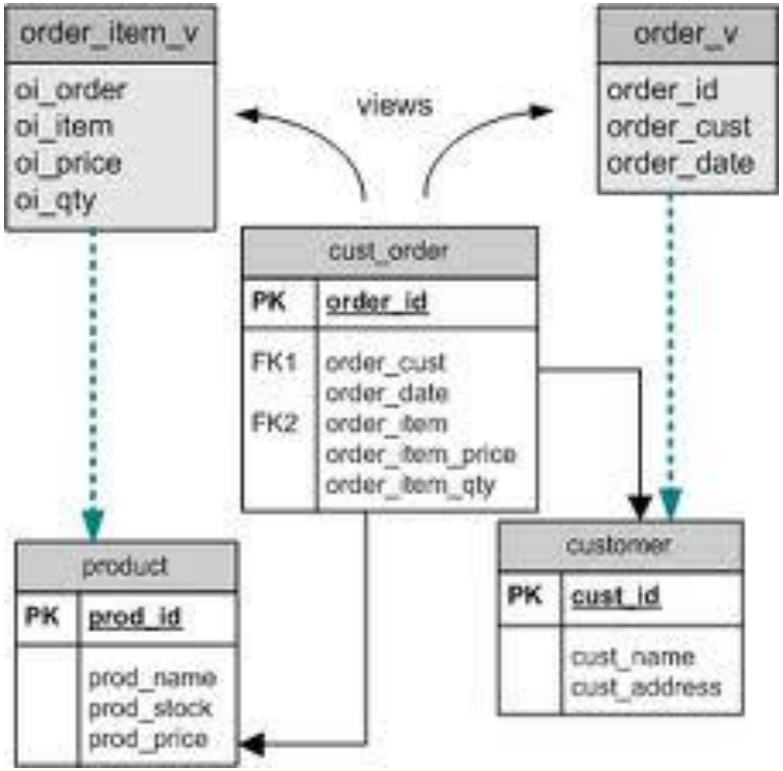
Data Modelling

Modelling

- Activities in information system analysis and design that identifies and develops a model of the information requirements for a proposed computer-based application
  - *Identifying and describing* the information requirements for an information system,
  - *Specifying* the data to be maintained by the data management system, and
  - *Specifying* the *data structures* to be used for data storage that best support the information requirements by providing *efficient* and *effective* information retrieval

User Requirement



Data Model

Data Model Type

- Data structure
  - data types and inter-relationships,
- Constraints
  - allowable values, relationships, and cardinalities, and
- Basic operations
  - for data storage, retrieval, modification, maintenance, and control.

Data structure

- *entity types:* real world phenomena about which data is collected.
  - *f.ex.:* person, teacher, student, course, report, …
- *relationship types:* between entity-types.
  - *associative relationships:* between different types of entities. *f.ex:* student *takes* course, teacher *teaches* course
  - *generic relationships:* which identify sub-classes or *roles* within an entity type. *f.ex* student *IS_A* person, teacher *IS_A* person, …
- *attribute types:* for each entity and associative relationship type.
  - *f.ex.: name* and *address* for the "person" entity type,
  - *semester* and *grade* for the "takes" relationship type.
- *data types:* to be used for recording the values of each attribute type,
  - *f.ex.* name ::= *character string*, picture ::= *image*
  - address::= {Street::=*string*, HouseNr::=*integer*, City::= *string*, PostCode::=*Pcode*}

Constraints

- rules for determining valid values and structures
- *entity identifier*, synonym: *primary key, PK*
  - one or more attributes whose value uniquely identifies an individual entity.
  - *f.ex:* a *social security number (SSN)* or the *person.Id, course.Id, and date* for a grade.
- *reference link*, synonym: *foreign key, FK,* uniquely identifies a related entity.
  - *f.ex:* person.*spouse*::= SSN that references a given person's spouse.
- *relationship cardinality:* for each entity to relationship type.
  - *f.ex.:* Student *(0:m)* takes *(0:m)* courses, meaning that a student may take more than 1 course and a course may be taken by more than 1 student.
- *participation constraints:* for generic hierarchies,
  - specify the relationship between the parent and sub-class entity-types
  - *f.ex.* a *partial, overlapping (p,o)* participation constraint for the person heirarchy indicates that each person (in the DB) may be a student, a teacher, both or neither.
- *domain constraints:* define the valid value sets for attributes,
  - *f.ex:* course.level must have a value between [1..6]

Basic operations

- define how data values are calculated when some event occurs.
  - *F.ex:* age and wage can be calculated upon request using a function definition:

