EFILA Meeting
Dublin
June 20, 1996

Z39.50 Digital Collections Profile
and
Z39.50 Attributes

Ray Denenberg
Library of Congress
ray@rden.loc.gov

Z39.50 Maintenance Agency

(Library of Congress)

http://lcweb.loc.gov/z3950/agency

- **The ANSI/NISO Z39.50-1992 document**
- **The ANSI/NISO Z39.50-1995 document**
- **Register of Implementors**
- **Profiles**

etc.

Approved Profiles:

- GILS,
- WAIS,
- ATS, and
- Z39.50 over TCP.

And under development:

- Use of Z39.50 Item Order Extended Service for Interlibrary Loan

- Z39.50 Profile for Access to Digital Collections

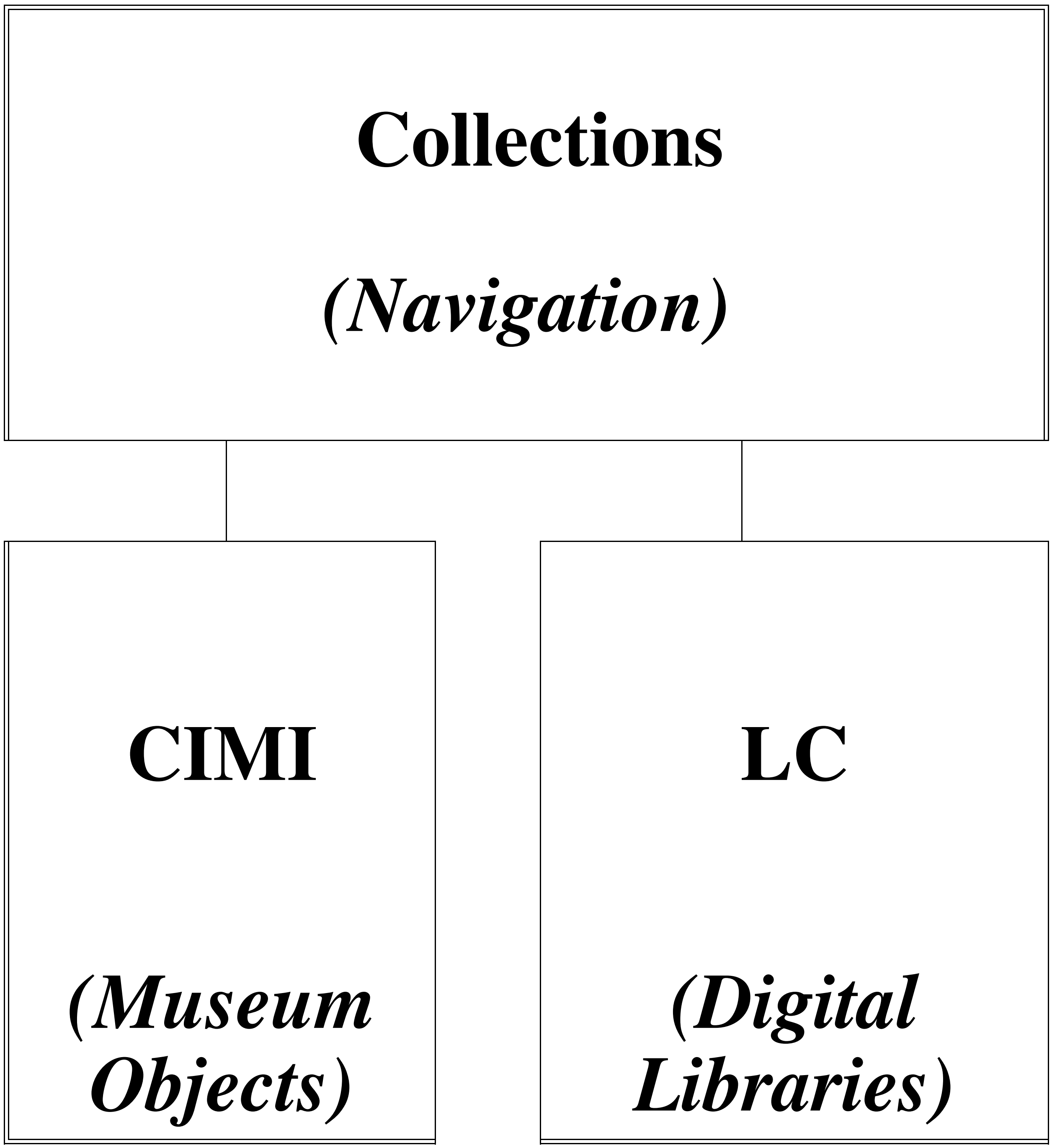
- Z39.50 Profile for Digital Libraries (LC)

- Z39.50 Profile for Museum Collections Information (CIMI)

- Geo profile

Participants in Development of Collections Profile

Getty
Berkeley
U. Michigan
U. California
OCLC
LC
RLG
Chemical Abstracts
IBM
FCLA
~~TRW~~
~~Knight Ridder~~
SilverPlatter
(Consultants)
(Liaisons)


Collections
(Navigation)
CIMI
(Museum Objects)
LC
(Digital Libraries)

Companion Profile

Compatible *extension* to or *subset* of Digital Collections profile, for a specific *application* or *class of information*.

