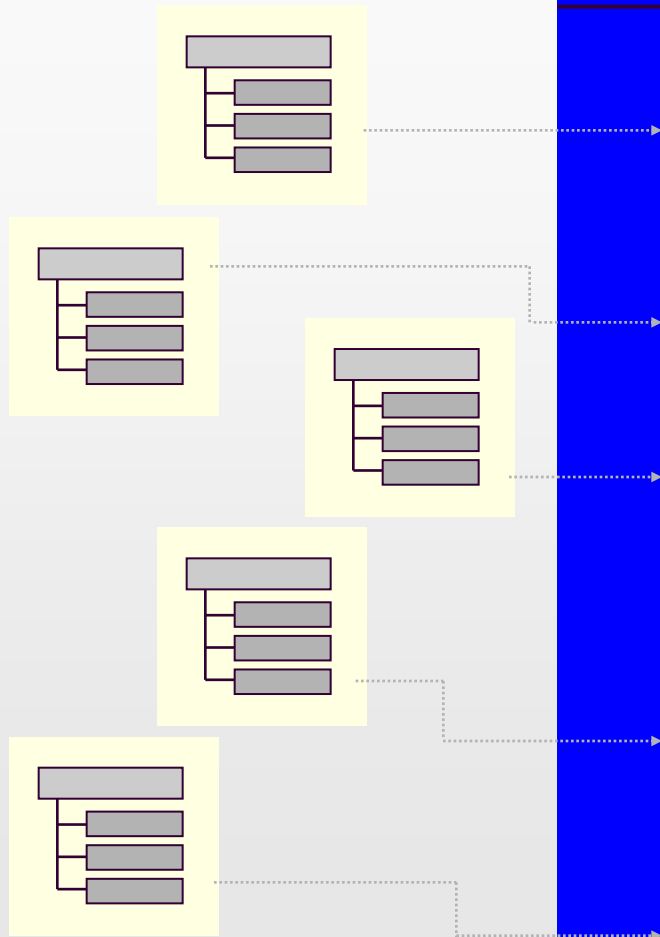


TIPS & FUN FACTS

Cape May Café is physically located at Disney's Beach Club Resort, just a short walk from Disney's Yacht Club Resort.

Goof it up at Goofy's Beach Club Breakfast Buffet hosted by Goofy, who is joined by **Chip 'n' Dale** and Minnie.

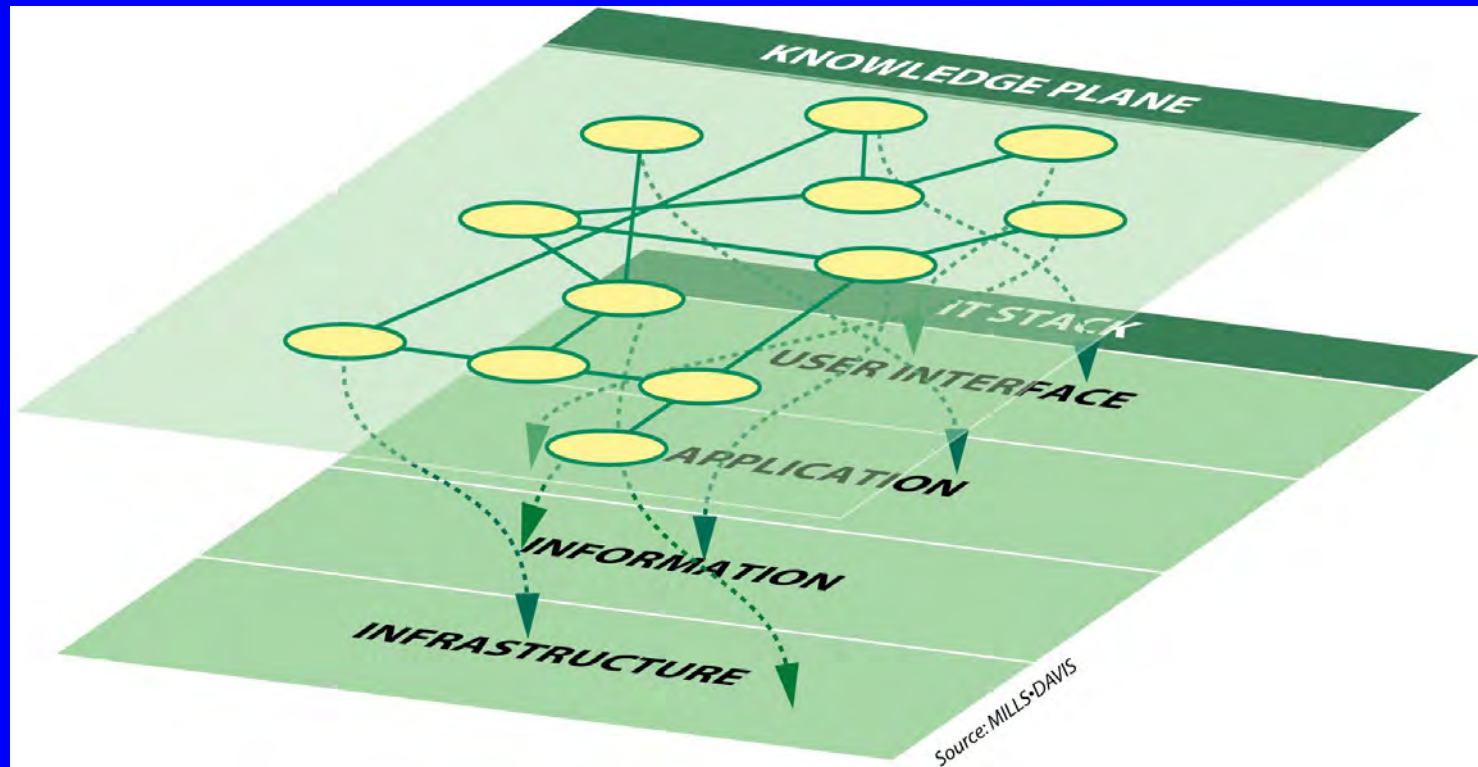
Typical Situation – Disconnected systems, inconsistent metadata