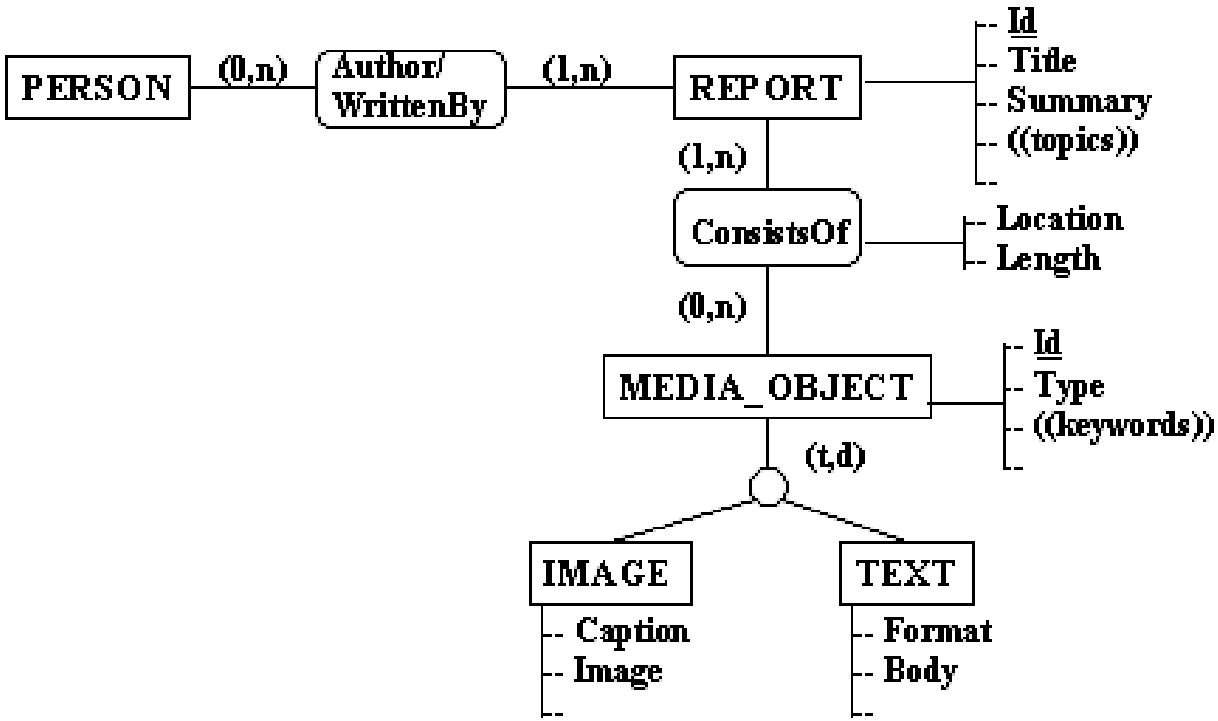
    person.age = todays_date - birth_date

    student.wage = hourly_wage * number_of_hours_worked
- ***Functions*** can also check that data constraints are maintained.
  - *F.ex:* ON INSERT person CHECK UNIQUE person.Id

Modelling Process

- Development of a *semantic model* of the information requirements for an information system.
- Translation of the semantic model to a *structural model* that describes the data types and structures using the concepts of a particular data model type (relational, semantic, object-oriented, ...).

Data Model


PERSON
(0,n)
Author/
WrittenBy
(1,n)
REPORT
Id
Title
Summary
((topics))
(1,n)
ConsistsOf
Location
Length
(0,n)
MEDIA_OBJECT
Id
Type
((keywords))
(t,d)
IMAGE
Caption
Image
TEXT
Format
Body

Meta Data

- Data about Data
- Uses
  - Specification and interpretation of user requests - queries - for information
  - Determining storage, indexing, and retrieval of the data to/from the database

Meta Data Spesification

- Database design
  - a semantic data model
  - metadata standard for description of user requirements, data structures, and constraints
- Data storage –
  - using a data definition language, DDL, to add metadata values.

Meta Data for Multimedia

- ***Semantic*** metadata
  - characterize the subject matter of the document,
- ***Context*** metadata
  - describe relationships to external (to the meaning of the document) objects, such as author and publisher
- ***Structural*** metadata
  - describe the internal structure and presentation layout for the media object.

Semantic Metadata

- Speficy *features* that describe the semantic content of the media object
  - Index terms or keywords
  - Shapes
  - Colors
  - Textures
  - Categories

Context Metadata

- specify the relationships that the media object has to its environment
  - author/creator, publisher, the date of creation/purchase/publication, and the current location of the object
  - typically represent 'facts' about the object that need to be determined and recorded manually
  - can be modelled as objects, attributes and/or relationships in a traditional data model

Structural Metadata

- describe the implementation, layout
  - spatial and temporal placement of objects within a document
  - presentation style of the media object
  - Language, length, presentation media

Multimedia Metadata

- Dublin Core
  - originally developed for description of text-based documents,
- Mpeg-7
  - developed for description of streamed multimedia such as film
- CIDOC/CRM
  - developed for description of museum artifacts

Dublin Core

- Metadata for Electronic Documents

Metadata Type	***DC element***
Semantic	Title, Subject, Description, Type, Coverage
Context	Creator, Contributor, Publisher, Date, Rights, Source, Relation
Structural	Type, Format, Language, Identifier

Dublin Core

1. TITLE A name give to the resource.

2. CREATOR An entiry primarly responsible for making the contents of the resource.

3. SUBJECT The topics of the contents of the result.

Dublin Core

4. DESCRIPTION A textual description of the content of the resource, including abstracts in the case of document-like objects; also may be a content description in the case of visual resources.

5. PUBLISHER The entity responsible for making the resource available in its present form, such as a publisher, university department or corporate entity.

6. CONTRIBUTORS Person(s) or organization(s) in addition to those specified in the CREATOR element, who have made significant intellectual contributions to the resource but on a secondary basis.

7. DATE The date the resource was made available in its present form.

8. TYPE The resource type, such as home page, novel, poem, working paper, technical report, essay or dictionary. It is expected that TYPE will be chosen from an enumerated list of types.