- **CIMI**
- **National Digital Library (LC)**
- **Federal Theater Project (LC)**
- **Geospatial Data**
- **Earth Observation Data**
- **Thesauri**

Properties of Collections

Collections:

- are organized thematically (subject, creator, historical period);
- with numerous, diverse objects;
- both digital as well as physical objects;
- are organized hierarchically;
- and may be distributed.

The *Problem*

Significant resources invested in digitization, aggregation, organ-ization, and description of the information in a collection.

But to user (or client) collection appears to be simply an accumulation of objects and undifferentiated data.

Reason for the Problem

No agreed-upon semantics for navigating the collection, to locate and retrieve objects of interest.

Key objectives

- **Distinguish content from description.**
- **Exploit existing descriptive aids.**
- **Model relationships among objects.**
- **Provide semantics for queries to be interpreted consistently across servers.**
- **Data structures that provide semantics for navigation, to locate and retrieve objects of interest.**

An Object

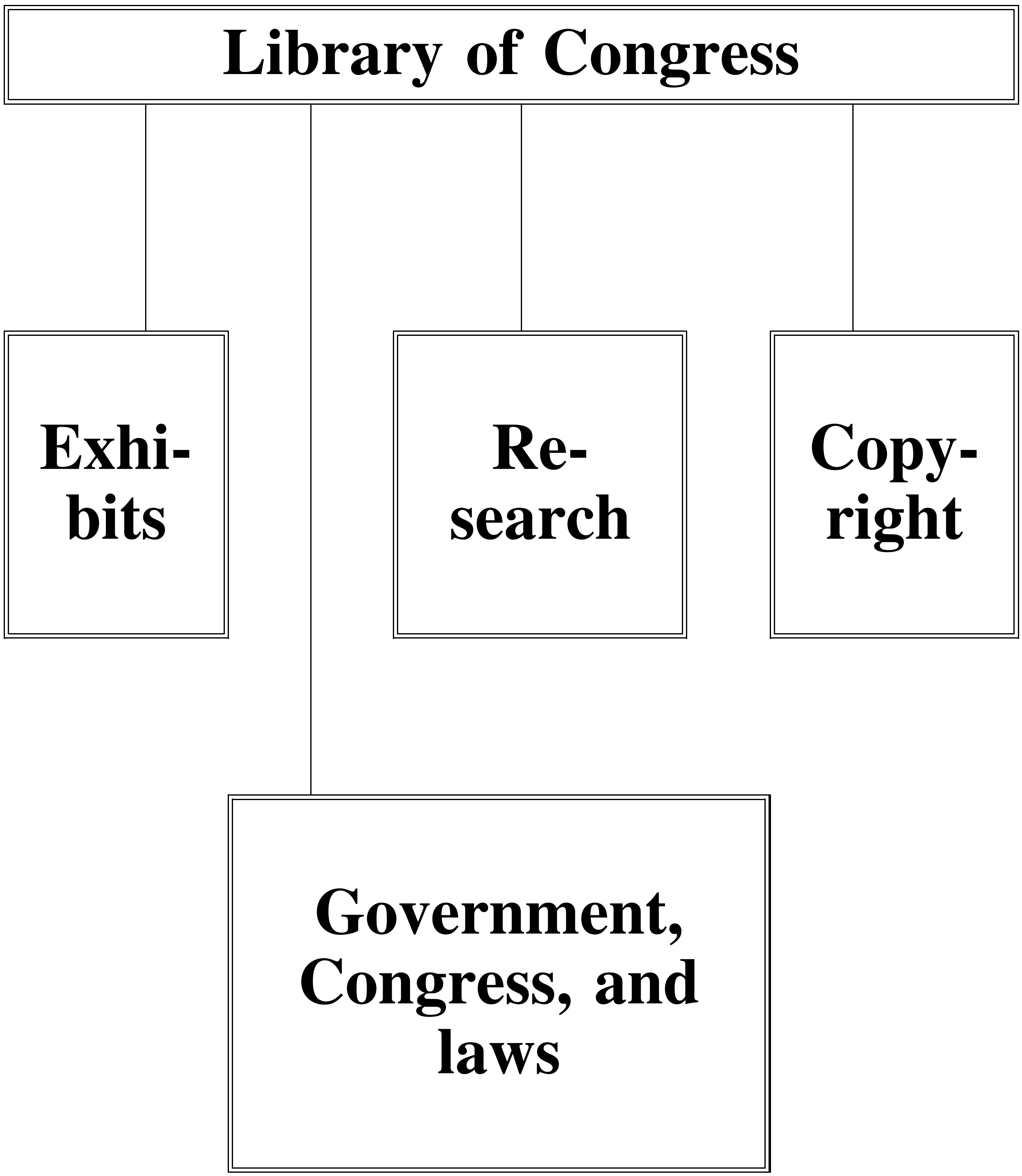
An *object* is a *Physical Object* or a *Digital Object*.

