

# Cataloguing and metadata in libraries

LIBR 509  
Week 8 Mar 8 2022

## 1 Lecture

### 1.1 Brief recap of cataloguing and metadata

“Cataloguing” and “metadata”: (kind of interchangeable terms) about how we fully describe an item for the sake of discovery and access in an online collection

- (“cataloguing” often used for resources where there is a print version, “metadata” often used for born-digital resources)

Key questions:

- Which aspects of the resource do we need to represent?
- What are the constraints placed on those representations?—enforce consistency, reduce subjectivity, to make the catalogue more useful

Central concepts for traditional information-bearing resources: **title, creator, revisions**

Two components to what we’re doing when we are creating a catalogue

- **Contents:** e.g., description, access points (which part of that record is going to be hyperlinked, indexed), call number
- **Format:** how are you encoding information so that people can read it, machines can read it, it can be searched / indexed / sorted (e.g., card format, Machine-readable Cataloguing Record (MARC), BIBFRAME)

### 1.2 FRBR as a conceptual model for contemporary content standards

**FRBR: Functional Rules for Bibliographic Relationships** is the model for how what we know about an information-bearing resource has an internal structure to itself, which impacts when and how we describe different things about it

- Published 1998 as a result of international collaborative effort at rethinking cataloguing at conceptual level

FRBR is the conceptual basis that is responsible for some of the biggest changes reflected in contemporary library content standards

- FRBR greatly informed development of RDA (Resource Description and Access)

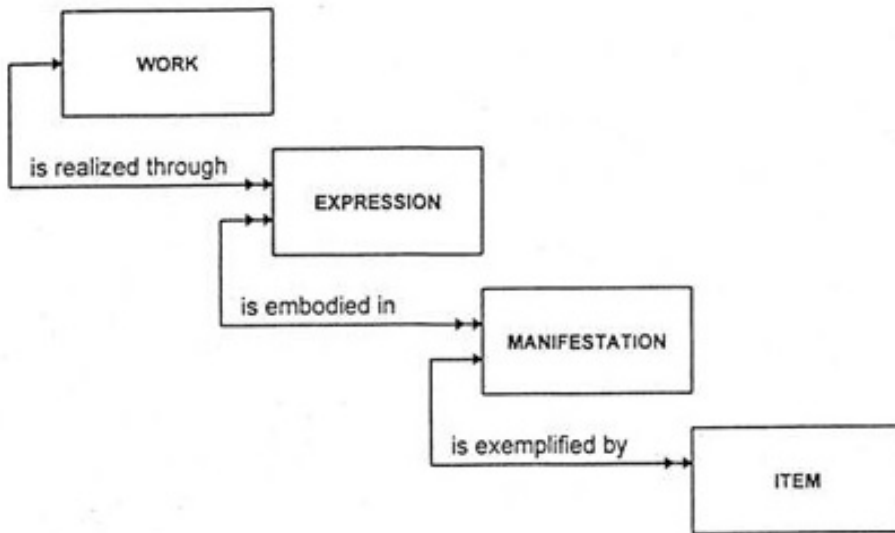
FRBR is trying to resolve: “there are relationships in the universe of books (and info bearing objects)”

- Among bibliographic entities **FRBR Group 1 Entities**
  - Relationships among the abstract intellectual work, its various versions, and their physical manifestations in items subject of today’s lecture
  - Equivalence, derivation, and part-whole relationships (chapter-book relationship)
- Responsibility relationships (agents and bibliographic entities; e.g., creator, publisher, owner) **FRBR Group 2**
- Subject relationships **FRBR Group 3**