9. FORMAT The data representation of the resource, such as text/html, ASCII, Postscript file, executable application or JPG image. FORMAT will be assigned from enumerated lists such as registered Internet Media Types (MIME types). MIME types are defined according to the RFC2046 standard.

Dublin Core

10. IDENTIFIER A string or number used to uniquely identify the resource. Examples from networked resources include URLs and URNs (when implemented).

11. SOURCE The work, either print or electronic, from which the resource is delivered (if applicable).

12. LANGUAGE The language(s) of the intellectual content of the resource.

13. RELATION The relationship to other resources. Formal specification of RELATION is currently under development.

14. COVERAGE The spatial locations and temporal duration characteristics of the resource. Formal specification of COVERAGE is also now being developed.

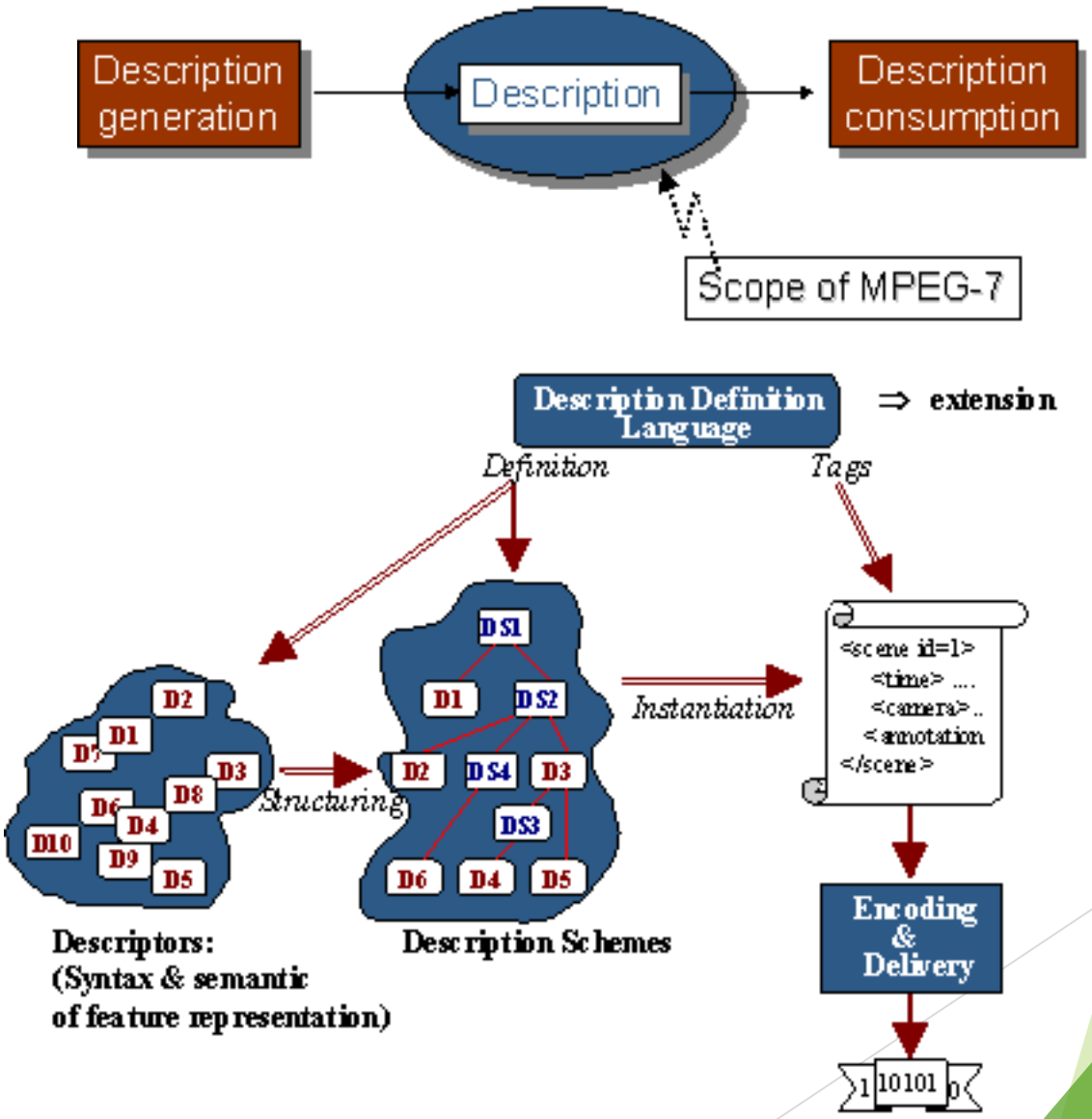
15. RIGHTS MANAGEMENT A link (URL or other suitable URI as appropriate) to a copyright notice, a rights-management statement or perhaps a server that would provide such information in a dynamic way.

