EFG
european film gateway

Exploring semantic relationships in filmographic data

May, 2009

Detlev Balzer
for Deutsches Filminstitut (DIF)

EFG
european film gateway

Isn't there enough semantics in conventional databases?

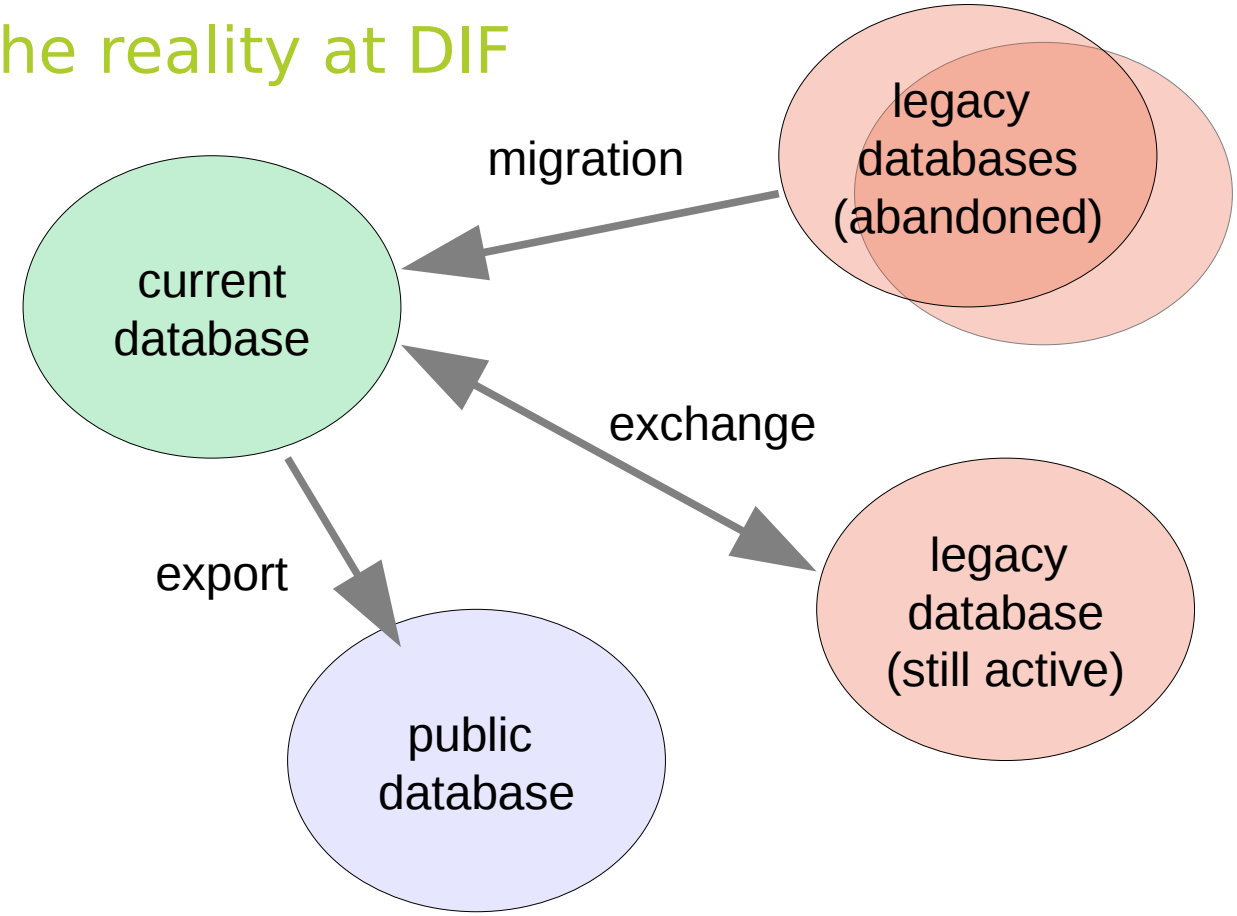
We have:

- Entities a.k.a. tables
- Attributes a.k.a. columns
- Relationships (one-to-many, many-to-many)
- Restrictions (e.g. referential integrity)