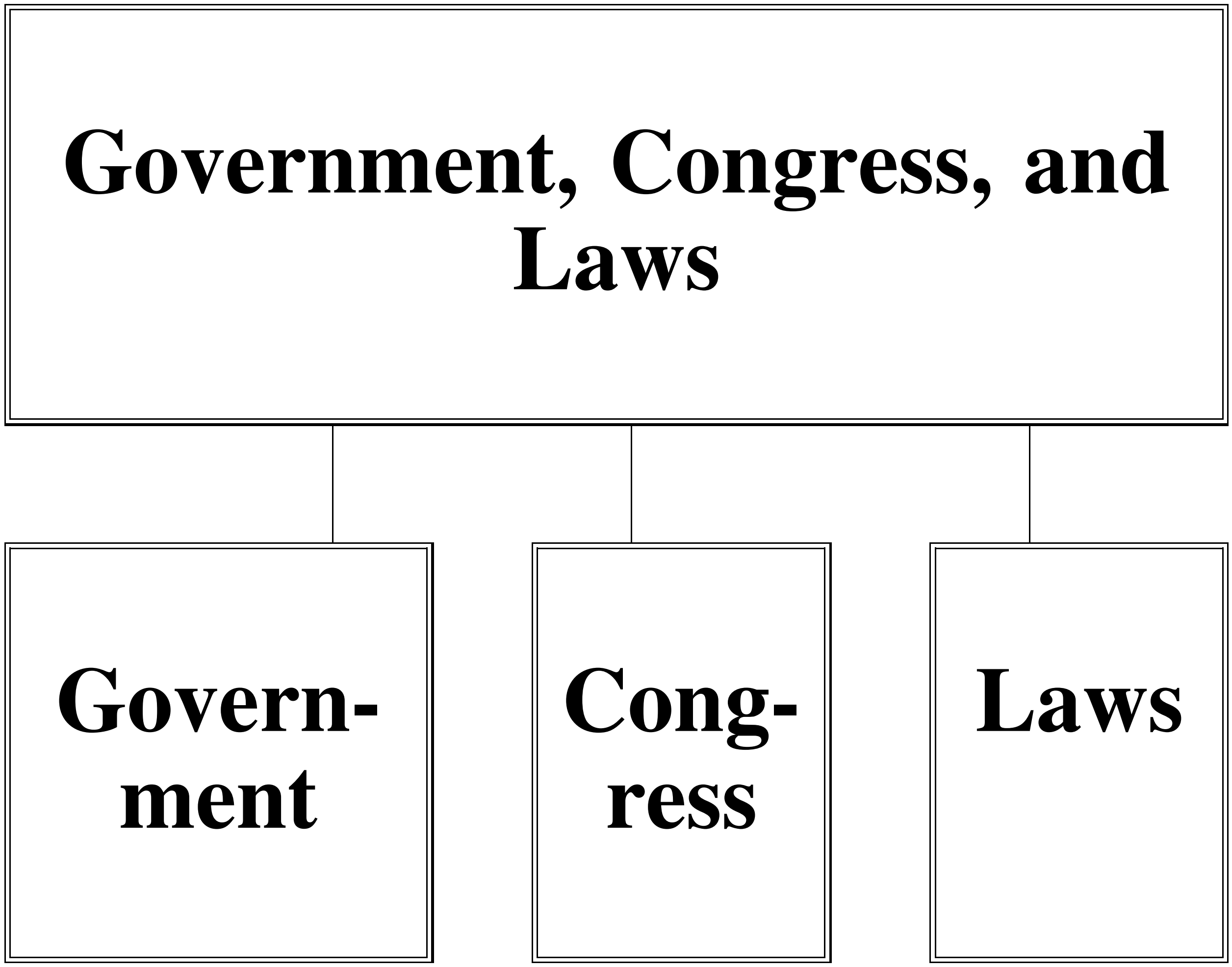
A Digital Object may be text, digitized image, audio, video, etc. In any case, its <u>opaque</u>.

