



FOUNDATIONS of DATA CURATION

Allen H. Renear, Cheryl A Thompson, Katrina S Fenlon, Myrna Morales



School of Information Sciences



University of Illinois at Urbana-Champaign



DATA CONCEPTS



4

THE WAY FORWARD: ROLES AND TYPES

Types and Roles

In the last video we saw how our FRBR-inspired model for representation encountered two problems

- First, it appeared that the two middle entity types should be collapsed into a single entity type

- Second, it appeared that the two middle entity types should be expanded into an indefinite number of entity types.

So apparently some considerable *refactoring* is required.

In this video we present the foundation for this refactoring, describing a distinction (*types* vs *roles*) that has become fundamental in ontological analysis.

The first problem: the collapse (again):

How about:

FRBR

Work

Expression

Manifestation

Item

Linguistic Representation

proposition

sentence

encoding

inscription

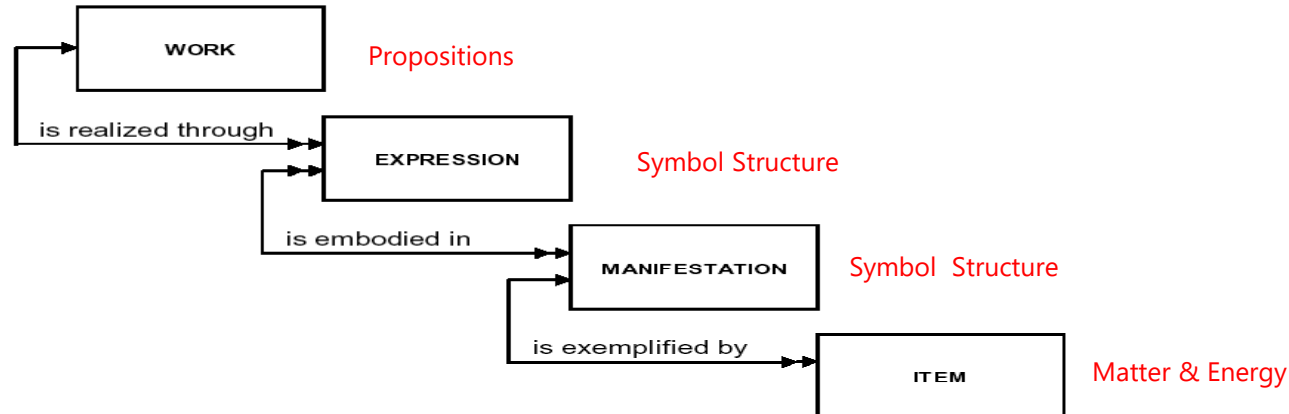
Entity Type

Proposition

Symbol Structure

Symbol Structure

Patterned Matter & Energy



The second problem: the proliferation (again):

	<i>Snow is white</i>	<i>a proposition</i>
can be expressed by	"Snow is white"	<i>a sentence</i>
can be encoded by:	S,n,o,w, ,i,s, ,w,h,i,t,e	<i>characters</i>
which can be encoded by:	<i>Snow is white.</i>	<i>glyphs</i>
which can be encoded by:	83, 110, 111...	<i>integers+*</i>
which can be encoded by:	53, 6E, 6F...	<i>numerals</i>
which can be encoded by	01010011 01101110 01101111...	<i>binary octets</i>
[But how many levels are there here, really? <i>There can be any number!</i>]		

The situation :

- 1) We have an indefinite number of symbolic encodings, not just one [or two]
- 2) the first level seems to be similar to a FRBR expression
- 3) the rest seem to be either encoding an expression, or encoding an encoding (!)

Role and types, intuitively

We can distinguish two sorts of properties,
those that are roles and those that are types

Very roughly*:

The property being a student is a **role**
because persons are students only in virtue of particular contingent circumstances,
(namely: being enrolled in a school).

The property being a person is a **type**
because persons are persons regardless of contingent circumstances

Other properties that are roles: parent, president, planet nearest the sun, etc.

Other properties that are types: number, color, physical object, copper, water, etc.

There's much more to this; see: Guarino & Welty (2000) A Formal Ontology of Properties.



Roles vs types

Rigidity can help distinguish roles and types.

Because types are rigid and roles are not.