All very good as long as we can define the meaning of everything at design time

EFG
european film gateway

The reality at DIF



Each is a conceptual universe of its own

Implicit semantics

		first screening
...	...	15 July 1967

- in which country?
  - may be inferred as long as this is not a multinational co-production
- what kind of screening?
  - may be inferred as long as all dates refer to public theatrical screenings

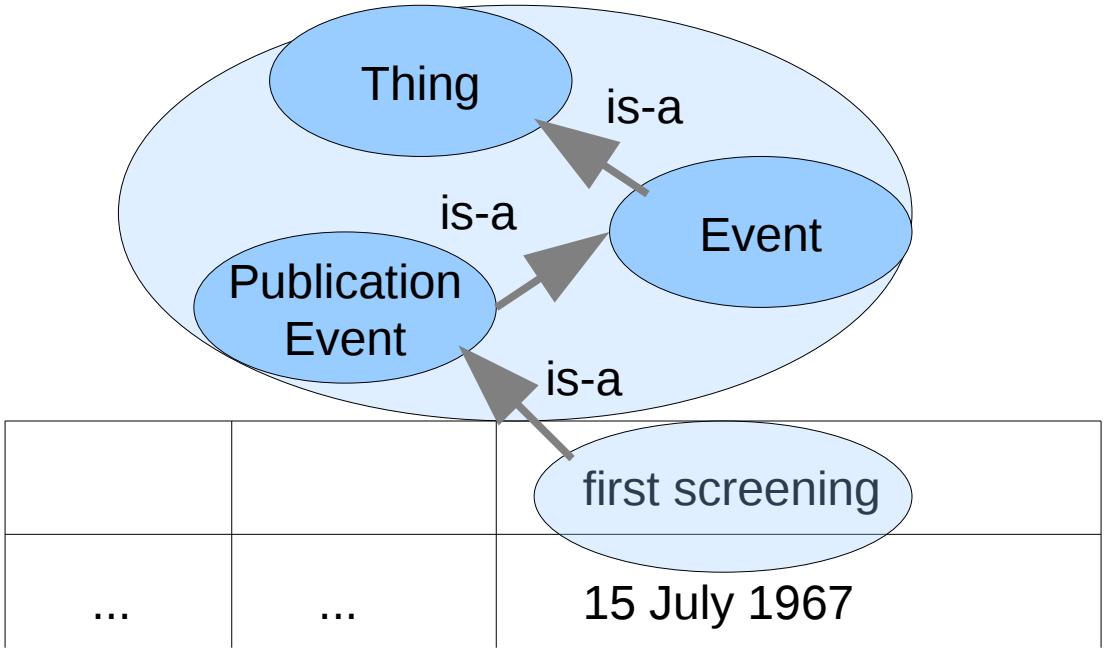
EFG
european film gateway

There is often more to tell ...
... but no sensible place to put it

		first screening	notes
...	...	15 July 1967	first screening in France: 09 Jul 1967; re-released in Germany, May 1975

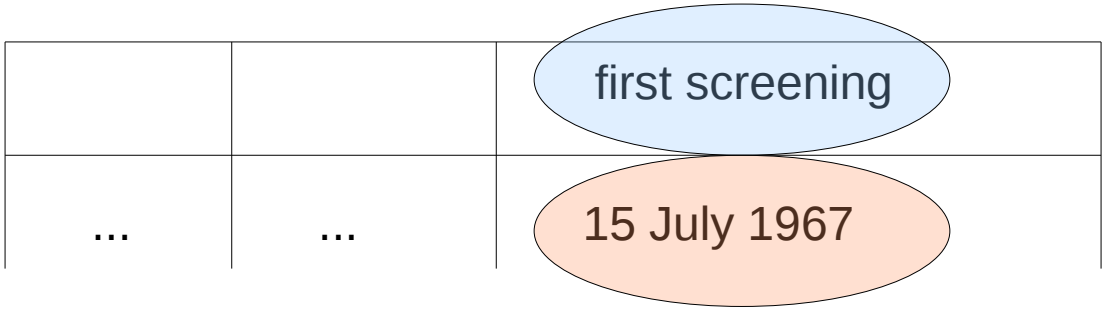
- Modifying a database model is expensive
- Ad-hoc solution is often „just add another column“