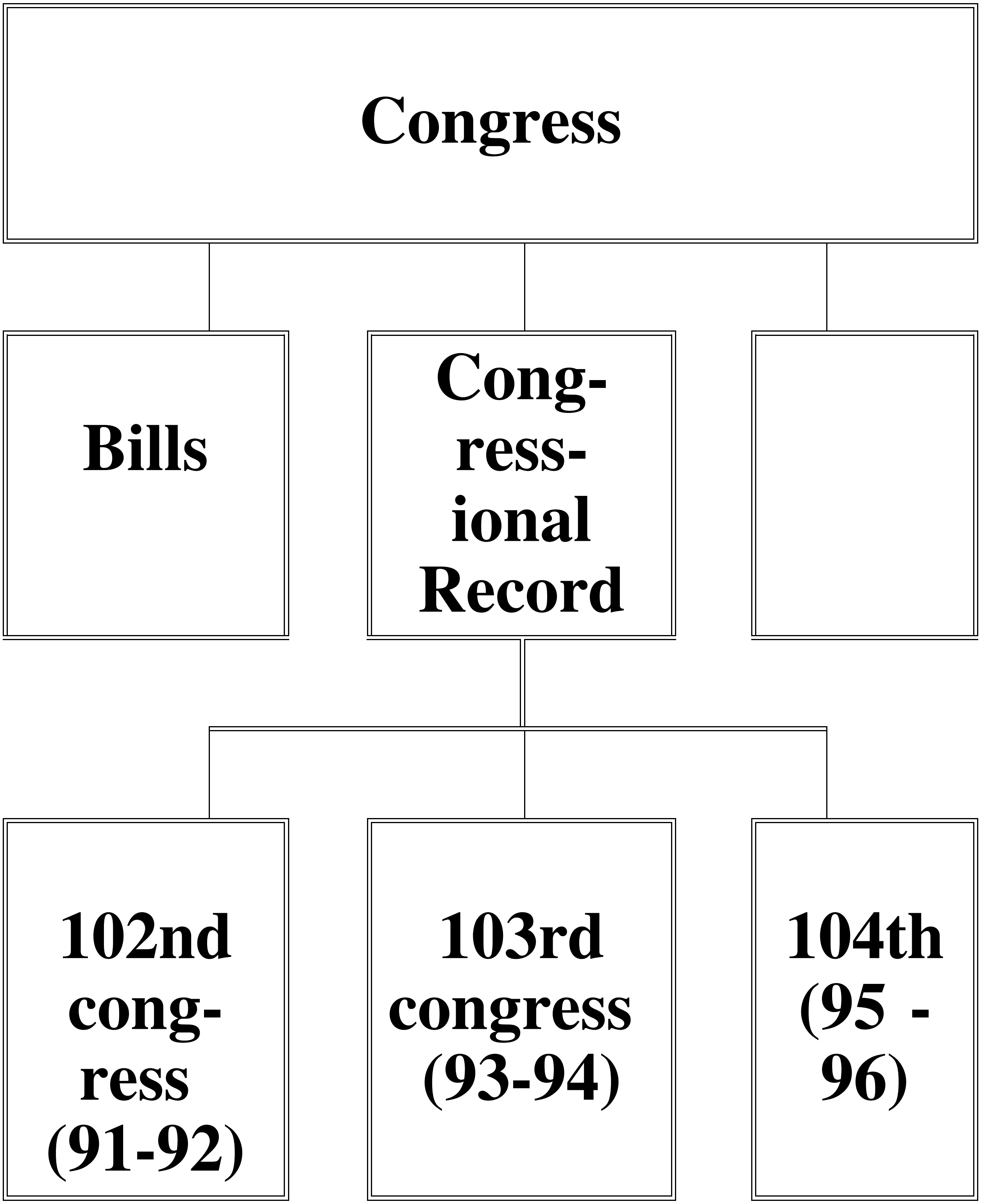
A Collection

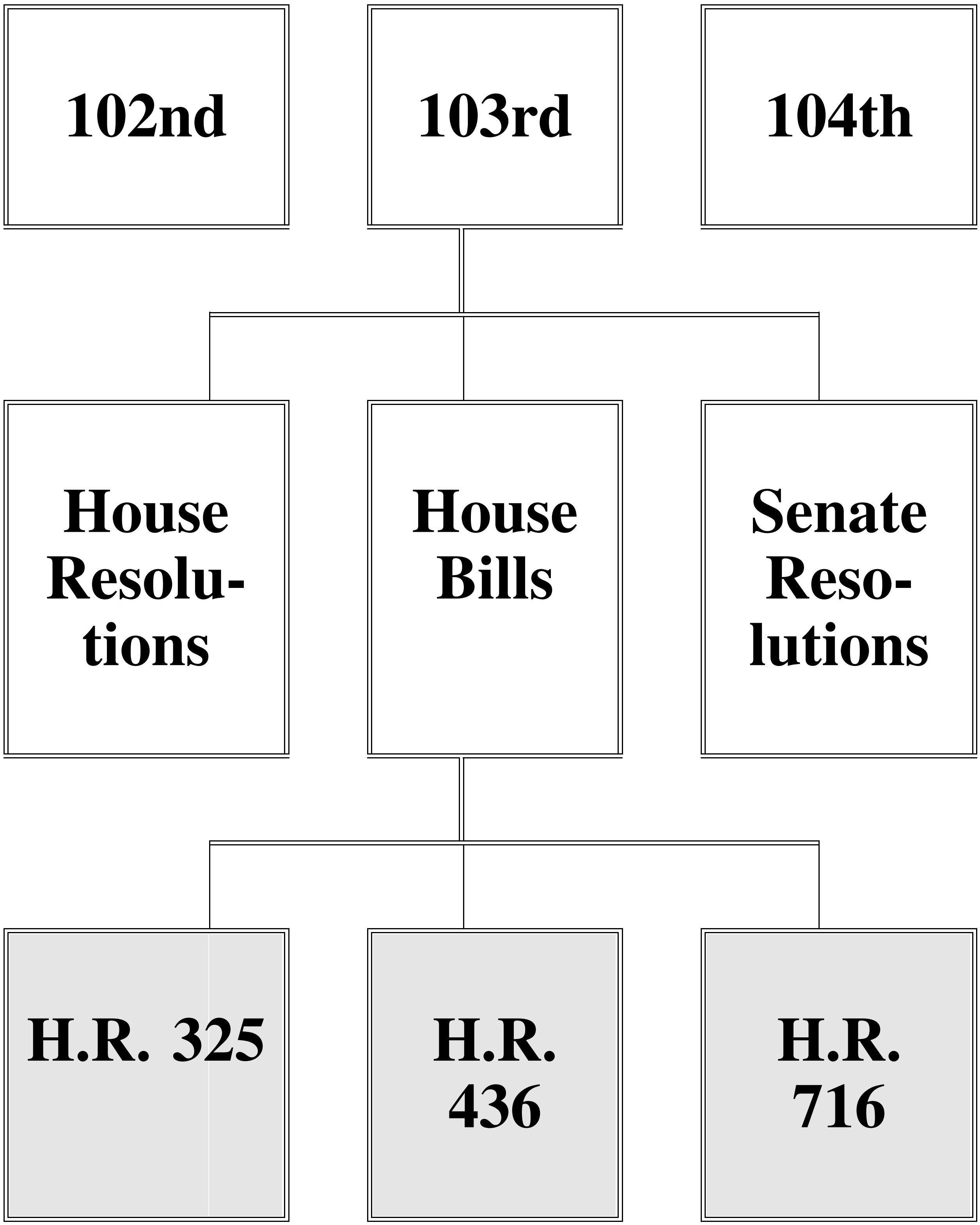
A *collection* is a group of related objects and/or collections: A tree, where leaf nodes are objects and non-leaf nodes are subcollections.