- Inconsistency, no standards
- Many integrations are point-to-point; brittle
- Lack of awareness of possible shared metadata

The “**knowledge plane**” connects resources in and across each layer of the stack. It can be queried, interpreted and reasoned over both by people and machines.

From Mills Davis: Semantic Wave 2006



Three elements of building a taxonomy

1. **Metadata model or schema** A set of data fields for describing content to enhance information access.
2. **Controlled Vocabularies**. An established list of terms from which an indexer or cataloguer may select when assigning descriptors to content.
3. **Application Profile**. Formal representation of metadata & vocabularies.



WHY am I listening to this?

The semantics: adding value

(not the same as “information management”)



Allegory of Fortune, Dosso Dossi

courtesy of the Getty Museum

- **Unique to rich media**
 - interpreting
 - translating

prosperity in life is transitory and dependent on luck

Metadata

more than just
data about the data:

a perspective

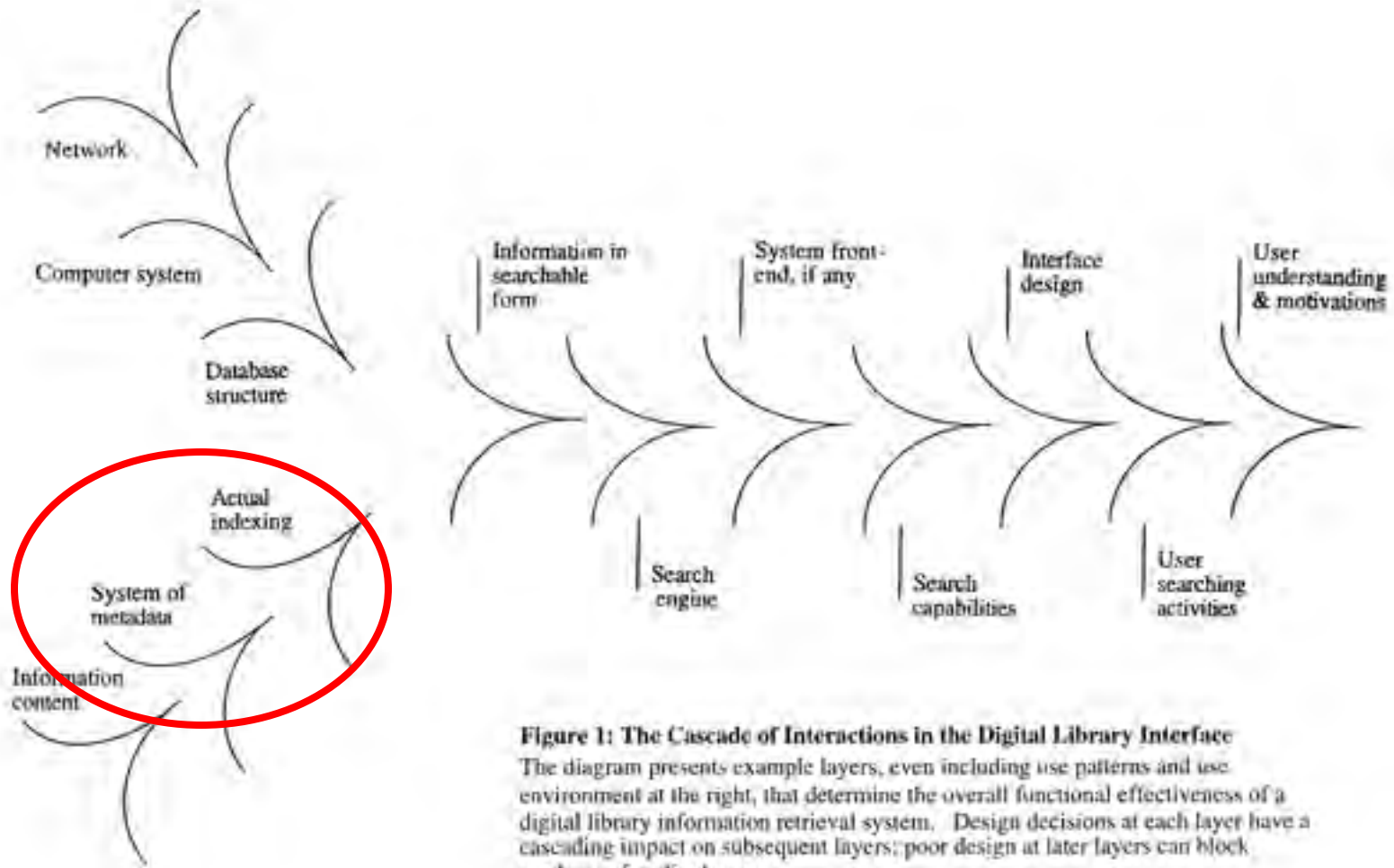
Types of metadata:

who, what, how, when

- **Administrative:** for managing and administering information resources. e.g. location information, version control, tracking. **Who** is responsible for this information/knowledge? **Where** is the repository?
- **Descriptive:** description or identification of information resources. e.g. specialized indexes, finding aids. **What** is the file? **What** kind of content is this?
- **Technical:** how a system functions or how the metadata behaves. e.g. hardware & software documentation, technical specifications of the content. **How** is this file represented?
- **Preservation:** preservation of information management. e.g. document refreshing, migration history. **What** is the lifecycle of this content? **When** did this happen? **What** platform was it created on?