Adding semantics: The first step



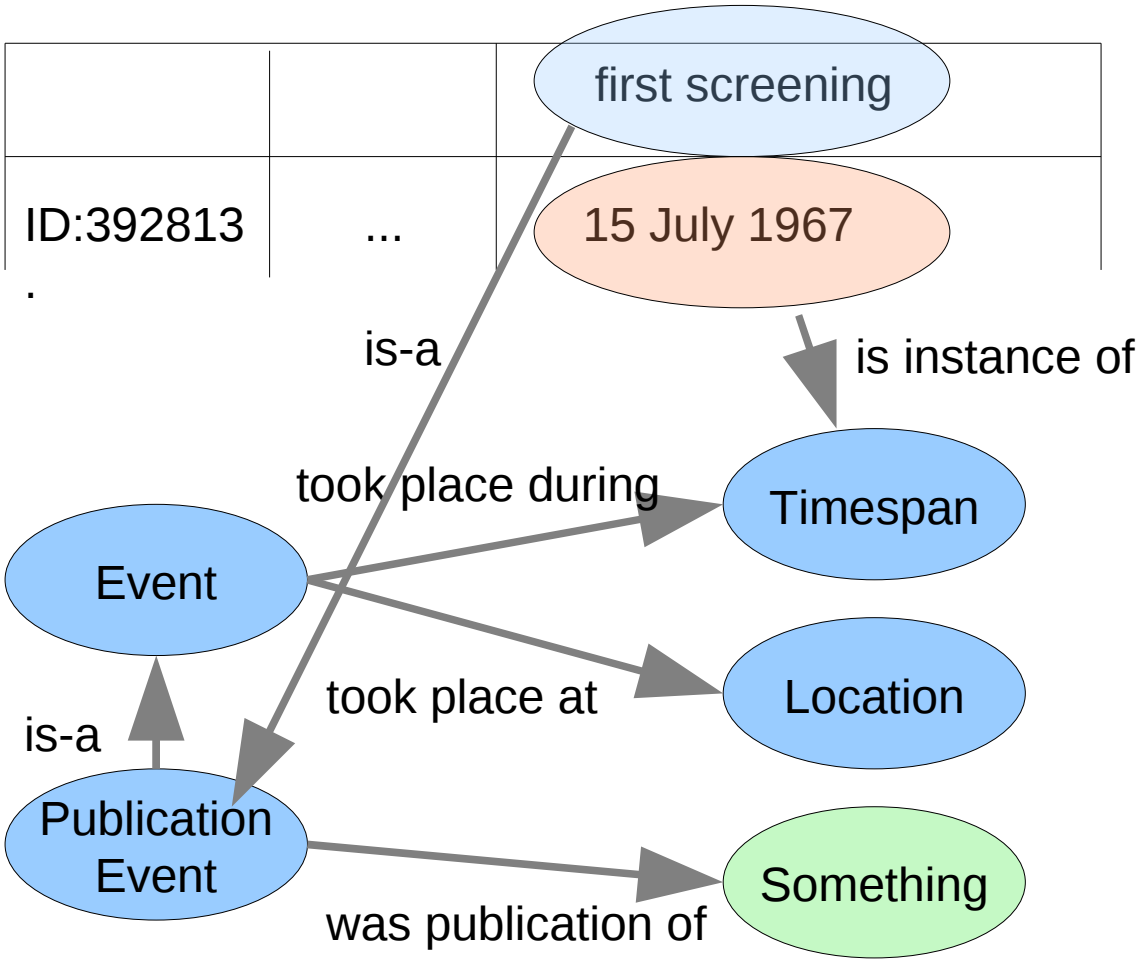
- Reference models help to explain what we mean
- We can explain it to us, and to others

How do these relate to each other?