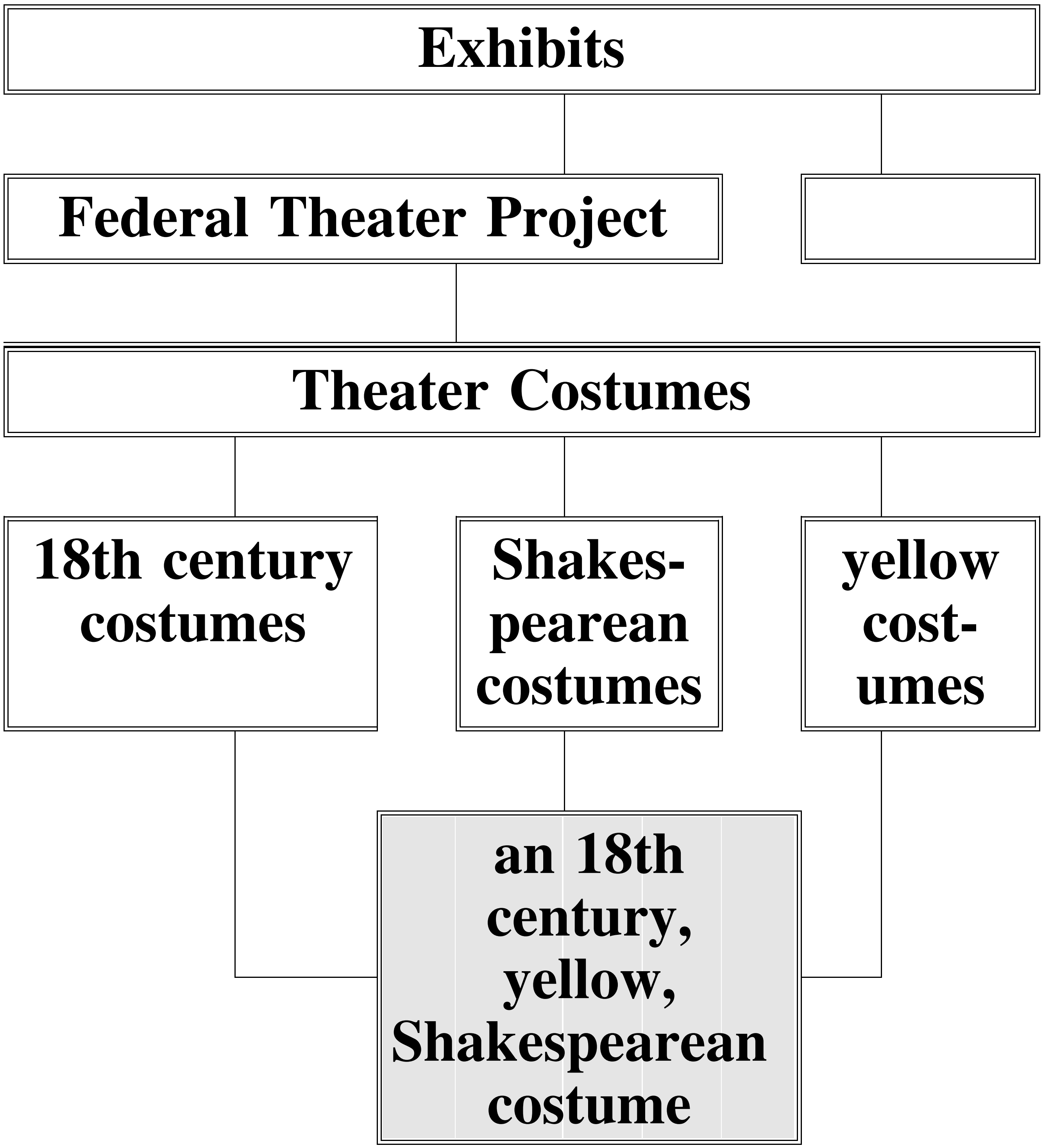
Examples


Library of Congress
Exhi-
bits
Re-
search
Copy-
right
Government,
Congress, and
laws