### 1.2.1 FRBR Group 1 Entities

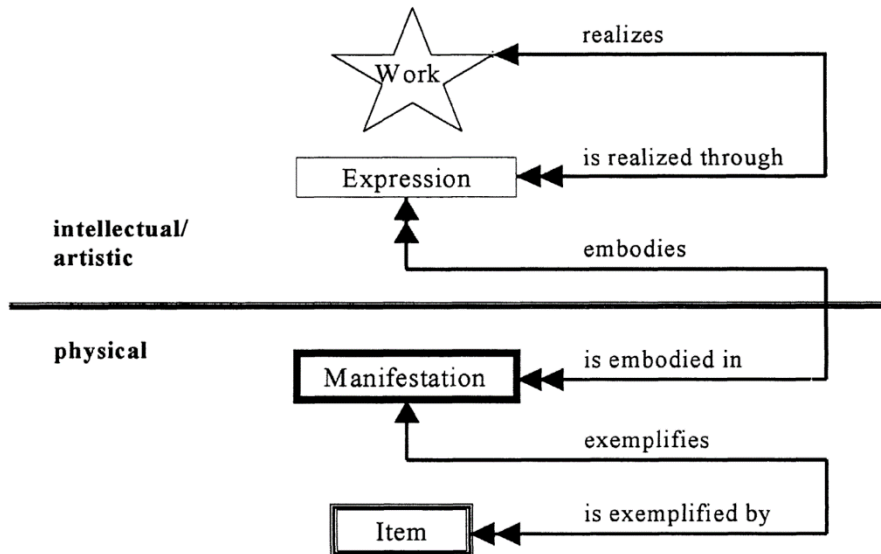
FRBR Group 1 Entities: about how you get from “works” (abstract / intangible intellectual creative content) to an “item” (a library can give access to)

Entity relationship diagram:



- “work” is realized through “expression” (e.g., text, play, novel—still no access until “manifestation”)
- “expression” is embodied in “manifestation” (tangible; represented by many different items; e.g., copy of publication is one “item” of that manifestation)
- Where there are 1 arrow, means only 1 thing goes in that direction; 2 arrows is “many” relationship”
  - A work can be expressed through many expressions (e.g., text of play, play as is performed, but that thing they are expressing is realization of one “work”)
  - “Expression” can be embodied in many “manifestations” (e.g., play text can be hand-written down, printed by publishers), and a “manifestation” can contain many “expressions” (e.g., “Collected Works of Shakespeare”)

Another version of the same diagram:



- Things that are digital are still “physical” (e.g., minerals and electricity)

#### 1.2.1.1 Work

**Work:** a distinct intellectual or artistic creation

- Tends to have: **creator**, **uniform title**, and **subject headings** (work is “about” things; at “work” level is where things gain their subjects)

#### 1.2.1.2 Expression

**Expression:** realization of a work

- Can have its own **Uniform title**
- Added contributor roles: e.g., **illustrators**, **translators**, **performers**, **language**
  - E.g., even though Shakespeare is creator of Hamlet, players are also responsible for expression of Hamlet, credited as performers
- **Edition statement** (which version is it?), **content type** (what type of material is this? Text? Performance? Etc.), **Word count** (about how long is the thing?) **Duration?**
- At this level you should know: who is responsible for the intellectual and creative work? What kind of way is it going to be expressed?

#### 1.2.1.3 Manifestation

**Manifestation:** a physical embodiment of intellectual and artistic content

- **Title**
- **Place and date of publication**

- **Publisher**, someone responsible for bringing into physical embodiment
- **Physical characteristics** e.g., Media Type (RDA), Carrier Type (RDA)

#### 1.2.1.4 Item

**Item:** the actual thing that someone can get a hold of

- **Call Number, Condition** (new, used, pages missing), **Provenance** (e.g., previously owned by xx), **Location** (“out on circulation”, “on 3<sup>rd</sup> floor”)

Example of description of our course textbook:

“Work & Expression” level	“Manifestation & Item” level
Creator: Robert Glushko Uniform title: The Discipline of Organizing Subject: information organization Language: English Edition statement: 4 <sup>th</sup> Content type: text	Place & date of publication: Sebastopol, California, 2016 Publisher: O’Reilly Media Carrier Type: Online resource Media type: Computer, PDF