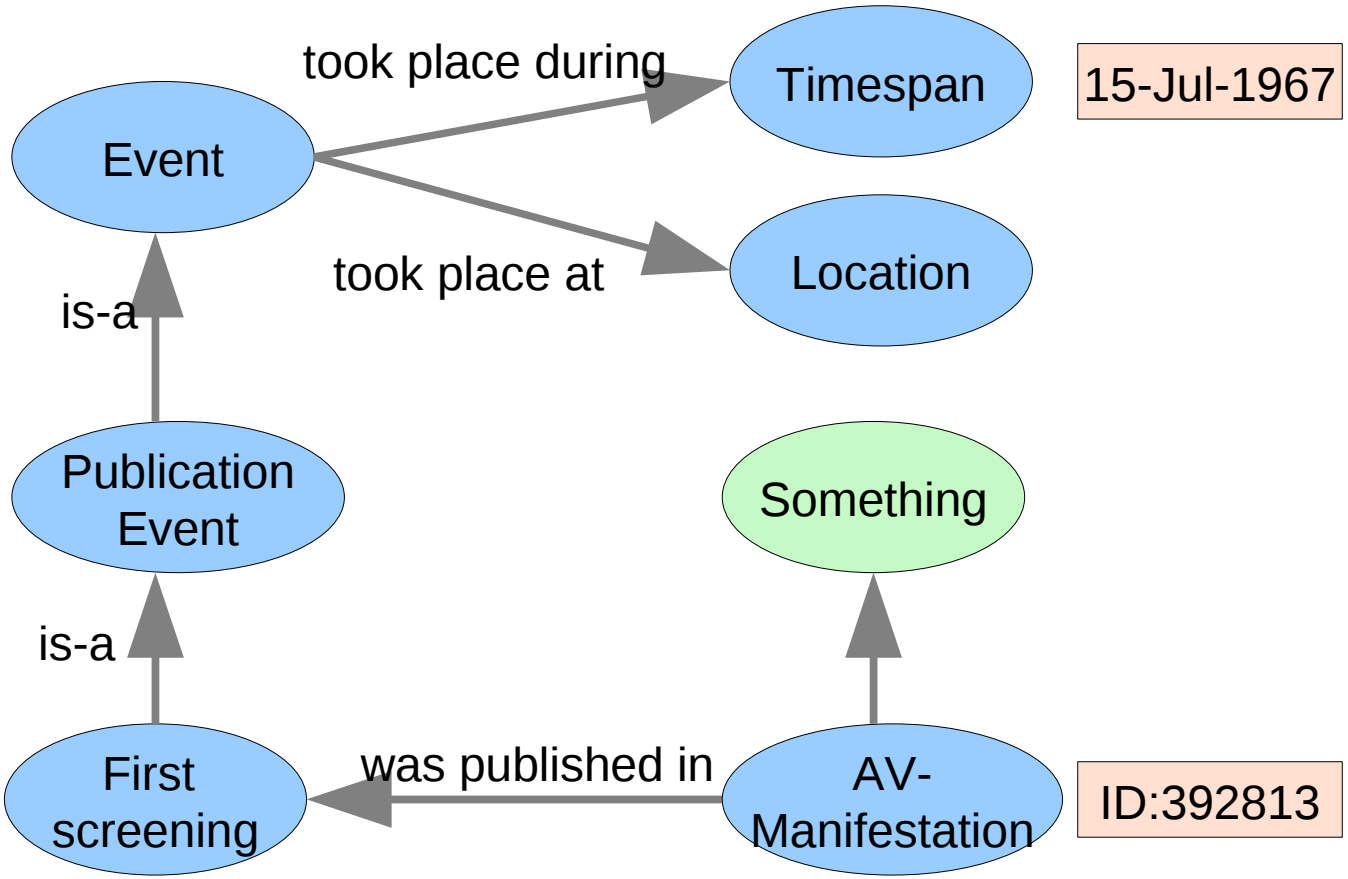


- In what way can „15 July 1967“ be a property value for „first screening“?
- What follows from „first screening“ being an event?