Government, Congress, and Laws
Govern-ment
Cong-ress
Laws


Congress
Bills
Cong-
ress-
ional
Record
102nd
cong-
ress
(91-92)
103rd
congress
(93-94)
104th
(95 -
96)


102nd
103rd
104th
House Resolutions
House Bills
Senate Resolutions
H.R. 325
H.R. 436
H.R. 716


Exhibits
Federal Theater Project
Theater Costumes
18th century costumes
Shakes-pearean costumes
yellow cost-umes
an 18th century, yellow, Shakespearean costume

An object or a collection may have one or more Associated Descriptions:

- **Finding Aid**
- **Encoded Archival Description**
- **Exhibition Catalog**
- **Cataloging Record (MARC)**
- **......**

The *Collections* Profile Models:

- **Objects**
- ***Digital* Objects**
- **Collections**
- **Associated Descriptions**
- **Descriptive Record**
- ***Collection* Descriptive Record**
- ***Object* Descriptive Record**

Collection
Descriptive Record

Brief text Description

Associated Description

Related Collections

Enumerated Members

Collection
Descriptive Record

Brief text Description

Associated Descriptions

- brief description
- category
- size, format
- actual Associated Description

Related Collections

Enumerated Members

Collection
Descriptive Record

Brief text Description

Associated Descriptions

Related Collections

- collection name
- relationship
- relative level
- description of relationship

Enumerated Members

Related Collections: Relationships

- Parent
- Superior
- Sibling
- Context
- "might be of interest"

Enumerated Members

First Object
Brief description
pointer to the Descriptive Record or Object

Second Object
Brief description
pointer

Third Object
Brief description
pointer

Object
Descriptive
Record

Brief Text Description

Associated Description

Related Collections

Digital Object

Digital Object

type of object

category of object

metadata: size, cost, format

actual object or a pointer

authoritative flag

Brief (Collection) Descriptive Record

Element	Occurrence
brief Description	include if available
collection Name	include
associated Description	optional, but include briefDescriptionOfAD and categoryOfAD only
related Collection	exclude
enumerated Members	exclude

Brief (Object) Descriptive Record

Element	Occurrence
briefDescription	include if available
category OfObject	include if available
Digital Object	include only actualDO, if available, and if so, include only variant information (no content)
associated Description	optional, but include briefDescriptionOfAD and categoryOfAD only
related Collection	Exclude

Navigation (Collection) Descriptive Record

Element	Occurrence
brief Description	exclude
collection Name	include
enumerated Members	include all occurrences
associated Description	exclude
related Collection	include all occurrences

"What sort of abstraction is this Descriptive Record anyway?"

(digression)

Z39.50 abstractions

- **Database**
- **Record**

For Searching:

- **Result Set**

For Retrieval:

- **Abstract Database Record**
- **Retrieval Record**
- **Diagnostic Record**

Result Set

Result set item 1	**Database:** ***Maps***	**Record:** ***10010***
Result set item 2	**Database:** ***Journals***	**Record:** ***20***
Result set item 3	**Database:** ***Image***	**Record:** ***10***
Result set item 4	**Database:** ***Bibliographic***	**Record** ***2000***

database record

Database Schema

abstract database record

Element Specification

abstract database record

Record Syntax

Retrieval Record

Z39.50 abstractions:

- **Database**
- **Record**
- **Result Set**

Collections Profile Abstractions:

- **Collection**
- **Digital Object**
- **Descriptive Record**

What is a "Record"?

- *Database* **record**
- *Abstract* **database record**
- *Retrieval* **record**

What sort or record, then, is the

- *Descriptive* **Record ?**

The Descriptive Record is:

For purposes of searching, a

- *Database* record

For purposes of describing the Descriptive record, an

- *Abstract* database record

And for purposes of retrieval, a

- *Retrieval* record

Pointers

Enumerated Members

First Object
Brief description
pointer to DR or Object

Object Descriptive Record

Digital Object
Actual Digital Object or pointer

Associated Description

Brief description
category
actual AD or pointer

Pointer

Element	Occurrence
database	occurs (in conjunction with recordId) if and only if alternativeIdentifier does not occur
recordId	occurs (in conjunction with database) if and only if alternativeIdentifier does not occur
alternative Identifier	occurs if and only if Database does not occur
typeOf Identifier	optional; occurs only if alternativeIdentifier occurs

Collection-1 Use Attributes

Record-id
Record-type

- **Descriptive Record**
- **Collection Descriptive Record**
- **Collection Descriptive Record for a context collection**
- **Object Descriptive Record**
- **Digital Object**

Object-type

- **Digital Object**
- **Digital Object for which there is a Physical Object**
- **Digital object for which there is no Physical Object**
- **Physical Object**

Parent-collection
Superior-collection
Related-collection
Collection-name

Conformance Levels

- **Basic Conformance**
  - V2
  - DR schema
  - Collection-1 attribute set
  - GRS-1

- **Basic V3 Conformance**
  - Basic + V3

- **Enhanced V3 Conformance**
  Adds:
  - Present parameter CompSpec
  - Element spec. eSpec-1
  - Variant set variant-1
  - Fragmentation
  - Metadata and appliedVariant of GRS-1

Scope limitations of Collections Profile, left for companion profiles

- Digital objects opaque.
- Associated Descriptions opaque.
- Modelling of complex relationships among objects.
- Guidelines to determine if information is description or content.
- organizing principles.