Defining rigidity (in natural language)

A property is rigid if and only if

Nothing that has it could exist and fail to have it
(in the past, present, or the future or in any alternative circumstances)

examples: person, number, color, physical object, copper. .

A property is a role if and only if

Anything that has it could exist and fail to have it
(in the past, present, or future or in any alternative circumstances)

examples: student, president, parent, planet nearest the sun

Adapted from Guarino & Welty (2000) A Formal Ontology of Properties.

A little more elucidation

The property of being a person is rigid

Nothing that is a person could exist and fail to be a person

so, a physical object could not have existed and not been a physical object
(e.g., have been a color or a number instead)

The property of being a student is not rigid

Things that are students could exist and not be a student

so, a student may be a student now, but if their life had gone otherwise
they might not have been a student.

Or:

Once you were not a student, now you are a student, soon you will not be a student.

But everything that is a person has been a person since it existed, and will always be a person as long as it exists. (same for numbers, colors, physical objects, etc)

The distinction in modal logic

In modal logic rigidity is defined:

A property ϕ is *rigid* **=df** $\Box(\forall x)(\phi x \rightarrow \Box\phi x)$

Or, in the model theoretic semantics for modal logic:

A property ϕ is *rigid* =df

if ϕ is had by some x in some possible world,
then x has ϕ in every possible world in which x exists

Guarino & Welty (2000) A Formal Ontology of Properties.

Are the middle entity types roles or types -- FRBR

With respect to the two models we are trying to align we can ask

"Are the middle entity types roles or types?"

Let's start with FRBR.

A **expression** is a symbol structure that realizes a work

A **manifestation** is a symbol structure that embodies an expression

But:

Symbols have their meanings only as a result of contingent social convention

In different circumstances symbol structures mean different things

➔ So both expressions and manifestations would seem to be *roles*, not *types*.

Are the middle entity types roles or types -- language

Now lets consider our simple model of linguistic representation:

A **sentence** is a symbol structure that expresses a proposition

An **encoding** is a symbol structure that encodes a sentence

But, again:

Symbols have their meanings only as a result of contingent social convention

In different circumstances symbol structures mean different things

➔ So both sentences and encodings would seem to be *roles*, not *types*.

The middle entity types are *roles*, not *types*!!

Our model of a type/role relationship:

A **student** is a *person* enrolled in a school

So *being student* is a role that things of a particular type (persons) have in particular contingent circumstances.

This is parallel to:

A **sentence** is a *symbol structure* that expresses a proposition

An **encoding** is a *symbol structure* that encodes a sentence

red=types
green=roles

So we can collapse the two middle entity types

Both are symbol structures, the difference is a difference of role, not type.

But what about the proliferation problem?

	<i>Snow is white</i>	<i>a proposition</i>
can be expressed by	"Snow is white"	<i>a sentence</i>
can be encoded by:	S,n,o,w, ,i,s, ,w,h,i,t,e	<i>characters</i>
which can be encoded by:	<i>Snow is white.</i>	<i>glyphs</i>
which can be encoded by:	83, 110, 111...	<i>integers+*</i>
which can be encoded by:	53, 6E, 6F...	<i>numerals</i>
which can be encoded by	01010011 01101110 01101111...	<i>binary octets</i>
[But how many levels are there here, really? <i>There can be any number!</i>]		

The situation :

- 1) We have an indefinite number of symbolic encodings, not just one [or two]
- 2) the first level seems to be similar to a FRBR expression
- 3) the rest seem to be either encoding an expression, or encoding an encoding (!)

Now we have a simple solution

Some new definitions:

A **sentence** is a **symbol structure** that expresses a proposition

An **encoding** is a **symbol structure** that encodes a [**sentence** or **encoding**]

So a single entity type can, with a recursive relationships,
represent an indefinite number of encoding levels

How cool is that?

You'll see in the next video.

FOUNDATIONS OF DATA CURATION (IS531)

Allen H. Renear, Cheryl A Thompson, Katrina S Fenlon, Myrna Morales

School of Information Sciences

University of Illinois at Urbana-Champaign

Includes material adapted from work by Carole Palmer, Melissa Cragin,
David Dubin, Karen Wickett, Bertram Ludæscher, Ruth Duerr and Simone Sacchi.

Comments and corrections to: renear@illinois.edu.