MPEG-7

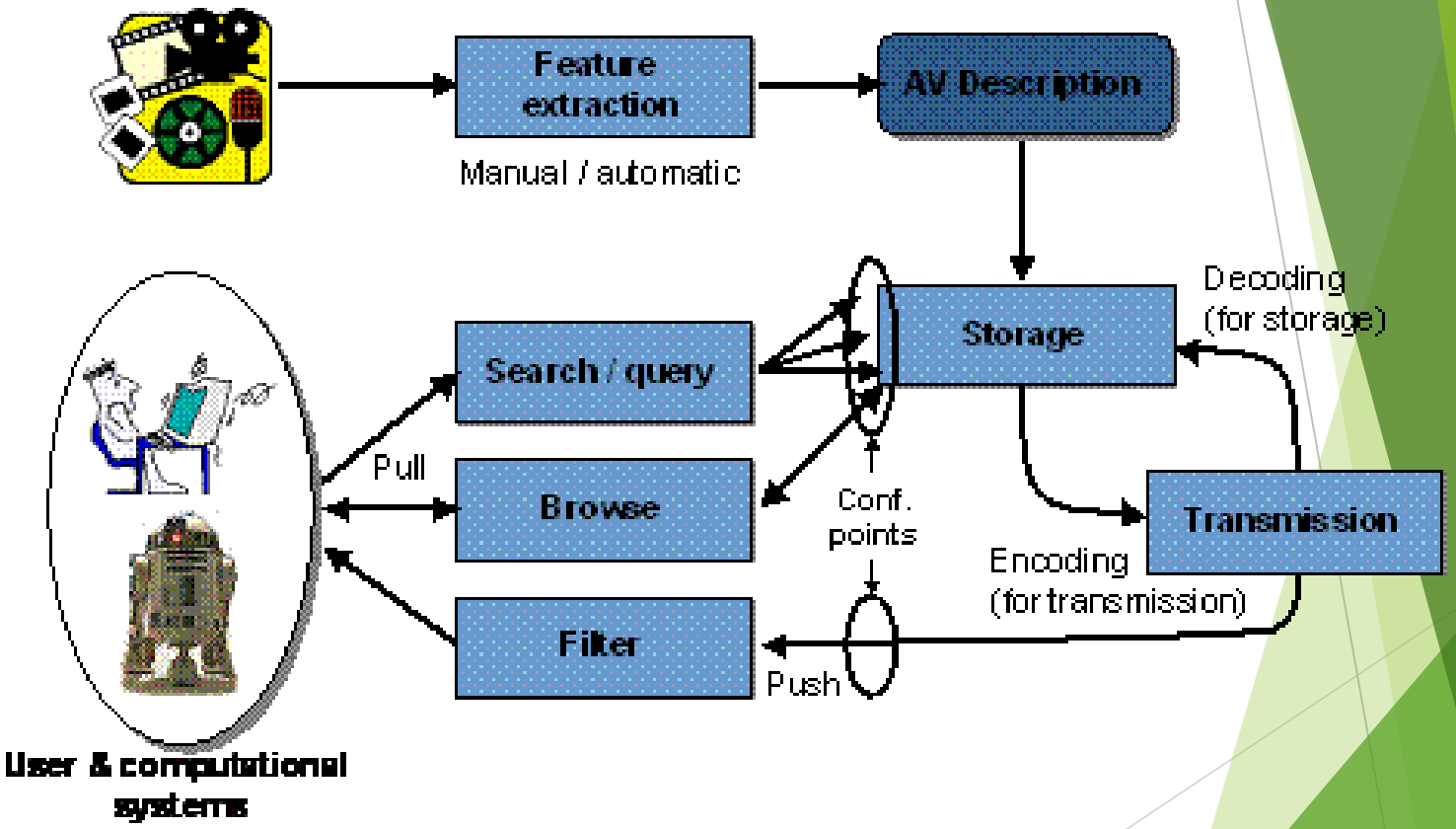
- Dublin Code
  - does not describe basic characteristics of moving pictures (streamed data), such as the temporal and sequencing requirements or the object identification and inter-relationships depicted in the film media.
- The Moving Picture Experts Groups
  - establishes standards for transmitting digital films and videos on the Internet
  - focused on describing the semantic and structural content of multimedia
  - formally named as "Multimedia Content Description Interface"

Elements of MPEG-7 Standards

- Description Tools: Descriptors (D), that define the syntax and the semantics of each feature (metadata element); and Description Schemes (DS), that specify the structure and semantics of the relationships between their components, that may be both Descriptors and Description Schemes;
- A Description Definition Language (DDL) to define the syntax of the MPEG-7 Description Tools and to allow the creation of new Description Schemes and, possibly, Descriptors and to allow the extension and modification of existing Description Schemes;
- System tools, to support binary coded representation for efficient storage and transmission, transmission mechanisms (both for textual and binary formats), multiplexing of descriptions, synchronization of descriptions with content, management and protection of intellectual property in MPEG-7 descriptions


Description generation
Description
Description consumption
Scope of MPEG-7
Description Definition Language
⇒ extension
Definition
Tags
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
Structuring
DS1
DS2
DS3
DS4
Instantiation
<scene id=1>
<time> ....
<camera>..
<annotation
</scene>
Encoding & Delivery
1 10101 0
Descriptors: (Syntax & semantic of feature representation)
Description Schemes

