



Mapping a Data Structure to the CIDOC Conceptual Reference Model

Martin Doerr

(ICS-FORTH, Crete, Greece)

Heraklion, Crete, April 2, 2002



What Means Mapping of One Schema to Another

- q Defining an (automated) transformation of each instance of schema 1 into an instance of schema 2 with the same meaning.
- q CRM Approach:
 - u Interpretation of schema 1 as semantic model (nodes and links),
 - u mapping each element of that to an equivalent CI DOC CRM path,
 - u such that each instance of an element of the semantic model 1 can be converted into a valid instance of the CI DOC CRM with the same meaning.
- q This is the **most simple theory**. Works for **good structures**



Interpreting a Schema as Semantic Model

1. Interpreting tables, columns as entities
 2. Interpreting records as entity instances
 3. Interpreting fieldnames as relationships and entities
 4. Interpreting field contents as entity instances
-
- q Each field is interpreted as entity-relationship-entity (e-r-e)
 - q The whole schema is decomposed into e-r-e's
 - q Each e-r-e is mapped individually to the CRM.



Interpreting a Schema as Semantic Model, Example

The field name stands for a relationship and the kind of contents

The field contents stand for an entity instance :

Object
1975-7309

has ID:

1975-7309

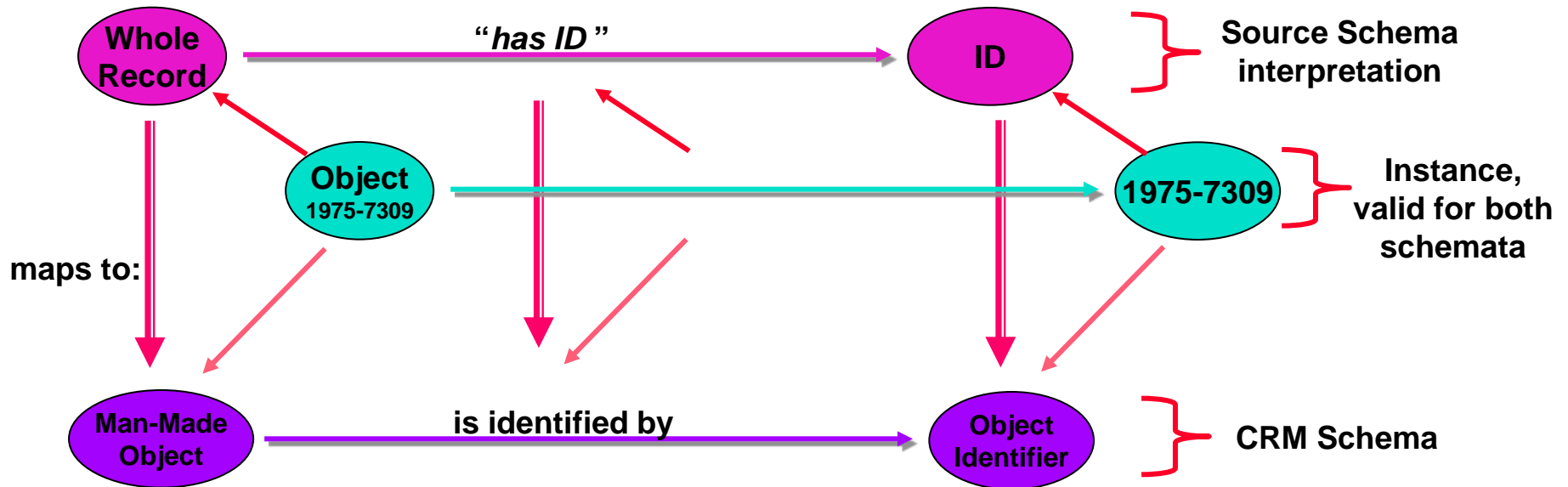
The whole record corresponds to one entity: It stands for one object which is not referred to