- **Permissions and use:** for articulating usage rights and availability. Includes user tracking and usability studies. **What** can I do with it? **What** have others done with it? **How** can I obtain this?

“Design decisions at each layer have a cascading impact on subsequent layers...”

Source: Marcia Bates, Department of Information Studies, Grad. School of Education and Information Studies, UCLA



A	B	C	D	E	F	G	H	I	J
A	B	C	D	E	F	G	H	I	J
Taxonomy Facet Name	Required Field	Recommended Field	Multi Value Field	Indexed	Controlled Vocabulary	Data Example	Description	Controlled Vocabulary Source	
Audience_PTC		*	*	*	*	Project Managers	Internal audience(s) targeted by the document	Audience vocabulary	
Discipline		*	*	*	*	Document Management	Describes a number of ways that consultants may think of customer challenges	Discipline vocabulary	
Document_Type	*	*	*	*	*	AAR	Content genre	Document Type vocabulary	
Engagement_Value		*		*	*	Small or <\$100,000.	Document describes project of specific size. It is likely that only SO's can be identified this way.	Engagement value vocabulary	
Industry		*		*	*	Automotive	Industry associated with customer discussed in document	Industry vocabulary	
PDS Stage		*	*	*	*	Change management	PDS project phase described by document	Phase vocabulary	
Product		*	*	*	*	ProEngineer	Documents associated with products developed and maintained by PTC for solution integration	Product vocabulary	
Product Release		*		*	*	8.0	Release/Version number of the product described	Release vocabulary	
Service Package		*		*	*	QuickStart	Document associated with fixed price packages sold by PTC that incorporate product (s) and consulting services.	Service vocabulary	
Project Lifecycle		*	*	*	*	Mobilization	Document associated with engagement process step(s)	Process vocabulary	