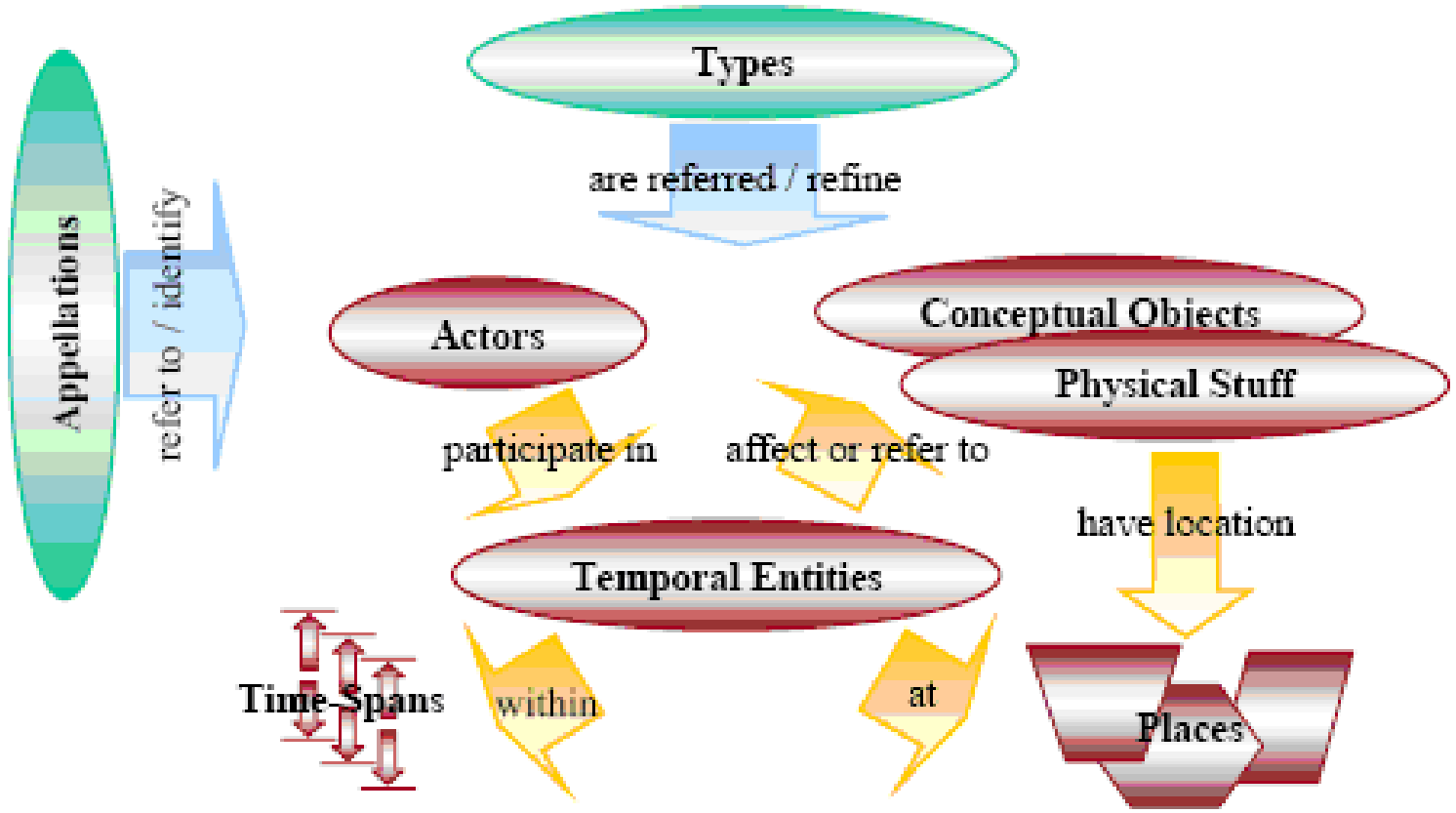
Possible Application



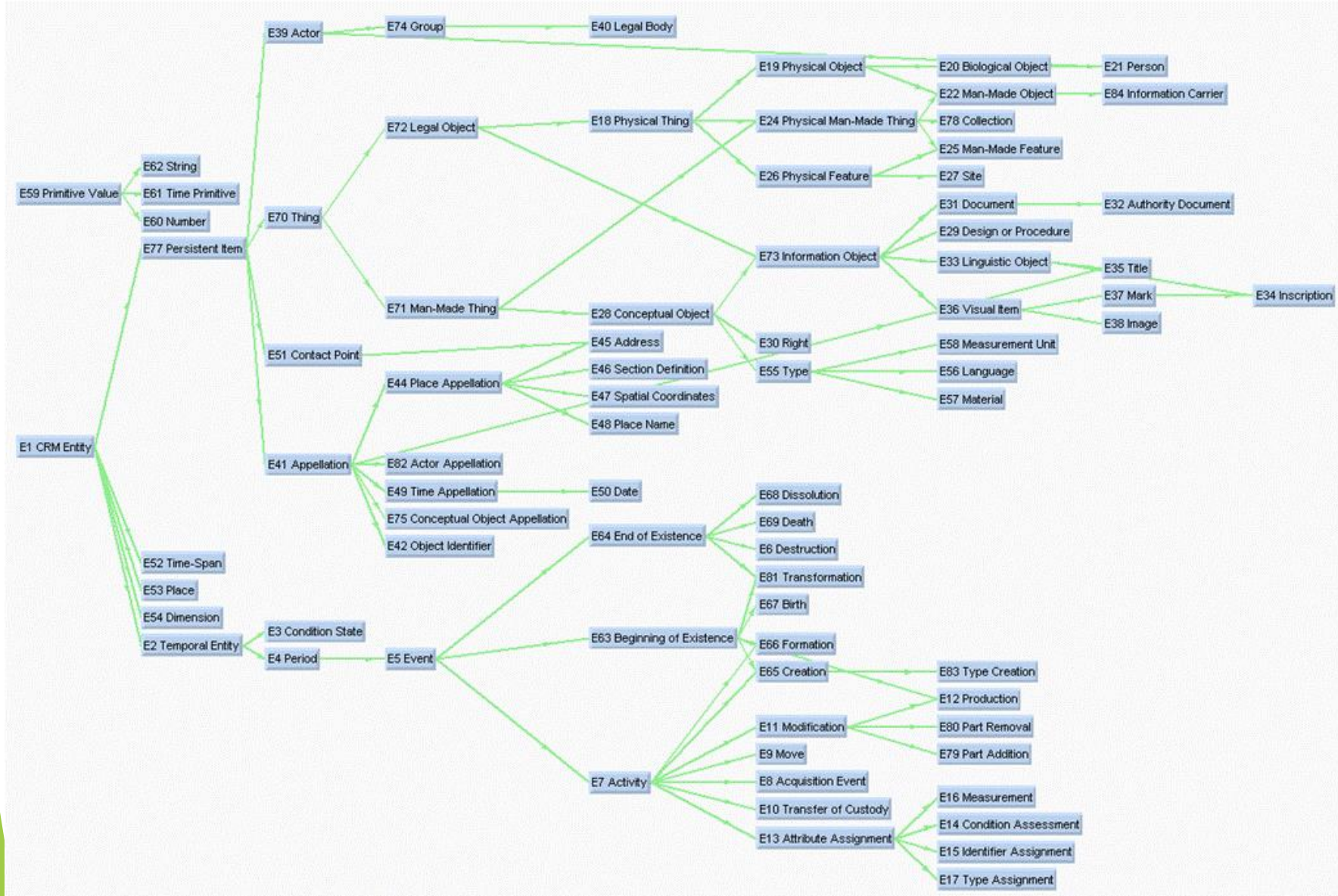
CIDOC-CRM

- Developed by The International Committee for Documentation, CIDOC, of international council of museums, ICOM
- for describing cultural heritage objects, particularly those found in museum collections
- descriptions include the spatial, temporal and event aspects of museum objects, in addition to the semantic, context and structural aspects emphasized in the Dublin Core standard

CIDOC-CRM Component Relationships



CIDOC-CRM Class Hierarchy



Multiple Media Documents

- A report, containing text, images, as well as structured data for title, author, date, ...
- A map, containing points, lines, areas or regions, as well as title, place names, facts (distances, heights), and icons, ...
- A film, containing image stream, and (multiple) audio streams as well as a title, actor list, producer, ...

Text Documents

- can be *described* from 3 perspectives:
  - Semantic content of the document, i.e. representation of its meaning,
  - Context of the document, e.g. its author, publisher
  - Structure of the document, e.g. its language, style, length, .

Images

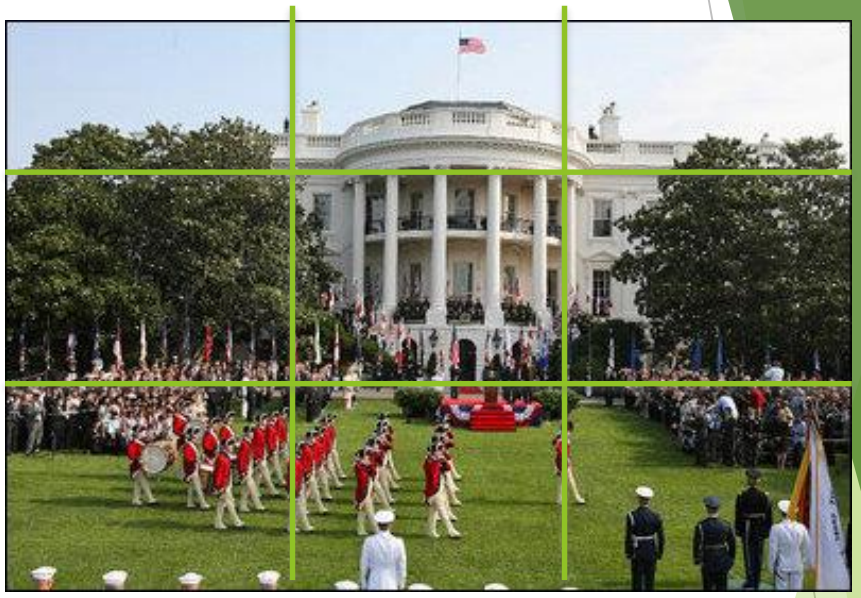
- do not have a standard 'vocabulary' or grammar that can be used for automatic interpretation of the semantic content, or meaning of the images
- Through text annotations
- Using content descriptors or features of images

Content Descriptors

Level	Feature	examples
1	structural	color, texture, shape location in image
2	Objects in image	building, trees, people, band spatial relationships between objects
3	Identification of objects in image	The White House, US army fife and drum corps
4	Event/action representation	marching, welcome ceremony
5	Emotion represented	formal, official, at 'attention'

Image color-signature example

low level features, such as color, texture and shape can be used to compose an *image signature*