CIMI

Consortium for the Computer Interchange of Museum Information

Project CHIO

Cultural Heritage Information Online

- **CHIO Structure**
  Development of SGML DTD for CIMI

- **CHIO Access**
  Z39.50 profile

Museum Objects

- Artist record
- Exhibition catalogues
- Wall labels
- Images with associated text

Highlights of the CIMI Profile

- Version 2
- Proximity searching
- Access control
- CIMI-1 Attribute set
  New attribute types:
    - context
    - authority
- "Chunk" model
  - GRS Hits
  - target token
  - highlighting

<u>Use</u> and <u>Context</u>

A Use attribute corresponds to an access point, and may be nested within a "context", represented by a Context attribute. (Contexts may be nested.)

(1) <p><topic access-point="award" value="Prix de Rome"></topic> ... some text about the Prix de Rome ... </p>

(2) <context CHIO="award"> ... some text about awards ... </context>

" Use = 'award' " means search the 'value' part within (1).

" Context = 'award' " means search the text in (2).

Nested Contexts:
(context="creation" (context="creator" (topic="person" with value = "de Lorenzo")))

Z39.50 Profile for the National Digital Library

Search Within:

Brief description.
Associated Description.
Associated Description of a specific category.
Digital Object.
Digital Object of a specific category.

Search for records of a specific:

Physical format (book, map, etc.).
Language.
Chronological *subset* (e.g. > 103rd Congress).

Associated Description Categories:

Cataloging Record
Archival Register
Header
Web Page

Digital Object Categories

Text
Sound
Still Image
Moving Image

LC Collections

Detroit Publishing Company Collection

About 25.000 photos. U.S. scenes (mostly), including buildings, towns, cities, universities, battleships yachts, resorts, natural landmarks, and industry.

Nation's Forum Collection

Recordings made between 1918 and 1920 to preserve the voices of prominent Americans; World War I topics, postwar issues, and the 1920 presidential election. (Includes three speeches by Eamonn DeValera, who led the fight to establish the Republic of Ireland.)

WPA Life Histories Collection

Life History Manuscripts from the Folklore Project, WPA Federal Writers' Project, 1936-40. 2900 documents from 300 writers from 24 states. 2,000-15,000 words each. 23,000 page images total. Histories describe informant's family education, income, occupation, political views, religion and mores, medical needs, diet and miscellaneous observations.

Finding Aids for Shirley Jackson Papers

Master American short-story writer and novelist of the mid-20th century her papers, given to LC in 1967, consist of diaries, journals, correspondence, literary manuscripts, and miscellaneous papers. Best known for modern Gothic horror, in particular for her classic story, The Lottery, 1948. She also wrote stories about contemporary domestic life.

Coolidge-Consumerism Collection

17,000 pages (images) of 1920s primary-source materials: manuscript, monograph, and serials Also photos and motion pictures. Documents various aspects of economic life in the U.S. during the 1920s. Includes the Calvin Coolidge Papers, focusing on the life of Calvin Coolidge during the six years he was president (1923-1929).

Legislative Information System

LC Congressional Research Service

Attribute Sets in Z39.50

Z39.50-1992

Bib-1 -- Bibliographic Databases

Z39.50-1995

Bib-1

Exp-1 -- Explain

Ext-1 -- Extended Services

Discipline or Profile Specific

GILS
STAS
Collection-1
CIMI-1
NDL

Bib-1 Use Attributes

Personal name	1	PA subject	26
Corporate name	2	LC subject heading	27
Conference name	3	RVM subject heading	28
Title	4	Local subject index	29
Title series	5	Date	30
Title uniform	6	Date of publication	31
ISBN	7	Date of acquisition	32
ISSN	8	Title key	33
LC card number	9	Title collective	34
BNB card no.	10	Title parallel	35
BGF number	11	Title cover	36
Local number	12	Title added title page	37
Dewey classification	13	Title caption	38
UDC classification	14	Title running	39
Bliss classification	15	Title spine	40
LC call number	16	Title other variant	41
NLM call number	17	Title former	42
NAL call number	18	Title abbreviated	43
MOS call number	19	Title expanded	44
Local classification	20	Subject precis	45
Subject heading	21	Subject rswk	46
Subject Rameau	22	Subject subdivision	47
BDI index subject	23	No. nat'l biblio.	48
INSPEC subject	24	No. legal deposit	49
MESH subject	25	No. govt pub.	50