top level hierarchies – the perspectives

Disney

- Asset Format Types
- Business Units
- Content Classifications
- Subject Keywords
- Main Titles/Properties
- Characters
- Personal Names

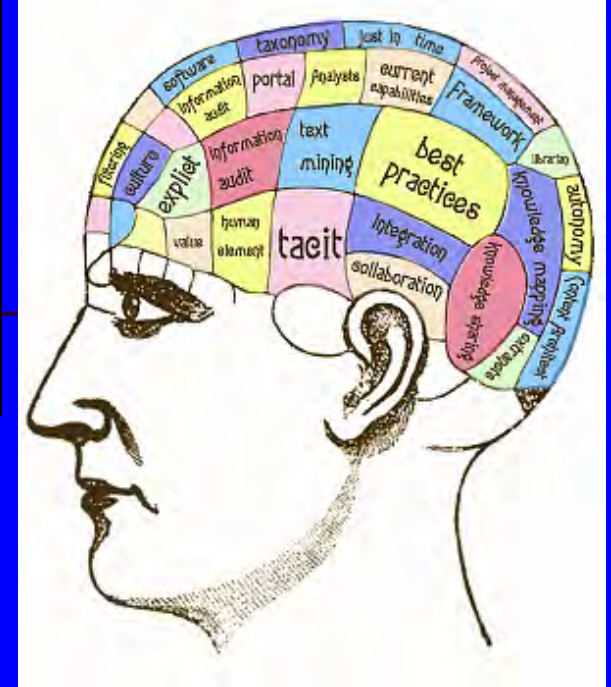
Getty Art & Architecture

- Associated Concepts Facet
- Physical Attributes Facet
- Styles and Periods Facet
- Agents Facet
- Activities Facet
- Materials Facet
- Objects Facet

NASA

- Access Security Requirements
- Audiences
- Business Purpose
- Competencies
- Content Types
- Industries
- Instruments
- Locations
- Missions and Projects
- Organizations
- Subject Categories

documentalist as *system* educator



Search / Browse

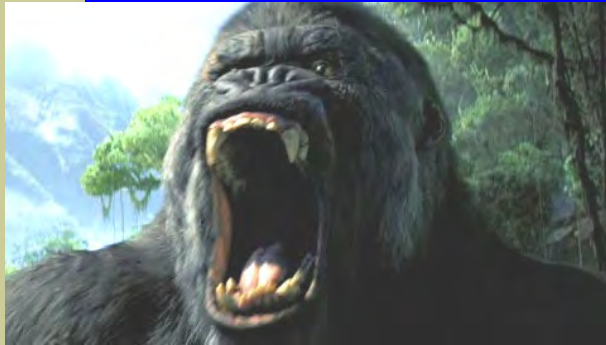
- Intelligent search and classification system
- Constantly 'feed' accurate information
- Provide syntax & entry conventions for precision