- ➔ Depending on why a user needs access to an item, some of the most important aspects of that item might be its intellectual / creative content, or its physical embodiment (manifestation / item)
- ➔ The reason FRBR exists: we can be more efficient in creating catalogue records if we understand there are things we only need to describe once about it (e.g., who created it, what it’s about) and other things need more care in describing at item level (call number)
  - If we can reliably distinguish those different types of labour that go into describing something, we can build off each other’s work better

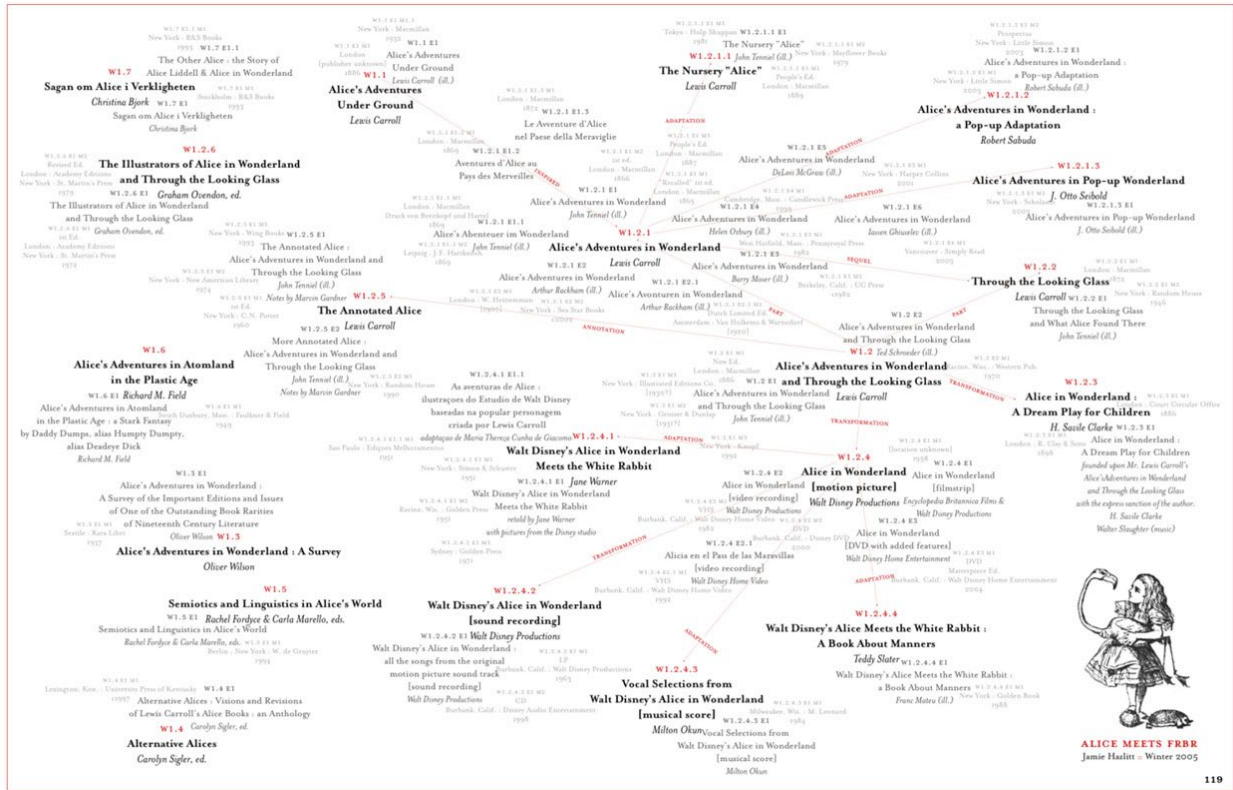
Examples of FRBR Group 1 Entities

Work	Berlioz <i>Symphonie fantastique</i>
Expression	1972 recording of a performance by the Chicago Symphony and Georg Solti
Manifestations	- Original Decca release in 1972 (London CS 6790) - Reissue on London Jubilee in 1985 (414 307-1)
Items	Individual copies of each manifestation at the Sibley Music Library

