Subject representation

6.1.1. What is a subject?

Overview

This module discusses the concept of subject and explains how subjects are represented. It makes the transition from bibliographic access, which deals generally with physical description of information objects, to intellectual access, which deals with representing intellectual content, or what information objects are about. An important part of understanding subject representation is being able to define and recognize a subject, so this module addresses the question: What is a subject?

What is a subject?

A subject is a representation of the intellectual content of an information object, or its aboutness, topic, theme, expressed concepts or ideas, area of interest, or knowledge.

Subject representations are . . .

- Secondary information objects that describe the intellectual content (aboutness) of primary information objects (documents and queries)
- Words, phrases, sentences and codes that serve as representations

Subject representations . . .

- Serve as access points for searching
- Inform searchers about the intellectual content of objects
- Are distinguished from physical descriptions

The traditional view of a subject is based on bibliographic conventions for representing textual objects, such as books. This view distinguishes between what an object is about and what an object is, that is, between subject description of intellectual content and physical description of the information container or package. Catalogers typically create the subject description, which involves only one or a few fields in a record, separately from the physical description, which involves all the other fields.
The traditional view of a subject also assumes that some identifiable intellectual content exists. In reality, subjects are difficult or impossible to identify for a few textual objects and most nontextual objects. This is because subject description depends on . . .

- A subjective interpretation based on ambiguous, emotional content
- The domain expertise of the person doing the subject description
- The kinds of information objects being described

Some kinds of objects do not lend themselves to simple subject representation. Think about works such as images and music that have no words. How can one represent their subjects when the interpretation is entirely personal and even emotional?

So why distinguish subject from physical representation?

- To provide intellectual access versus bibliographic access
- To provide more access points for searching
- To distinguish between work (the intellectual entity) and text (a version of the work)
- To clarify representations of various kinds of subjects

---

**Distinguishing between subject and physical description**

While at first glance, it may seem easy to tell the difference between what an object is *about* and what an object *is*, this really is one of the classic challenges of representation. Here are some guidelines for sorting out the gray areas.

**Project Alert!** In section 3, the first thing you do is decide how many kinds of subjects you have

**How does one identify subjects?**

It depends on what is being represented.

For example, this abstract image . . . might be represented in a record with these fields.

These fields are all *physical description* because the image has no identifiable subject (although that is not true of all images).
However . . .

If the image is described in a book about art . . . the book might be represented by these fields.

Fields 1 to 3 are **physical description** of the book and fields 4 to 6 are **subject description** of what the book is about (the image).

To add to the confusion, elements or fields with the same or similar names can be either physical or subject description. This is evident in the record above, where both book and image are represented by fields for creator, title, and format.

**How does one decide whether element/field is subject or physical description?**

It depends on what they represent. Here are some typical elements/fields. In these examples the elements/fields represent physical description, or what an information object *is*:

- **Title (name of object)**
- **Name**
  - Person (in role such as creator)
  - Corporate body (in role such as creator)
  - Geographic area (as place of origin)
- **Time (as date created or published)**
- **Format (as physical container)**

In the next examples, the elements/fields represent intellectual content, or what an object is *about*:

- **Topic (meaningful content)**
- **Name**
  - Person (as in a biography)
  - Corporate body (as in a prospectus)
  - Geographic area (as in a travel guide)
  - Named entity (about artworks, buildings, departments, etc.)
- **Time (as in history; period)**
- **Form (literary: poetry, essays)**
Taylor (2009) discusses the kinds of subjects above and the difficulties of identifying subject concepts. She says it is particularly hard to distinguish between *form* and *subject*, when form may strongly influence subject. Taylor uses the term *form* to refer to physical description in a general sense. In cataloging and shelving, she says, "Because it was often difficult to separate the ideas of 'aboutness' from 'form,' as in the case of 'history,' which seems to incorporate elements of both, the concept of form has only recently begun to be treated differently in surrogate records".

**Note:** Two common element/field names that make this subtle distinction are *form* and *format*. In this course, *form* usually refers to intellectual content or purpose, whereas *format* refers to the physical container or technical function. Here are some examples.