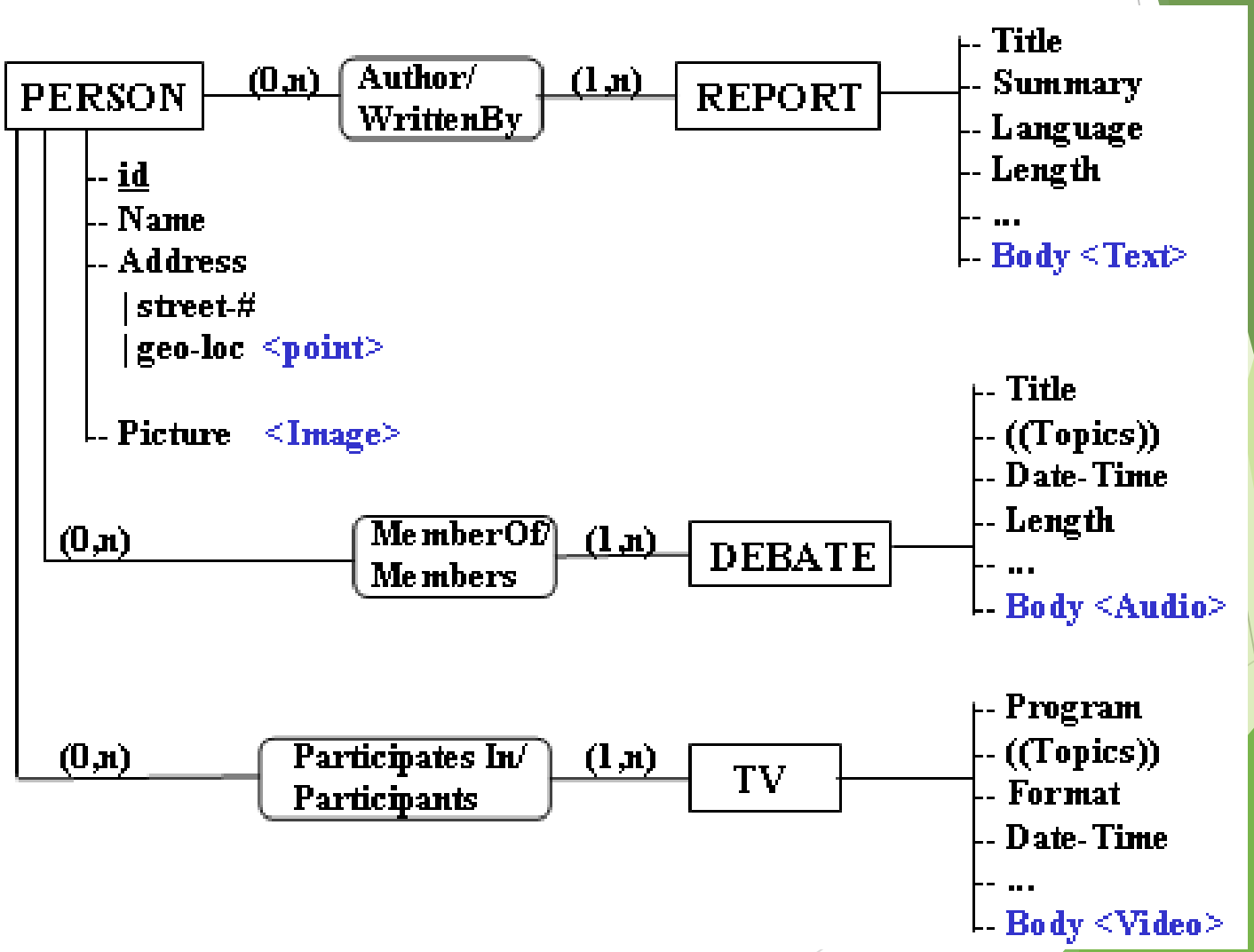


Color	*Cell-1*	*Cell-2*	*Cell-3*	*Cell-4*	*Cell-5*	*Cell-6*	*Cell-7*	*Cell-8*	*Cell-9*
red	**-**	**1**	**-**	**1**	**5**	**1**	**20**	**20**	**-**
green	**40**	**2**	**30**	**89**	**20**	**85**	**70**	**75**	**70**
blue	**60**	**48**	**20**	**-**	**-**	**1**	**-**	**-**	**25**
white	**-**	**49**	**50**	**10**	**75**	**13**	**5**	**5**	**5**