One of reasons it matters to be able to describe intellectual / creative work from the physical manifestation is to distinguish relationships between these things; e.g.,

- X is index to Y
- X is sequel to Y
- X is an analysis of Y
- X is a choreography of Y

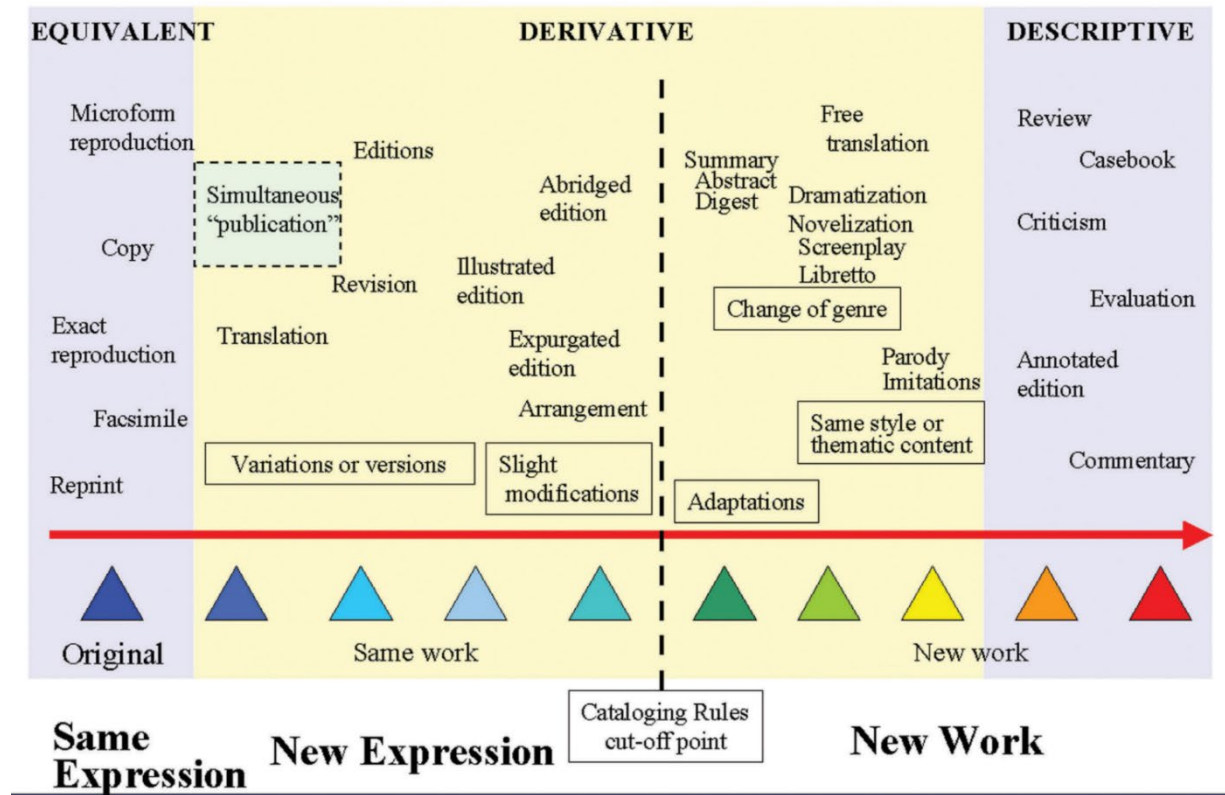
One example:



Once we have decided these things (what kind of information is info about the work, about the manifestation), then this needs to be embodied in library systems

- Where should we put creator information?
- Where should we put subject information?
- Is someone putting a HOLD on an item? Manifestation? Expression?
- Where should we put library user REVIEWS? (item? Manifestation? Expression? Work?)
- How else can we streamline cataloguing?