<i>ID</i>	1975-7309		
<i>Category</i>	NRM - Railway furniture		
<i>Description</i>	Armchair, Upholstered in blue moquette with curved, buttoned back & scroll arms. Wooden legs		
<i>Item name(s)</i>	armchairs (AAT Hierarchy: Furnishings)		
<i>Part</i>	<i>Aspect</i>	<i>Term</i>	<i>(AAT Hierarchy)</i>
overall	physical descriptor	upholstering	Processes & techniques
overall	material	moquette	Materials
overall	colour	blue	Color
legs	material	wood	Materials
back	physical descriptor	buttoning	Processes & techniques
back	shape	curved	Physical attributes
arms	shape	scrolled arms	Components

(data example from the Science Museum of London)



Mapping the First Element: Creating an Equivalent Proposition



Possible Mapping Annotation:

Whole Record = E22 Man-Made Object
ID = E42 Object identifier
Whole Record->ID = P47 is identified by

Possible CRM instance Annotation:

Object 1975-7309 (E22: Man-Made_Object)
is_identified_by **1975-7309** (E42 Object_Identifier)



Mapping the Interpreted Schema to the CRM

- q Each Entity-link-entity can be instantiated as self-explanatory, context independent proposition
- q The mapping allows to create sets of propositions equivalent to the meaning of each source document, but in terms of the CIDOC CRM.
- q As the CRM-compatible propositions are self-explanatory, they can be merged into huge knowledge pools and the document boundaries can be ignored.
- è **buzz words:** Data warehouses, Semantic Web



Interpreting a Schema: Advanced Stuff: Value Dependency

The first field name stands for a relationship and the kind of contents

The field contents stands for an entity instance :



Mapping condition:

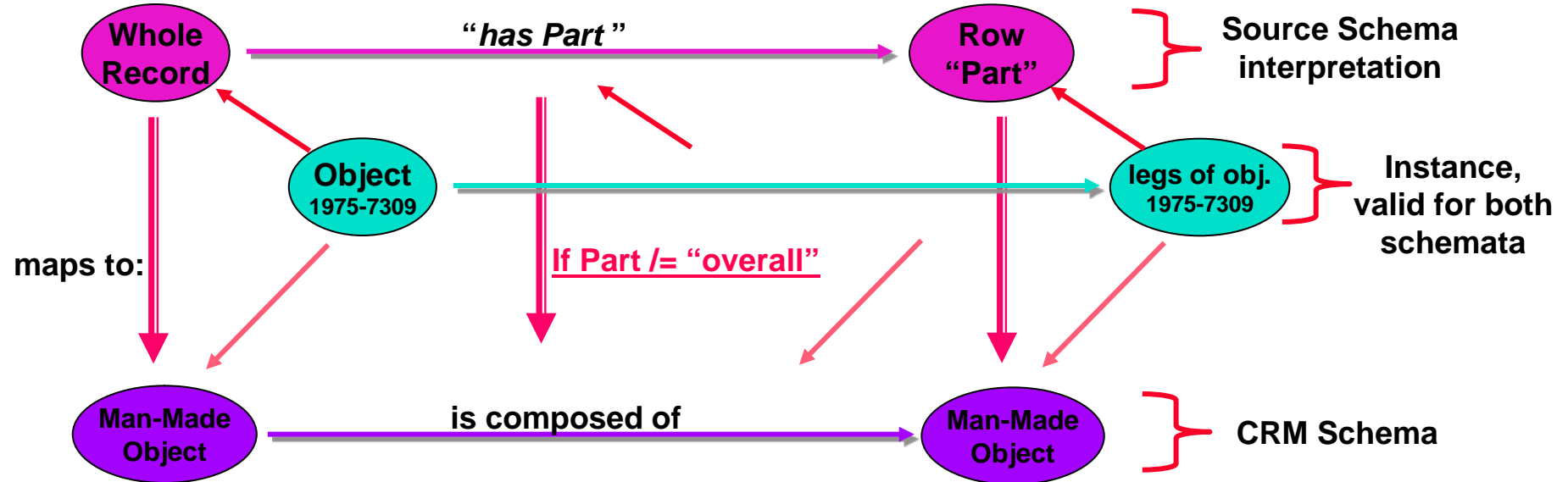
If part = overall,
it stands for the whole

