Schemas

Custom, Dublin Core, VRA, and MODS

What is a metadata schema?

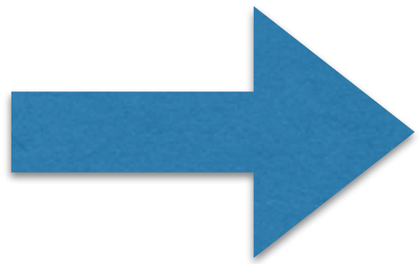
- an element set
- “…a logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata specifically as regards the semantics, the syntax and the optionality (obligation level) of values.” — NISO

Encoding

- Metadata schemas can use encoding schemes for specific elements and also the entire dataset
- Encoding schemes are “Controlled lists of all the applicable values in natural language and/or as a syntax-encoded text string designed for machine processing.”—NISO
- Ex. Rules for entering data like dates, names of people, etc.

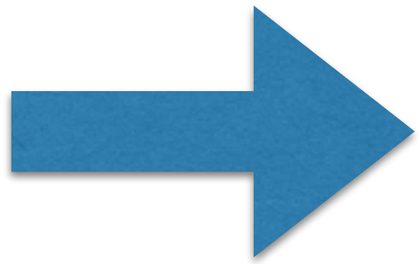
Some schemas require specific **encoding** standards for the entire dataset; others can be expressed in multiple forms

Ex. MODS
TEI



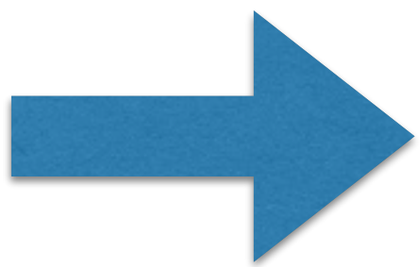
XML

Ex. RDF



XML, N-Triples
Turtle, JSON

Ex. Dublin Core



XML, CSV, etc

How to choose a metadata schema

Assess the needs of your collection and the platform you'll be using.

Ask yourself…

- Who will be using the collection?
- Who is the collection cataloger?
- How much time/money do you have?
- How will your collection be accessed?
- How is your collection related to other collections?

—From Marie Kennedy's "Nine questions to guide you in choosing a metadata schema"

- What is the scope of your collection?
- Will your metadata be harvested?
- Do you want your collection to work with other collections?
- How much maintenance and quality control do you wish?

—From Marie Kennedy's "Nine questions to guide you in choosing a metadata schema"

Schemas we're going to talk about today

- Custom
- Dublin Core
- VRA Core (VRA = Visual Resources Association)
- MODS (Metadata Object Description Schema)

Custom schema

Pros:

- Gives you the most flexibility to describe your objects
- You define your metadata elements and can choose whether you want a flat or hierarchical

Cons:

- Will have limited interoperability unless you have mappings to other schemas or ontologies
- Limited interoperability means your metadata will be difficult for **machines** to share
- You are responsible for documenting your schema

Tips for creating custom schema*

- think about all items in the dataset; look for patterns in the data
- make decisions for data that doesn't fit patterns; try to be consistent in treatment
- make decisions for dealing with messy data; spelling variants, abbreviations, etc.
- be iterative in developing the schema
- document your decisions
- add value with controlled vocabularies, geographic coordinates, etc

*These tips can also apply when using a standard schema

Exercise 1:
Create Some Metadata About A Historical Object

(Custom Schema)

Dublin Core:

- one of the most established and long-standing metadata schemas

- consists of 15 core elements (See #3 in Table of Contents at http://dublincore.org/documents/dcmi-terms/#H3)

- **Elements:** contributor; coverage; creator; date; description; format; identifier; language; publisher; relation; rights; source; subject; title; type

Dublin Core

Pros:

- Can quickly provide quality description for individual objects as well as collections
- Simple and easy to use because schema is very flat (non-hierarchical)
- Flexible; can be qualified/modified depending on need
- Has mappings to popular ontologies (ex. SKOS)
- Has robust, stable community and lots of documentation
- DC Terms is one of the most common namespaces for RDF
- Is commonly used in library systems like institutional repositories; if you want to archive and share your work, you'll probably describe it in DC

Dublin Core

Cons:

- Is very flat schema (non-hierarchical), which makes it difficult to describe complex objects

- Limited number of elements; you may have data that does not map cleanly to any of the DC elements

Dublin Core Elements (in-depth)

dc.creator: “An entity primarily responsible for making the resource”; often personal or corporate name; can be taken from controlled vocabulary such as LCNAF or VIAF

dc.contributor: “An entity responsible for making contributions to the resource”; often personal or corporate name; can be taken from controlled vocabulary such as LCNAF or VIAF

My recommended format for uncontrolled names:
surname, first name, life dates

Ex. Hansen, Carolyn Marie, 1981-

Dublin Core Elements (in-depth)

dc.coverage: “The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant”; best practice is to use controlled vocabulary from the Getty for geo names (http://www.getty.edu/research/tools/vocabulary/tgn/index.html)

Ex. - Seattle (inhabited place)

- 20th century

Dublin Core Elements (cont.)

dc.date: A point or period of time associated with an event in the lifecycle of the resource. Best practice is to format dates according to ISO 8601 (https://en.wikipedia.org/wiki/ISO_8601).