No. music publisher	**51**
Number db	**52**
Number local call	**53**
Code-- language	**54**
Code-- geogr. area	**55**
Code--institution	**56**
Name and title	**57**
Name geographic	**58**
Place publication	**59**
CODEN	**60**
Microform gen.	**61**
Abstract	**62**
Note	**63**
Author-title	**1000**
Record type	**1001**
Name	**1002**
Author	**1003**
Author-name -- personal	**1004**
Author-name -- corporate	**1005**
Author-name -- conference	**1006**
Identifier	**1007**
Subject -- LC children's	**1008**
Subject name -- personal	**1009**
Body of text	**1010**
Date/time added	**1011**
Date/time modified	**1012**
Authority/ format id	**1013**
Any	**1016**
Server-choice	**1017**
Publisher	**1018**
Record-source	**1019**
Editor	**1020**
Bib-level	**1021**
Geographic-class	**1022**
Indexed-by	**1023**
Map-scale	**1024**
Music-key	**1025**
Related-periodical	**1026**
Report-number	**1027**
Stock-number	**1028**
Thematic-number	**1030**
Material-type	**1031**
Doc-id	**1032**
Host-item	**1033**
Content-type	**1034**
Anywhere	**1035**
Author-Title-Subject	**1036**

Type	Value	Type	Value
Use	1	Structure	4
Relation	2	Truncation	5
Position	3	Completeness	6

Relation	Value	Relation	Value
less than	1	not equal	6
less or equal	2	phonetic	100
equal	3	stem	101
greater or equal	4	relevance	102
greater	5	alwaysMatches	103

Structure	Value	Structure	Value
phrase	1	name (un-norm.)	102
word	2	structure	103
key	3	urx	104
year	4	free-form-text	105
date (normalized)	5	document-text	106
word list	6	local number	107
date (un-norm.)	100	string	108
name (normalized)	101	numeric string	109

Problems

Duplication of common attributes

Tag-space

Rules of inheritance

Attribute set proliferation (interoperability)

Semantic ambiguities
- **Repeated attributes**
- **Omitted attributes**

Term typing/ Representation/normalization

Attribute set inter-relationships/interactions

Dimensionality

Version 3 Enhancements

- **Attributes from Multiple attribute sets in same query;**
- **...even in same operand.**
- **Term typing.**
- **String Attributes.**
- **Complex attributes.**
- **Semantic indicator.**
- **Additional bib-1 semantics.**

ASN.1

-- Definitions for RPN query

```
RPNQuery ::= SEQUENCE{
   attributeSet      AttributeSetId,
   rpn               RPNStructure}
```

```
RPNStructure ::= CHOICE{
  op              [0] Operand,
  ......
```

```
Operand ::= CHOICE{
  attrTerm
AttributesPlusTerm,    .......
```

```
AttributesPlusTerm ::= [102]
    IMPLICIT SEQUENCE{
    attributes        AttributeList,
      term            Term}

Term ::= CHOICE{
  general  [45] IMPLICIT OCTET STRING,
                -- values below may be used
               -- only for version 3
  numeric   [215]  IMPLICIT INTEGER,
  characterString .. IMPLICIT
                    InternationalString,
    oid...
    dateTime...
    external...
    integerAndUnit....
    null....


  AttributeList ::= [44]  IMPLICIT
  SEQUENCE OF AttributeElement
```

```
AttributeElement ::=    SEQUENCE{

    attributeSet [1] IMPLICIT
                 AttributeSetId OPTIONAL,
        -- Must be omitted in version 2. If
        -- included, overrides value of
        -- attributeSet in RPNQuery above,
        -- but only for this attribute.

  attributeType   [120] IMPLICIT INTEGER,

  attributeValue    CHOICE{
    numeric [121] IMPLICIT INTEGER,
        -- must select For version 2

   complex [224] IMPLICIT
                 SEQUENCE{
                   list      [1] IMPLICIT
                               SEQUENCE OF
                             StringOrNumeric,
               semanticAction [2]
                      IMPLICIT SEQUENCE
                      OF INTEGER
                         OPTIONAL}}
```

Bib-1 Rules

If a target does not support a given attribute list, fail the search and supply an appropriate diagnostic.

A given attribute type may appear in an attribute list zero, once, or more than once.

- If zero, no default behavior.
- If exactly once, the attribute value specifies the preferred target behavior with respect to that attribute type.
- Recommended an attribute type not occur more than once unless semanticAction is included.

SemanticAction is a single integer or a sequence of two integers.

For first integer:

1. **May not substitute another attribute. If none in the list is supported, fail the search.**
2. **May substitute another attribute, but only if none in the list is supported.**
3. **May substitute another attribute at target discretion (even if one or more in the list is supported).**

Second integer supplied iff there are multiple attributes in the list.

1. **Select the first supported attribute in the list.**
2. **Select the best attribute in the list.**

Attribute Architecture

- **Assume version 3**
  Compatibility with version 2 should not confine the development.

- **ZIG Role**
  ZIG should only maintain attribute sets that play a structural role:
    ~~bib-1~~
    exp-1
    ext-1
    core set?
  However, architecture guidelines should come from the ZIG

- **Phase out bib-1**

- **De-couple attributes from Z39.50**

- **Attribute set class**

  Template for attribute sets. sets in a class have same types; types have common general semantics. Inheritance rules, relationships among attribute sets, apply *within a class*; semantics of attributes from multiple sets within a single query.

  E.G. Attribute class for attributes that explicitly correspond to retrieval schema elements.

- **Architectural primitives**

  Specific attribute types (e.g Use, Relation, Structure attributes) that transcend specific classes and deserve definition within the standard itself.

- **Examine relationship of attributes to:**
  - **data element dictionaries**
  - **record interchange formats**
  - **document structure definitions**
  - **metadata interchange structures**