The whole row corresponds
to one entity:
It stands for one part

<i>ID</i>	1975-7309		
<i>Category</i>	NRM - Railway furniture		
<i>Description</i>	Armchair, Upholstered in blue moquette with curved, buttoned back & scroll arms. Wooden legs		
<i>Item name(s)</i>	armchairs (AAT Hierarchy: Furnishings)		
<i>Part</i>	<i>Aspect</i>	<i>Term</i>	<i>(AAT Hierarchy)</i>
overall	physical descriptor	upholstering	Processes & techniques
overall	material	moquette	Materials
overall	colour	blue	Color
legs	material	wood	Materials
back	physical descriptor	buttoning	Processes & techniques
back	shape	curved	Physical attributes
arms	shape	scrolled arms	Components



Mapping under condition: Creating an equivalent statement



Possible Mapping Annotation:

Whole Record = E22 Man-Made Object
 Row "Part" = E22 Man-Made Object
 If (in Row "Part", Part != "overall") then
 Whole Record-> Row "Part" = P46 is composed of

Possible CRM instance Annotation:

Object 1975-7309 (E22: Man-Made_Object)
 is_composed_of **legs of 1975-7309** (E22: Man-Made_Object)



Interpreting a Schema: Advanced Stuff: Values as Properties

The field "Aspect" contents
state a relationship

The field contents stands for
an entity instance :



Value based mapping

If part = overall,
AND
Aspect = material

<i>ID</i>	1975-7309		
<i>Category</i>	NRM - Railway furniture		
<i>Description</i>	Armchair, Upholstered in blue moquette with curved, buttoned back & scroll arms. Wooden legs		
<i>Item name(s)</i>	armchairs (AAT Hierarchy: Furnishings)		
<i>Part</i>	<i>Aspect</i>	<i>Term</i>	<i>(AAT Hierarchy)</i>
overall	physical descriptor	upholstering	Processes & techniques
overall	material	moquette	Materials
overall	colour	blue	Color
legs	material	wood	Materials
back	physical descriptor	buttoning	Processes & techniques
back	shape	curved	Physical attributes
arms	shape	scrolled arms	Components



Interpreting a Schema: Advanced Stuff: Mapping to Paths

The field "Aspect" contents
state a relationship

The field contents stands for
an entity instance :

Object
1975-7309

has physical descriptor:

upholstering

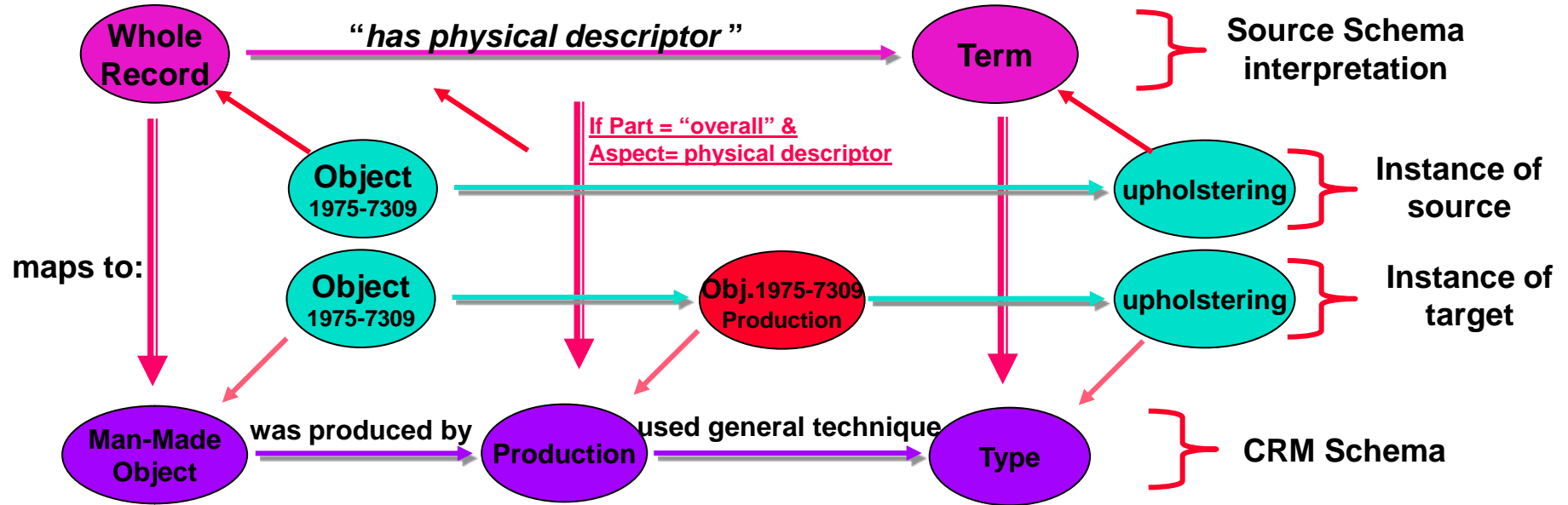
Value based mapping

If part = overall,
AND
Aspect = physical descriptor

<i>ID</i>	1975-7309		
<i>Category</i>	NRM - Railway furniture		
<i>Description</i>	Armchair, Upholstered in blue moquette with curved, buttoned back & scroll arms. Wooden legs		
<i>Item name(s)</i>	armchairs (AAT Hierarchy: Furnishings)		
<i>Part</i>	<i>Aspect</i>	<i>Term</i>	<i>(AAT Hierarchy)</i>
overall	physical descriptor	upholstering	Processes & techniques
overall	material	moquette	Materials
overall	colour	blue	Color
legs	material	wood	Materials
back	physical descriptor	buttoning	Processes & techniques
back	shape	curved	Physical attributes
arms	shape	scrolled arms	Components



Mapping to Paths: Introducing an intermediate node



Possible Mapping Annotation:

Whole Record	=	E22 Man-Made Object
Term	=	E55 Type
If Part = "overall" & Aspect= physical descriptor		
Whole Record-> Term	=	P108 was produced by – E12 Production - P32 used general technique

Possible CRM instance Annotation:

Object 1975-7309 (E22: Man-Made_Object)
was_produced_by **Obj. 1975-7309 Production** (E12: Production)
used general technique **upholstering** (E55 Type)



Interpreting a Schema: Advanced Stuff: Nested Structures

The contents of field "Aspect"
state a relationship

The field contents stands for
an entity instance :