== Identification / Authentication

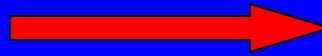
Instantiation / Discovery

Rights Management

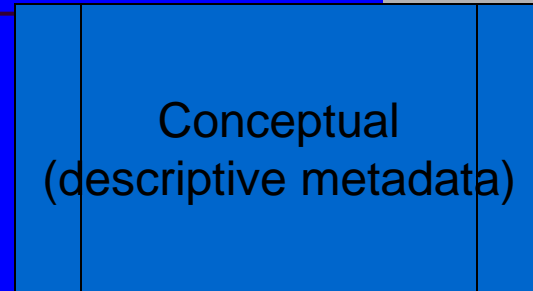
Entity Relationships



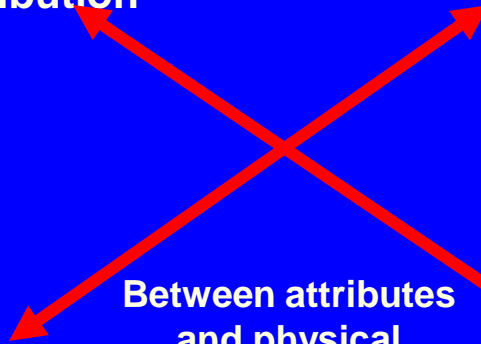
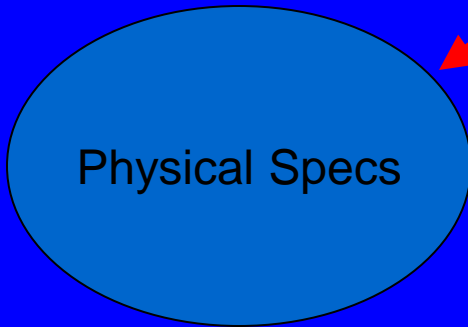
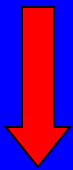
Between classes of things



Between entity associations



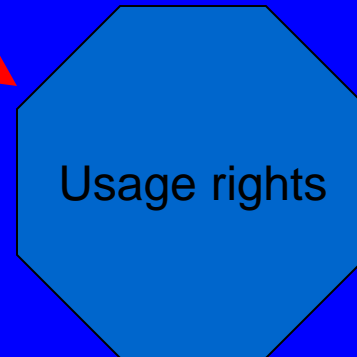
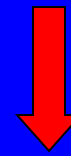
Creation/Production/Distribution