Note: ISO does not allow circa dates, but does allow questionable dates and date ranges

ISO Format: YYYY; YYYY-MM; YYYY-MM-DD

Ex.: June 5, 2016 = 2016-06-05

Dublin Core Elements (cont.)

dc.description: The most general and flexible of all DC fields. Basically a free-text note field for anything you want to say about the object. Can include things like abstracts, tables of contents, historical/biographical information, etc.

dc.format: The file format, physical medium, or dimensions of the resource (ex. size or duration). Best practice is to use a controlled vocabulary such as Internet Media (MIME) types (http://www.iana.org/assignments/media-types/)

Dublin Core Elements (cont.)

dc.identifier: An unambiguous reference to the resource within a given context (does not need to be persistent, i.e., a URI). Examples include accession numbers or local internal ordering systems.

dc.language: Language or the resource. Best practice is to use a controlled vocabulary, such as ISO 639-2 (https://www.loc.gov/standards/iso639-2/php/code_list.php)

Dublin Core Elements (cont.)

dc.publisher: An entity responsible for making the resource available. Often a personal or corporate name. Can be taken from controlled vocabulary such as LCNAF or VIAF.

dc.relation: A related resource. This is often a larger collection that the item belongs to.

dc.rights: Information about rights held in and over the resource. We'll talk about this more later.

Dublin Core Elements (cont.)

dc.source: A related resource from which the described resource is derived.

dc.subject: Topic of the resource. Best practice is to use a controlled vocabulary.

dc.title: Title of the resource.

Dublin Core Elements (cont.)

dc.type: The nature or genre of the resource. Best practice is to use a controlled vocabulary like DCMI Type Vocabulary (http://dublincore.org/documents/dcmi-type-vocabulary/). To describe the file format, physical medium, or dimensions of the resource, use the Format element.

Ex. - StillImage

- Text

- MovingImage

- PhysicalObject

What about Qualified Dublin Core?

Exercise 3:
Create Dublin Core Metadata

Note: For now, ignore everything we just talked about re: controlled vocabularies; just try to match information to the DC field

VRA Core:

- data standard for description of images and works of art and culture

- consists of 19 core elements, which can be modified by subelements and attributes (https://www.loc.gov/standards/vracore/VRA_Core4_Outline.pdf)

- **Elements:** work, collection or image; agent; culturalContext; date; description; inscription; location; material; measurements; relation; rights; source; stateEdition; stylePeriod; subject; technique; textref; title; worktype

VRA

Pros:

- Has more granular level of description with subelements and attributes that work well for describing visual objects and their digital surrogates

- Mappings exist to other ontologies, although mappings are not as robust as for Dublin Core

- Has robust, stable community and lots of documentation

Cons:

- May be more difficult to use than Dublin Core because the schema is hierarchical

- Does not work as well for textual materials or collections (ex. archival collections containing mostly text)

VRA Elements (in-depth)

Work, Collection, or Image

Attributes: id

This element is important when expressing VRA in XML. This is an upper-level wrapper element that contains the rest of the description (i.e. the other VRA elements). Here you define your piece as a **Work** (a build or created object), **Collection** (an aggregate of such objects), or **Image** (a visual surrogate of such objects)

VRA Elements (in-depth)

AGENT

Subelements:

name
Attributes: type
culture
dates
Attributes: type
earliestDate
latestDate
role
attribution

The names of an individual, group, or corporate body that has contributed to the design, creation, production, manufacture of the work or image.

Agent Example

VRA Core Element	XML element	XML sub-element	*XML attribute*	XML sub-element	Data example (display value in bold)
AGENT	**agent**				**School of Peter Paul Rubens (1577-1640)**
		name			Rubens, Peter Paul
			type		personal
			vocab		ULAN
			refid		500002921
		culture			Flemish
		dates			**1577-1640**
			type		life
				earliestDate	1577
				latestDate	1640
		dates			
			type		activity
				earliestDate	1590
				latestDate	1640
		role			painter (artist)
			vocab		AAT
			refid		300025136
		attribution			School of

https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf

Agent Example (XML)

```
<agentSet>
    <display>School of Peter Paul Rubens (1577-1640)</display>
        <agent>
          <name type="personal" vocab="ULAN" refid="500002921 "dataDate="2006-09-22">Rubens,
    Peter Paul</name>
          <culture>Flemish</culture>
          <dates type="life">
                <earliestDate>1577</earliestDate>
                <latestDate>1640</latestDate>
          </dates>
          <dates type="activity">
                <earliestDate>1590</earliestDate>
                <latestDate>1640</latestDate>
          </dates>
          <role vocab="AAT" refid="300025136">painter (artist)</role>
          <attribution>School of</attribution>
        </agent>
</agentSet>
```

https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf

VRA Elements (in-depth)

Cultural Context

The name of the culture, people, or adjectival form of a country name from which a Work, Collection, or Image originates, or the cultural context with which the Work, Collection or Image has been associated.