### 1.2.2 Families of Works



- Dotted line is cutoff: how is something just another expression of the same work, or should be treated as a distinct intellectual / creative work?

### 1.2.3 “Superwork” or “Work-Set”

Connects works with a common origin; there is a super-work of Alice in Wonderland things that are not all derivations of work, but have spawned intellectual / creative things of their own

Other examples: “Romeo and Juliet”, “The MCU”, “Alice in Wonderland”

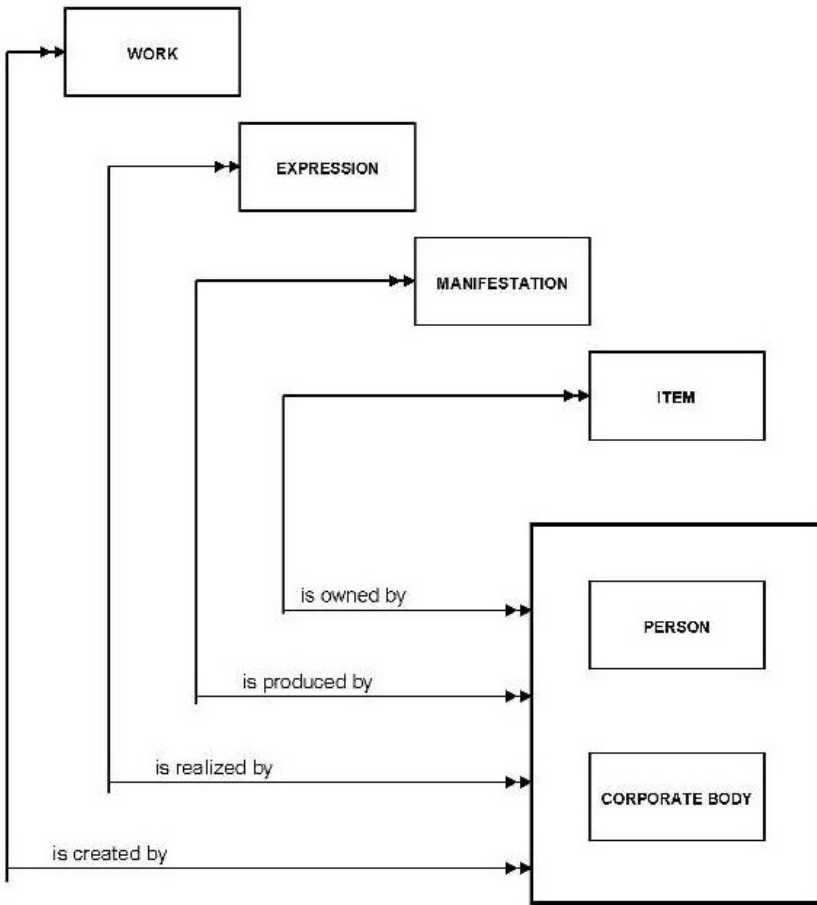
- Note: not yet figured out how to do this in library catalogue