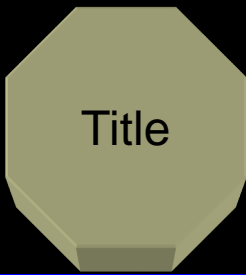


Between attributes and physical mediums



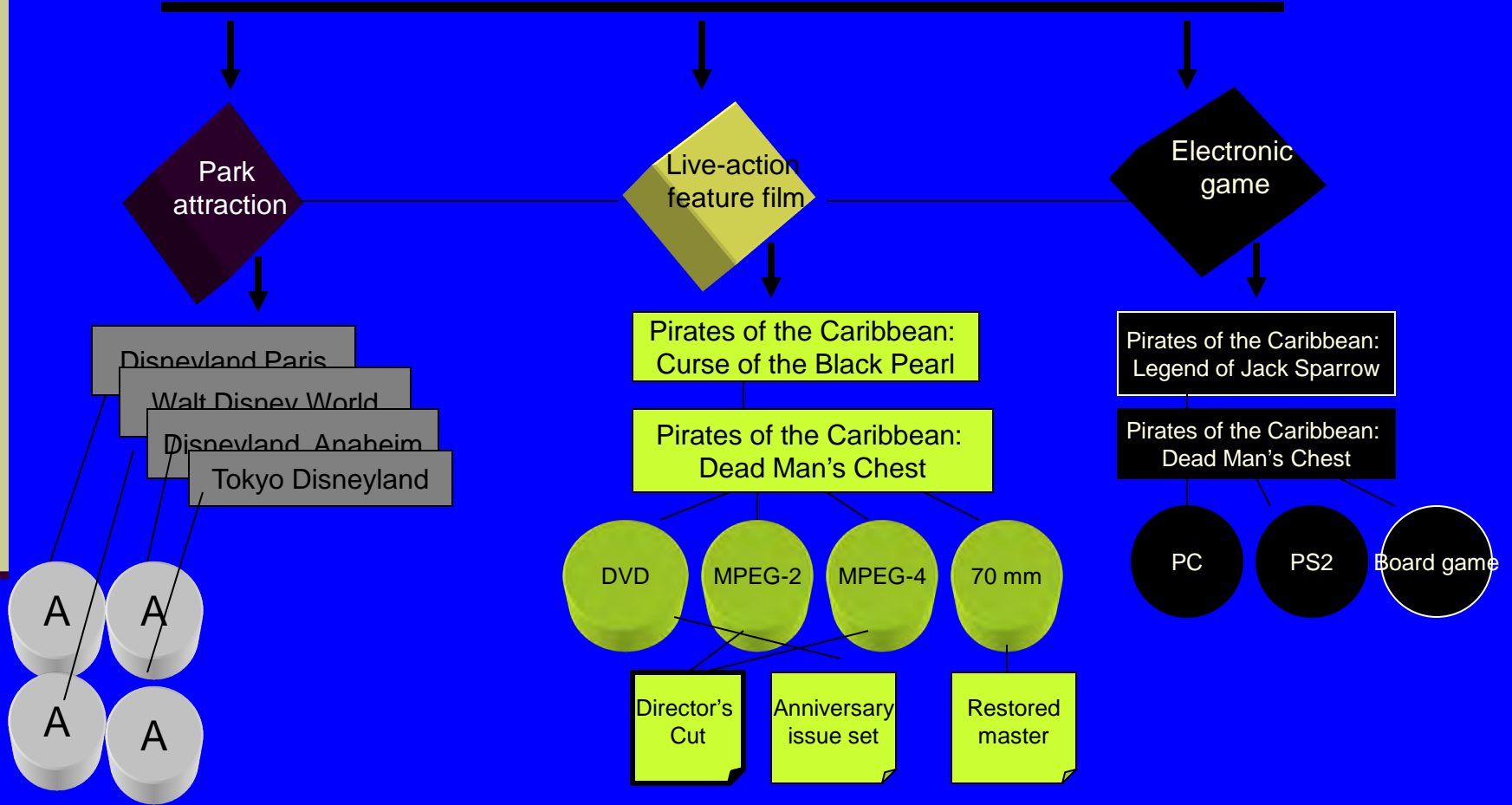
Permissions of Use





Title

Pirates of the Caribbean



Adding intelligence to content

Metadata schema: A defined set of metadata elements

- Identification / Authentication
- Searching / Browsing / Finding /
- Instantiation
- Access and Usage

Taxonomy: A list of terms that enable classification of information

- Process intelligent searches through a classification system
- Relate language to information objects
- Provide inferred meaning through context