Ex. English

VRA Elements (in-depth)

DATE

Attributes:

type

Subelements:

earliestDate

circa

latestDate

circa

The names of an individual, group, or corporate body that has contributed to the design, creation, production, manufacture of the work or image.

Date Example

VRA Core Element	XML element	*XML attribute*	XML subelement	Data example (display value in bold)
DATE	**date**			**created 1520-1525**
		type		creation
			earliestDate	1520
			latestDate	1525
		source		Grove Dictionary of Art Online
		href		http://www.groveart.com
		dataDate		2005-06-08

https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf

VRA Elements (in-depth)

Description

A free-text note about the Work, Collection, or Image, including comments, description or interpretation, that gives additional information not recorded in other categories.

VRA Elements (in-depth)

INSCRIPTION

Subelements:

author
position
text

Attributes: *type*

All marks or written words added to the object at the time of production or in its subsequent history, including signatures, dates, dedications, texts, and colophons…

Inscription Example

VRA Core Element	**XML element**	**XML subelement**	*XML attribute*	**Data example (display value in bold)**
INSCRIPTION	**inscription**			**Inscribed, on side of table at left "El Sueño de la Razon Produce Monstruos" (The Sleep of Reason Produces Monsters)**
		author		Goya, Francisco de
			vocab	ULAN
			refid	500118936
		position		on side of table at left
		text		El Sueño de la Razon Produce Monstruos
			type	text
			xml:lang	es
		text		The Sleep of Reason Produces Monsters
			type	translation
			xml:lang	en

https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf

VRA Elements (in-depth)

LOCATION

Attributes:
type
Subelements:
name
Attributes: type
refid
Attributes: type

The geographic location and/or name of the repository, building, site, or other entity whose boundaries include the Work or Image.

Location Example

VRA Core Element	**XML element**	*XML attribute*	**XML subelement**	*XML attribute*	**Data example (display value in bold)**
LOCATION	**location**				**Musée du Louvre (Paris, FR) Inv. MR 299**
		type			repository
			name		Musée du Louvre
				type	corporate
				xml:lang	fr
			refid		Inv. MR 299
				type	accession
			name		Paris
				type	geographic
				vocab	TGN
				refid	7008038
				extent	inhabited place
			name		France
				type	geographic
				vocab	TGN
				refid	1000070
				extent	nation

https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf

VRA Elements
(in-depth)

MATERIAL

Attributes:

type

The substance of which a work or an image is composed (this can get VERY granular)

Material Examples

VRA Core Element	XML element	*XML attribute*	Data example (display value in bold)
MATERIAL			**oil paint on canvas**
	material		oil paint
		type	medium
		vocab	AAT
		refid	300015050
	material		canvas
		type	support
		vocab	AAT
		refid	300014078
MATERIAL			**graphite on paper**
	material		graphite
		type	*medium*
		vocab	*AAT*
		refid	300011098
	material		paper (fiber product)
		type	*support*
		vocab	*AAT*
		refid	300014109

https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf

VRA Elements (in-depth)

MEASUREMENTS

Attributes:

type
unit

The physical size, shape, scale, dimensions, or format of the Work or Image. Can include measurements like volume, weight, or running time.

Material Examples

VRA Core Element	XML element	*XML attribute*	Data example (display value in bold)
MEASUREMENTS			**Base 3 cm (H) x 36 cm (W) x 24 cm (D)**
	measurements		3
		type	height
		unit	cm
		extent	base
	measurements		36
		type	width
		unit	cm
		extent	base
	measurements		24
		type	depth
		unit	cm
		extent	base

https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf

Material Examples

VRA Core Element	XML element	*XML attribute*	Data example (display values in bold)
MEASUREMENTS	**measurements**		**72**
		type	resolution
		unit	ppi
	measurements		**650**
		type	width
		unit	px
	measurements		**123**
		type	duration
		unit	min

https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf

Exercise 4:
Create VRA Core Metadata

MODS:

- bibliographic element set, designed particularly for library applications

- is XML-based, more granular than Dublin Core, with many elements, subelements and attributes (see: http://www.loc.gov/standards/mods/)

- The metadata community (in general) is moving away from MODS in favor of RDF and/or less complex schemas like Dublin Core

- We will work with MODS later in the week during exercises using NYPL's APIs