### Usually considered to be subject description:

- Literary forms: poetry, essays
- Popular genres: romance (fiction, film), jazz (music)
- Type of information: correspondence, bibliography, statistics
- Arrangement of information: calendar, outline, dictionary
- Style or technique related to purpose or audience: comedy, drama, persuasion, cartoon
- Style or technique related to time period: Baroque (music), Impressionism (painting)

### Usually considered to be physical description:

- Physical media format: book, video, CD, photo, map
- Artifact format: sculpture, figurine, toy, shirt
- Communication mode: text, image, video, audio
- Technical digital format: ASCII/plain text, HTML, .pdf, .gif
- Version/part of work: edition, translation, chapter

**Project Alert!** These are only guidelines: decisions for your system may vary.

---

**Subjects in the Dublin Metadata Core Element Set**

The Dublin Core, as it is commonly called, is a set of 15 generic elements intended to represent any kind of digital information object. The elements were developed beginning in 1995 by teams of experts from many fields, including librarians, publishers, and system designers. As you can imagine, this was a slippery undertaking, and in fact the details are still being debated.

As a final guide to your thinking about subjects, take a look at five of these elements (Dublin Core Metadata initiative, 1999). The first three seem to be clearly related to intellectual content.
Subject

- **Definition**: The topic of the content of the resource.
- **Comment**: Typically, a Subject is expressed as controlled-vocabulary or natural-language terms or classification codes that describe a topic of the resource.
- Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

Description

- **Definition**: An account of the content of the resource.
- **Comment**: Description may include but is not limited to an abstract, table of contents, reference to a graphical representation of content, or a natural-language account of the content.

Coverage

- **Definition**: The extent or scope of the content of the resource.
- **Comment**: Coverage typically includes spatial location (place name or geographic coordinates), temporal period (period, date, or date range) or jurisdiction (such as a named administrative entity).
- Recommended best practice is to select a value from a controlled vocabulary (e.g., Thesaurus of Geographic Names) and, where appropriate, to use named places or time periods instead of numeric identifiers such as sets of coordinates or date ranges.

The next element seems to be clearly related to the physical object.

Format

- **Definition**: The physical or digital manifestation of the resource.
- **Comment**: Typically, Format may include the media type or dimensions (size, duration) of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource.
- Recommended best practice is to select a value from a controlled vocabulary, for example: HTML, ASCII, jpeg, mpeg.
This next element can go either way: the term *genre*, which seems to relate to subject, is used in the definition and comment, but the best-practice examples seem to relate to physical description.

**Type**

- **Definition**: The nature or genre of the content of the resource.

- **Comment**: Type includes terms describing general categories, functions, genres, or aggregation levels for content.

- Recommended best practice is to select a value from a controlled vocabulary, for example: collection, dataset, event, image, interactive resource, service, software, sound, text.

**Note**: In her discussion of all 15 Dublin Core elements, Taylor (2004, p. 172) says the elements Title, Source, and Relation relate to information content. This conflicts with conventional definitions of physical description elements/fields used in this course. For a more complete picture of the Dublin Core, see other assigned readings on metadata.

---

**Project Alert!** The names of Dublin Core Elements are intended to be generic. Description and Type in particular are poor choices of element/field names for the IOP because they could apply to anything. Use more specific names, such as Subject and Format.

---

**Making choices**

As you can see, the task of determining what is a subject is not at all simple! This is why subject description presents such an fascinating challenge.

Any information organization system has three alternatives for subject representation:

- **One subject element/field**. This is the most common solution. The subjects are general topics or themes. The information objects are usually text-based but may be nontext in nature.

- **Multiple subject elements/fields**. This solution is often used for text-based objects with a great deal of complexity in intellectual content. Typically, a single element called Subject translates into two or more separate fields. An example is a collection of history books for which the database records have three subject fields called Subject, Period, and Country.

- **No subject element/field**. Usually this solution involves nontext objects that have no intrinsic aboutness (i.e., they just *are*; they don't cover any topic). However, it may be appropriate to create subject-like elements/fields such as historical context. For INFO 5200 purposes, you must have a project that has at least one subject field.

**Cites & sites**