Term: knights

The Knights Templar Burned in the Presence of Philip the Fair and His Courtiers,

Boucicaut Master and Workshop, Paris about 1415



<Keywords>



<Agents>

<agents by state or condition>

royalty

kings

→knights

princes

princesses

queens



<Agents Facet>

<People>

<people by state or condition>

landholders

→knights

landowners

tenants



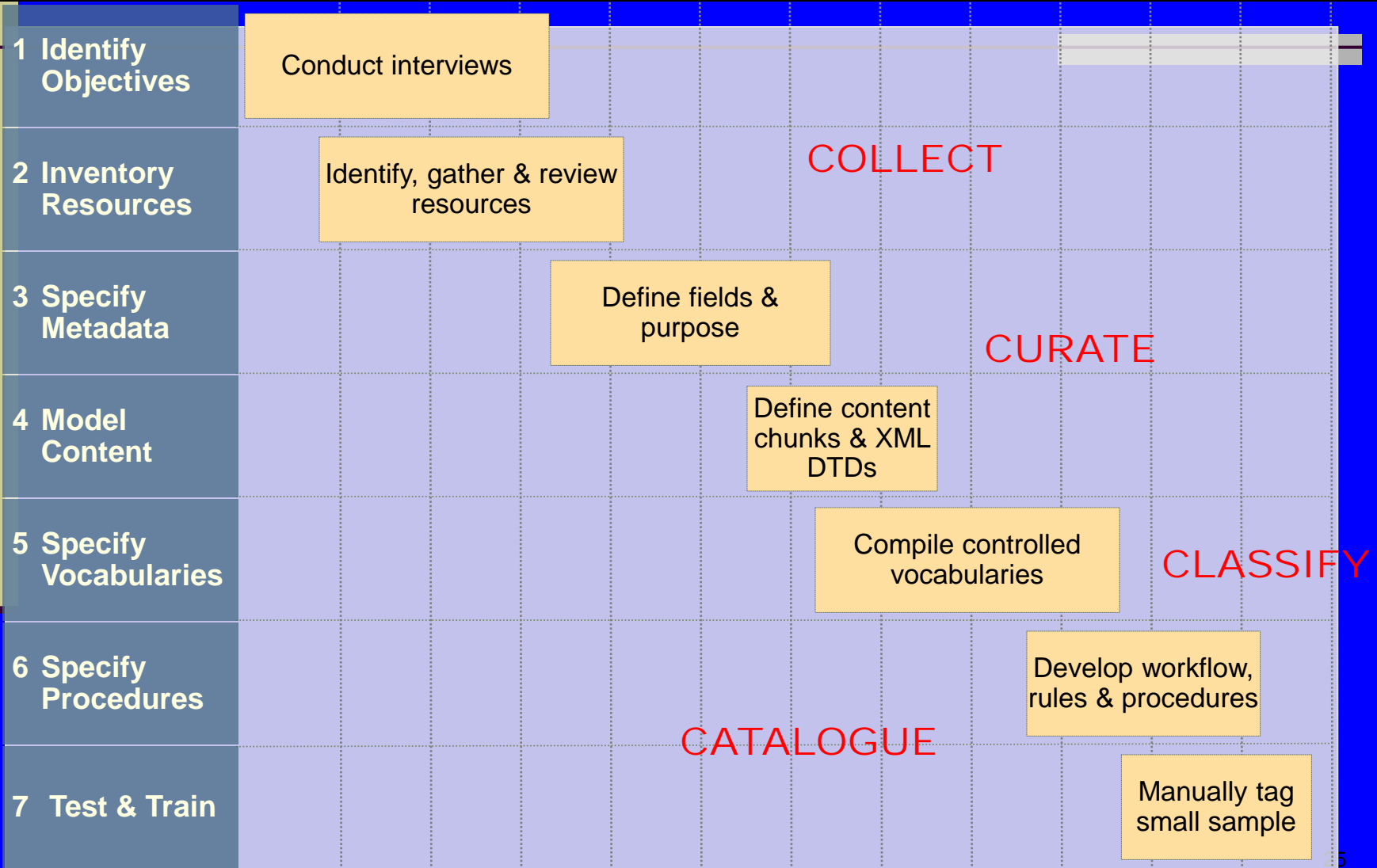
courtesy of Disney Consumer Products

Taxonomy derivation

- **Capture:** What are the current cultural practices? What are the users thought patterns?
- **Curate:** determine what is important in knowledge representation
- **Classify:** Represent knowledge in a structure that supports meaning
- **Catalogue:** validate, test functionality, identify processes

Seven phases of taxonomy development

Source: Taxonomy Strategies taxonomystrategies.com



Application framework

- **What tools are available?**
 - Search mechanisms, content management systems, portal framework, etc

- **How do those tools utilize and apply metadata?**
 - Document and content containers
 - Custom metadata applied to content
 - Thesaurus management
 - Term expansion
 - Mapped search (tuned search)

- **How will metadata be passed from one system to another?**

- **How will updates to taxonomy terms be applied ?**

Taxonomy Maintenance

- **Who owns the taxonomy?** *The consortium of stakeholders.*
- **Who maintains the taxonomy?** *The taxonomy team:*
 - technical specialist*
 - taxonomy specialist*
 - content specialist*
 - domain experts – content custodians – asset managers*
- **How are candidate or proposed terms submitted?** *Through different editorial roles depending on your application.*
- **How are they approved or accepted?** *By user and literary warrant; through the application by the taxonomist.*



News Front Page

World

UK

England

Northern Ireland

Scotland

Wales

Business

Politics

Health

Education

Science/Nature

Technology

Entertainment

Also in the news

Video and Audio

Have Your Say

Magazine

In Pictures

Country Profiles

Special Reports

RELATED BBC SITES

SPORT

WEATHER

Last Updated: Thursday, 19 April 2007, 10:01 GMT 11:01 UK

E-mail this to a friend

Printable version

Colour barcode system to hit DVDs

A colour barcode system holding more data than current codes will find its way onto DVDs later this year.

The four and eight-colour geometric patterns can hold up to two-pages of data, double the amount of traditional black and white, striped barcodes



The new barcode will appear on games later this year

Developed by Microsoft, it is one of a number of competing products hoping to find their way onto packaging.

Microsoft has said consumers could interact with the new barcodes, using webcams and mobile phones with cameras.

Gavin Jancke, the Microsoft Research engineering director who developed the so-called High Capacity Color Barcode (HCCB), said the aim was not to replace the current barcode system, called UPC.