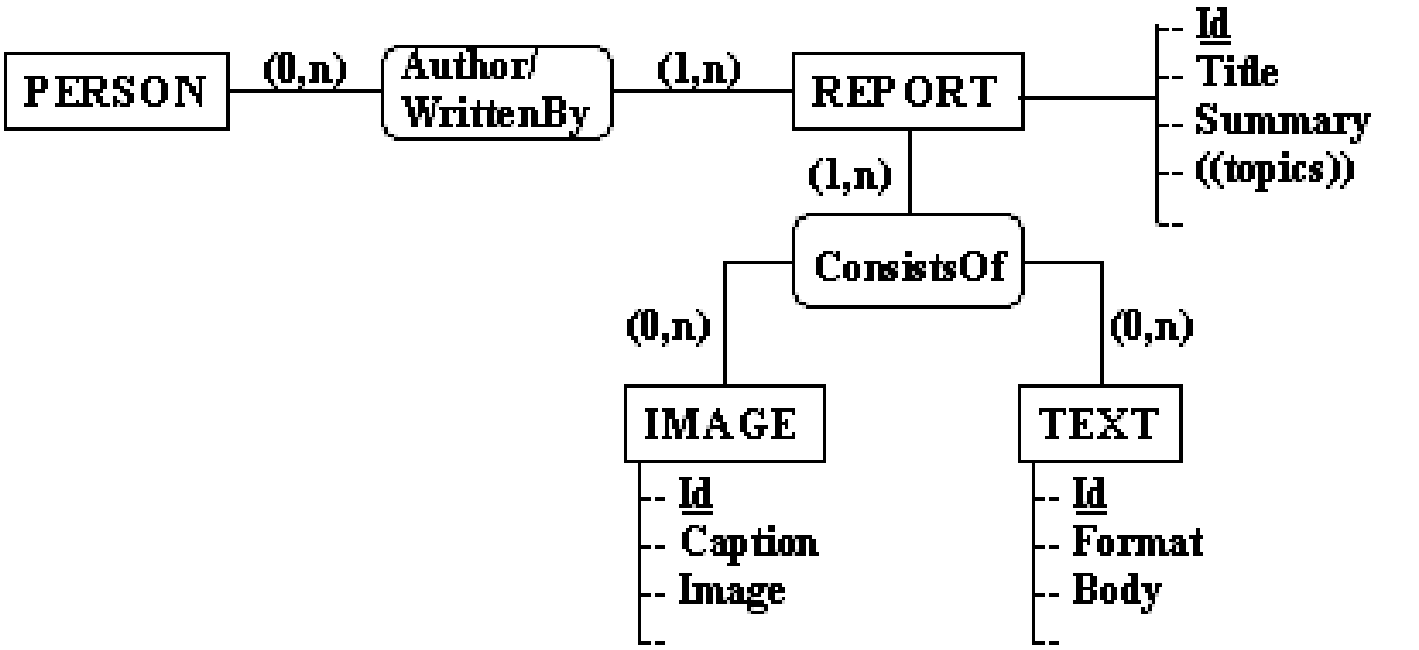
Modelling Multimedia Data in SSM

- media objects can be modelled as:
  - An attribute of an entity-type: ex. Person.picture or Person.geographic_location
  - An entity-type: ex. Report with such attributes as {id, title, keywords, summary, content}
  - A set of related entity-types: ex. Report consists of Media Objects of text and image type.

Media Objects as Attributes

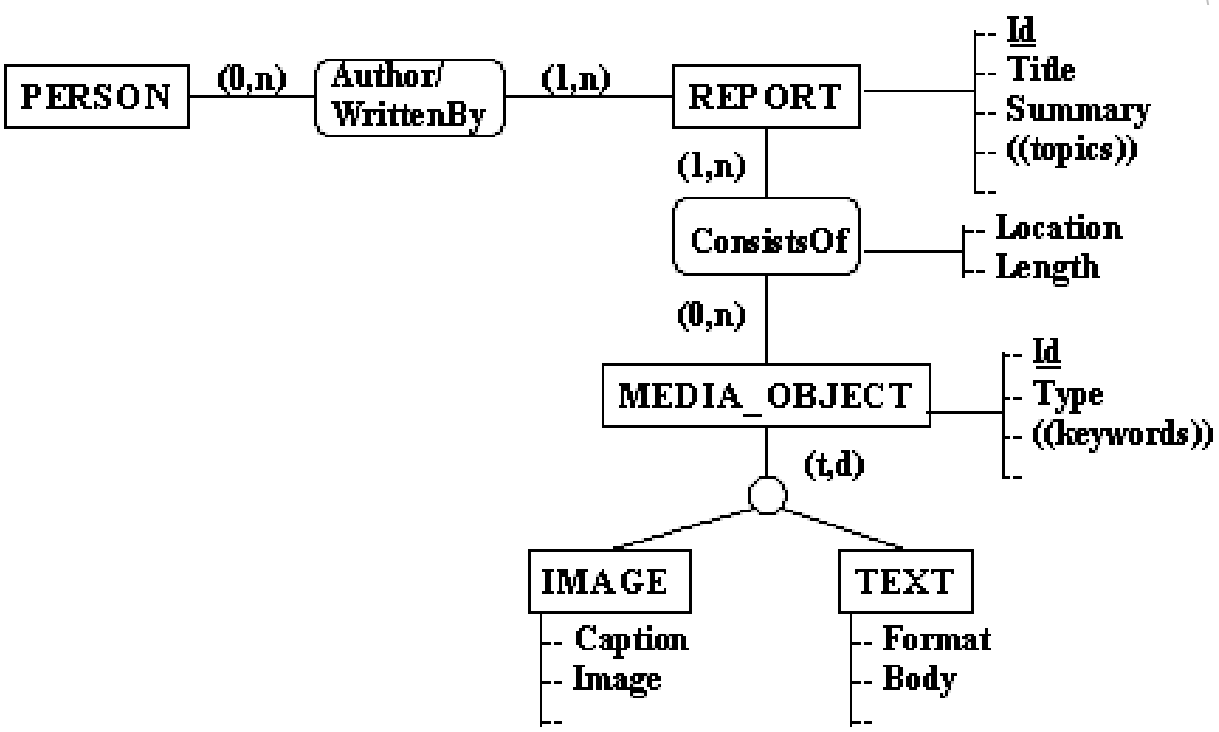


Media Objects as Ternary Relationship



- useful when a balanced set of related media objects occur in the 'parent' multimedia object

Media Objects as a Classification Hierarchy



- useful when the sets of media objects are relatively independent and used in multiple multimedia objects