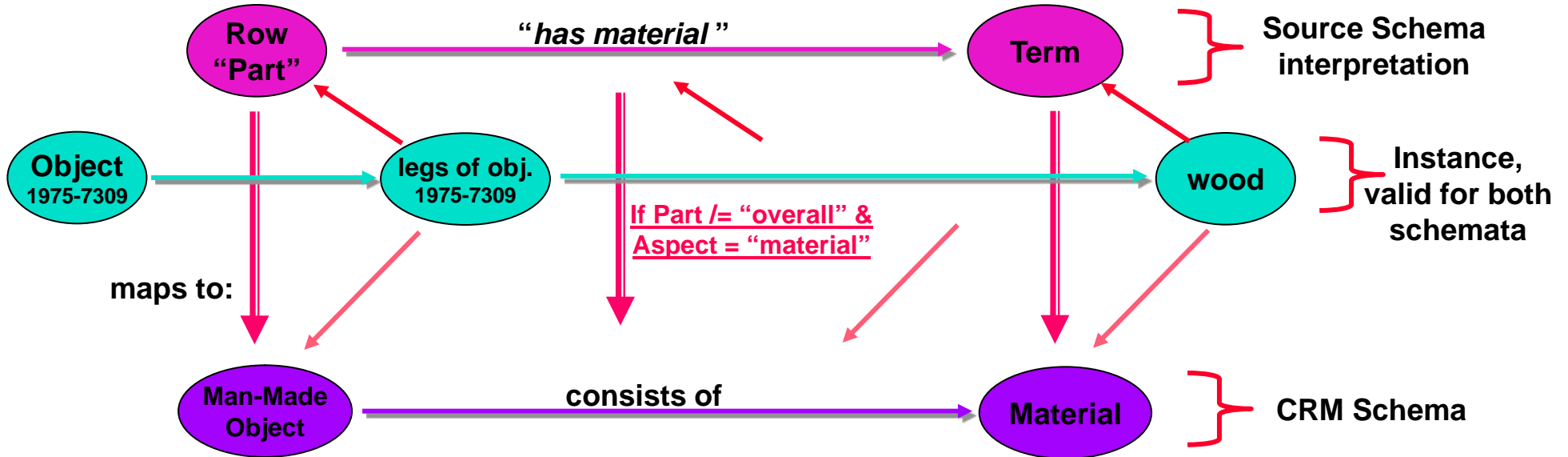
Value based mapping
If Aspect = material

The whole row corresponds
to one entity:
If part != overall
it stands for one part

<i>ID</i>	1975-7309		
<i>Category</i>	NRM - Railway furniture		
<i>Description</i>	Armchair, Upholstered in blue moquette with curved, buttoned back & scroll arms. Wooden legs		
<i>Item name(s)</i>	armchairs (AAT Hierarchy: Furnishings)		
<i>Part</i>	<i>Aspect</i>	<i>Term</i>	<i>(AAT Hierarchy)</i>
overall	physical descriptor	upholstering	Processes & techniques
overall	material	moquette	Materials
overall	colour	blue	Color
legs	material	wood	Materials
back	physical descriptor	buttoning	Processes & techniques
back	shape	curved	Physical attributes
arms	shape	scrolled arms	Components



Mapping Nested Structures : Continuing on a Range Entity



Possible Mapping Annotation:

Row “Part”	=	E22 Man-Made Object
If Aspect= “material”		
Term	=	E57 Material
Row “Part” -> Term	=	P45 consists of

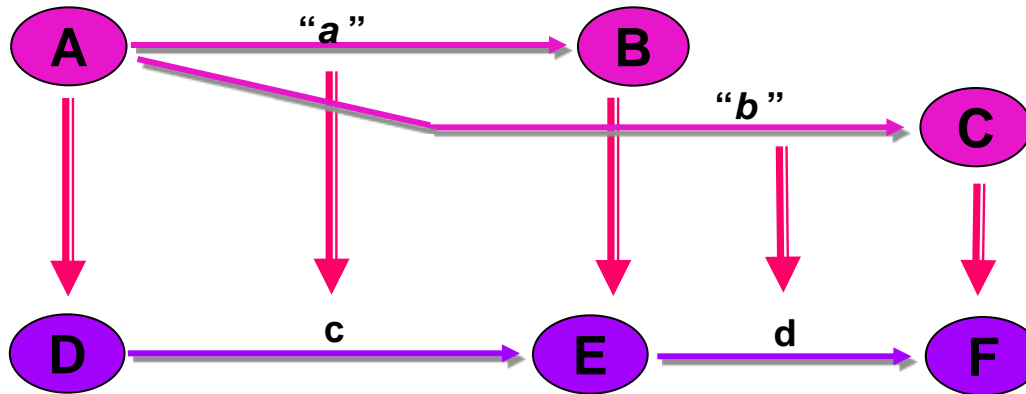
Possible CRM instance Annotation:

Object 1975-7309 (E22: Man-Made_Object)
is_composed_of legs of 1975-7309 (E22: Man-Made_Object)
consists_of wood (E57 Material)



Other Forms of Maps: Cases of Heterogeneity

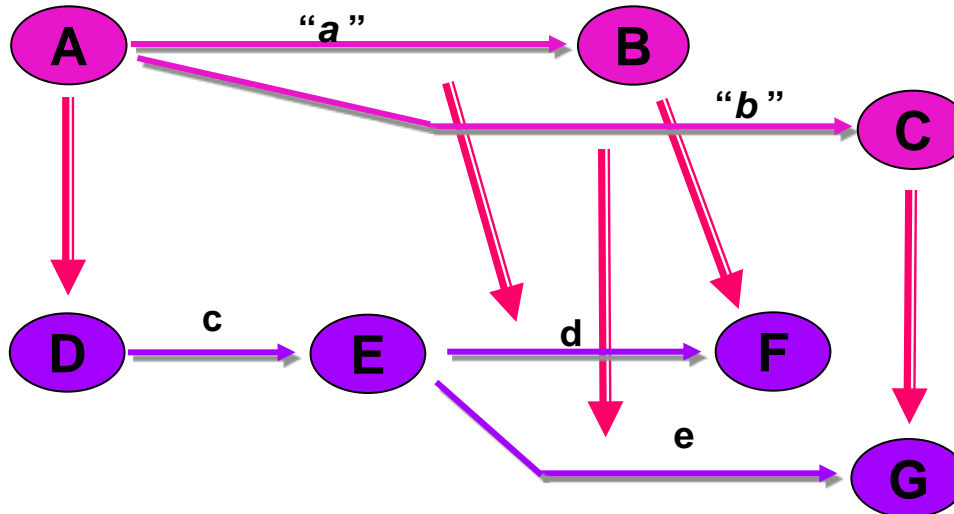
Parallel
to nested:



Source Schema
interpretation

CRM Schema

Parallel
to intermediate-
parallel:
(frequent with
events!)



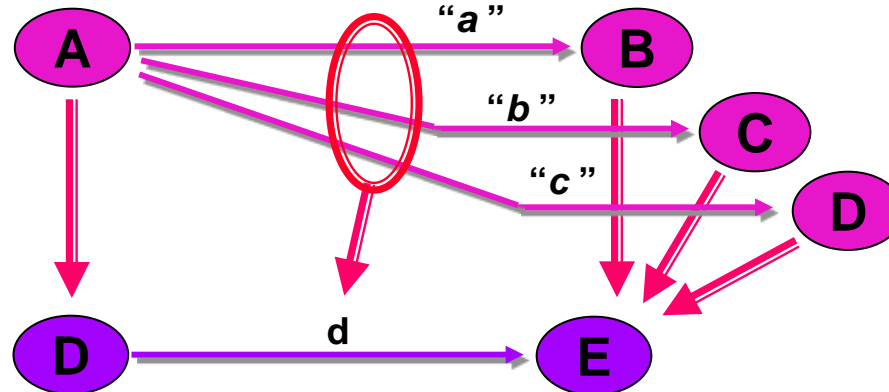
Source Schema
interpretation

CRM Schema



Other Mapping Forms: Cases of Heterogeneity

Compound contraction:
(frequent with addresses, species names etc!)



Source Schema interpretation
B,C,D are parts
of an identifier
for one real-life thing

CRM Schema



Mapping to the CRM: Conclusions

- q Mapping to the CRM can serve just as guide for good-practice data structures.
- q It can be used to create a Semantic Web of cultural knowledge.
- q It can be used to preserve data in a neutral form.
- q Even though mapping can become weird, good data structures transform easily, and there are commercial tools.
- q No tool can guess all the experts intention in a data structure: **Domain experts must assist the mapping.**