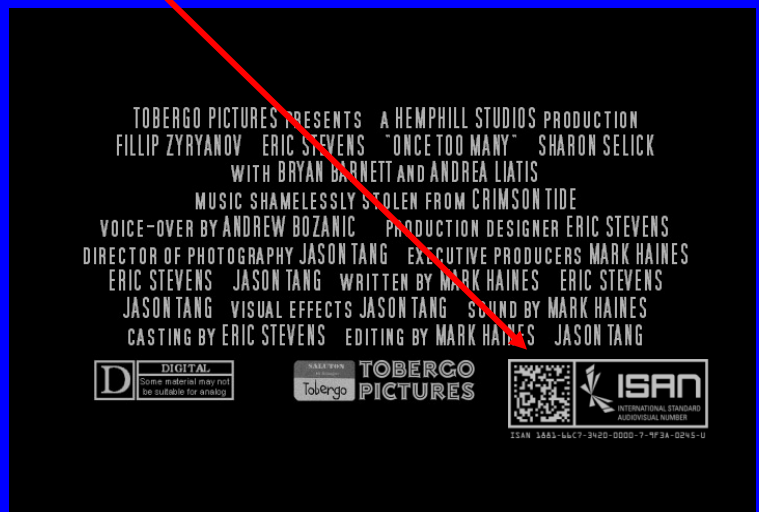
What is ISAN?

The ISAN (International Standard Audiovisual Number) is a voluntary numbering system for the **identification of audiovisual works**, not publications or broadcasts

What ISBN is for books...



... ISAN is for audiovisual works



Approved Registration Agencies

- Agence Française ISAN (France)
- Aribsan (Spain & Latin America)
- ISAN Australasia (Australia and New Zealand)
- ISAN Berne (Switzerland and neighboring countries)
- ISAN NL (Netherlands)
- ISAN RA Deutschland (Germany)
- Microsoft Studios (USA)
- Secure Path (USA)
- Aribsan (Spain and Latin America)
- ISAN NL (Netherlands)
- ISAN IFTA (IFTA Members)
- ISAN UK
- ISAN Sverige
- ISAN PL (Poland)
- ABRISAN (Brazil)
- ISAN Iran

Bob le flambeur

Discuss "Bob le flambeur"



- ≡ **Types** [Film \(Film\)](#)
- ≡ **Also known as** *double-click to add "Also known as"*
- ≡ **Initial release date** 1955
- ≡ **Directed by** [Jean-Pierre Melville](#)
- ≡ **Produced by** [Jean-Pierre Melville](#), [Serge Silberman](#)
- ≡ **Screenplay By** [Auguste Le Breton](#), [Jean-Pierre Melville](#)
- ≡ **Story By** *double-click to add "Story By"*
- ≡ **Cinematography** *double-click to add "Cinematography"*
- ≡ **Edited by** *double-click to add "Edited by"*
- ≡ **Costume Design By** *double-click to add "Costume Design By"*
- ≡ **Music by** [Eddie Barclay](#), [Jo Boyer](#)
- ≡ **Genres** [Film noir](#), [Crime](#)
- ≡ **Runtime In Minutes** 98.0
- ≡ **Languages** [French](#)
- ≡ **Country of origin** [France](#)
- ≡ **Estimated budget** *double-click to add "Estimated budget"*
- ≡ **Rated** *double-click to add "Rated"*
- ≡ **IMDb profile page** <http://www.imdb.com/title/tt0047892>
- ≡ **Soundtrack** *double-click to add "Soundtrack"*

Page History

Created by [Metaweb](#) Oct 23, 2006 11:22am

Last edited by [mw_prop_bot](#) Oct 5, 2007 11:31pm

≡ **Web Link(s)**
double-click to add "Web Link(s)"

≡ **Performances**

[Roger Duchesne](#)

[Isabelle Corey](#)

[Guy Decomble](#)

≡ **Personal Appearances**
double-click to add "Personal Appearances"

≡ **Dubbing performances**
double-click to add "Dubbing performances"

≡ **Featured Film Locations**

In summary

Three elements to building a taxonomy:

1. Metadata schema or model
2. Controlled vocabularies
3. Application Profile

Taxonomy take away:

- Never build in a vacuum
- Be prepared to change with semantics
- Don't need much to do a lot
- Exploit what is abundant

Merci!

Questions?



Madi Solomon
(44) 772 662 8153
madi@taxonomysociety.com