### 1.2.4 Other FRBR Groups

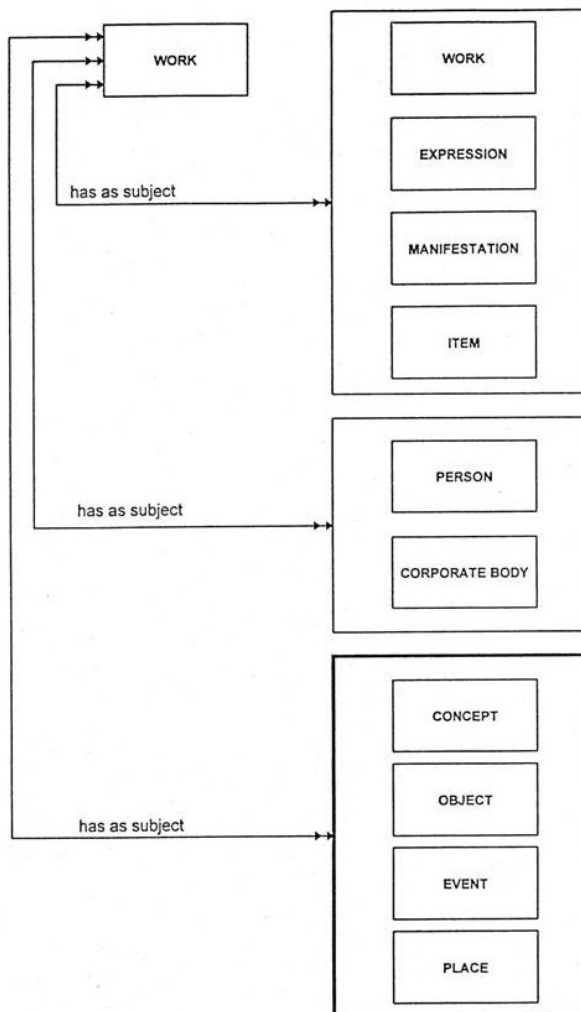
Group 1 entities are about intellectual / creative / physical embodiment of the thing; types of user needs are addressing are things like

- “Do you have Seamus Heaney’s translation of Beowulf?” (a request for an **expression**)
- “Do you have Stephen Hawking’s A Brief History of Time?” (a request for a **work**)

Group 2 entities are about responsibility relationships, agents (**persons, families, corporate bodies**) that can be related to physical work in different ways



Group 3 Entities: **concepts, objects, events, places**



### 1.2.5 Some thoughts / reminders

FRBR is very hard and remains a topic of debate, angst, and exasperation among professional libraries  
 Current cataloguing standard for libraries, RDA, is built upon FRBR as an underlying model for what gets described

## 2 Readings

### 2.1 Discipline of Organizing Section 5.2: An Overview of Resource Description

**Descriptor:** (in library profession) one of the terms in a carefully designed language that can be assigned to a resource to designate its properties, characteristics, meaning, or its relationship with other resources

- Also called: keywords, index terms, attributes, attribute values, elements, data elements, data values, vocabulary, labels, tags



“Descriptor” is a more inclusive / general concept than these two other well-established terms in information resources:

- **Bibliographic descriptions:** describe information resources and other entities of the bibliographic universe (e.g., works, editions, authors, subjects); typically manifests as a structured record in a standardized format (e.g., book catalogues)
- **Metadata:** (“data about data”) any kind of structured description for information resources; most useful in a metadata schema that defines the elements in the structured description
- **Tags:** extension of the concept of metadata; applying labels to content in order to describe and identify it (some problems with tagging: vocabulary problem, morphological inconsistencies, etc.)

**Resource Description Framework (RDF):** a standard model for making computer-processable statements about web resources (specifically for resources that have been given a URI [Uniform Resource Identifier])

- RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. (<https://www.w3.org/RDF/>)
- RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “**triple**”). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.

**Aggregation:** a set of information objects that, when considered together, compose another named information object (e.g., web page, web site, album)

- How to express an aggregation in the information architecture and technicalities?
- Boundaries of information objects in the digital realm can be fuzzy (e.g., is a web page one resource, or 3—HTML, CSS, and JPEG? What about a web site?)
- Some technical approaches to the aggregation problem:
  - **Kahn-Wilensky Digital Object Framework**
  - **Warwick Framework**
  - **METS**
  - **OAI-ORE**

The variety of resources that might be described means there are a variety of frameworks that people have proposed to think about resource descriptions. Some questions:

1. Should multiple resource descriptions be combined, or kept separate?
2. What syntax should be used to encode descriptions?

But most importantly: for who / what purpose are we describing? “Resource description is not an end in itself”