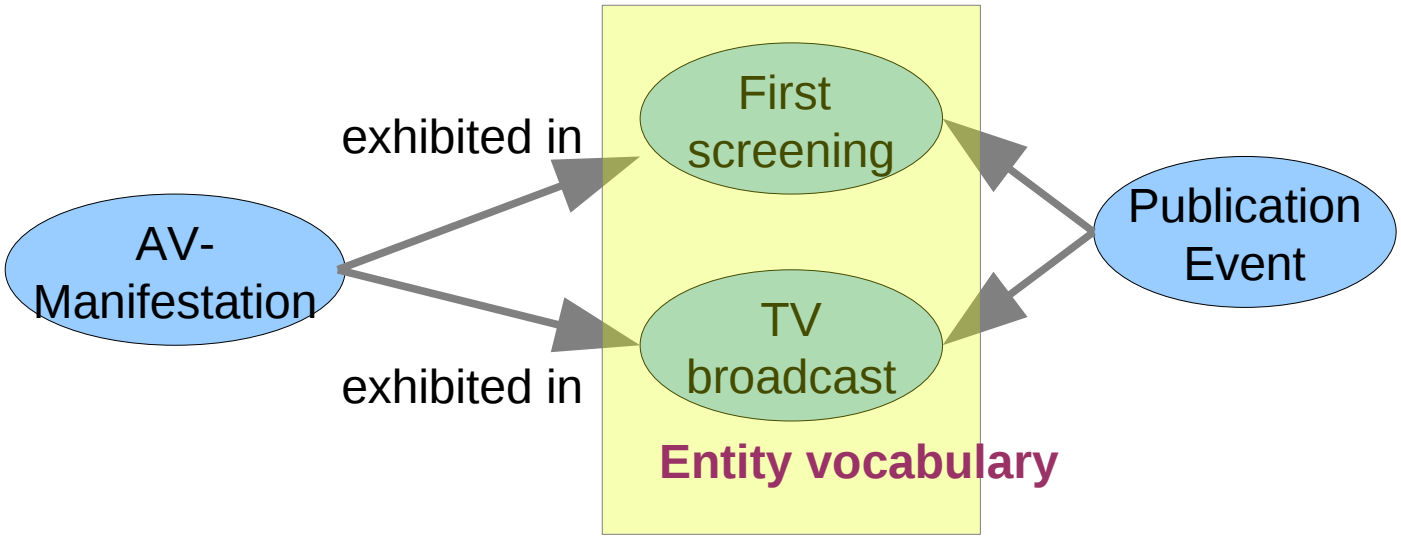
Reference model to the rescue, again!

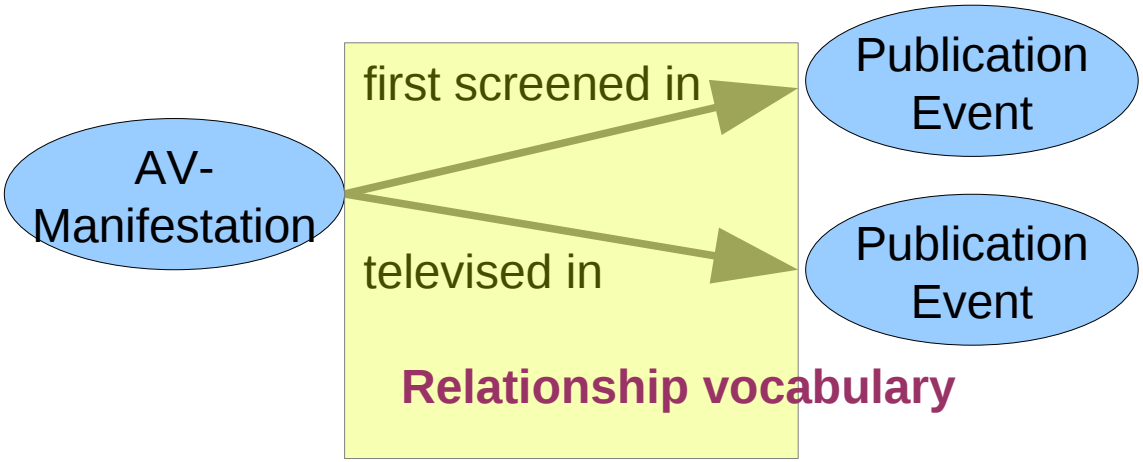


Re-arranging our interpretation



Entities or relationships?





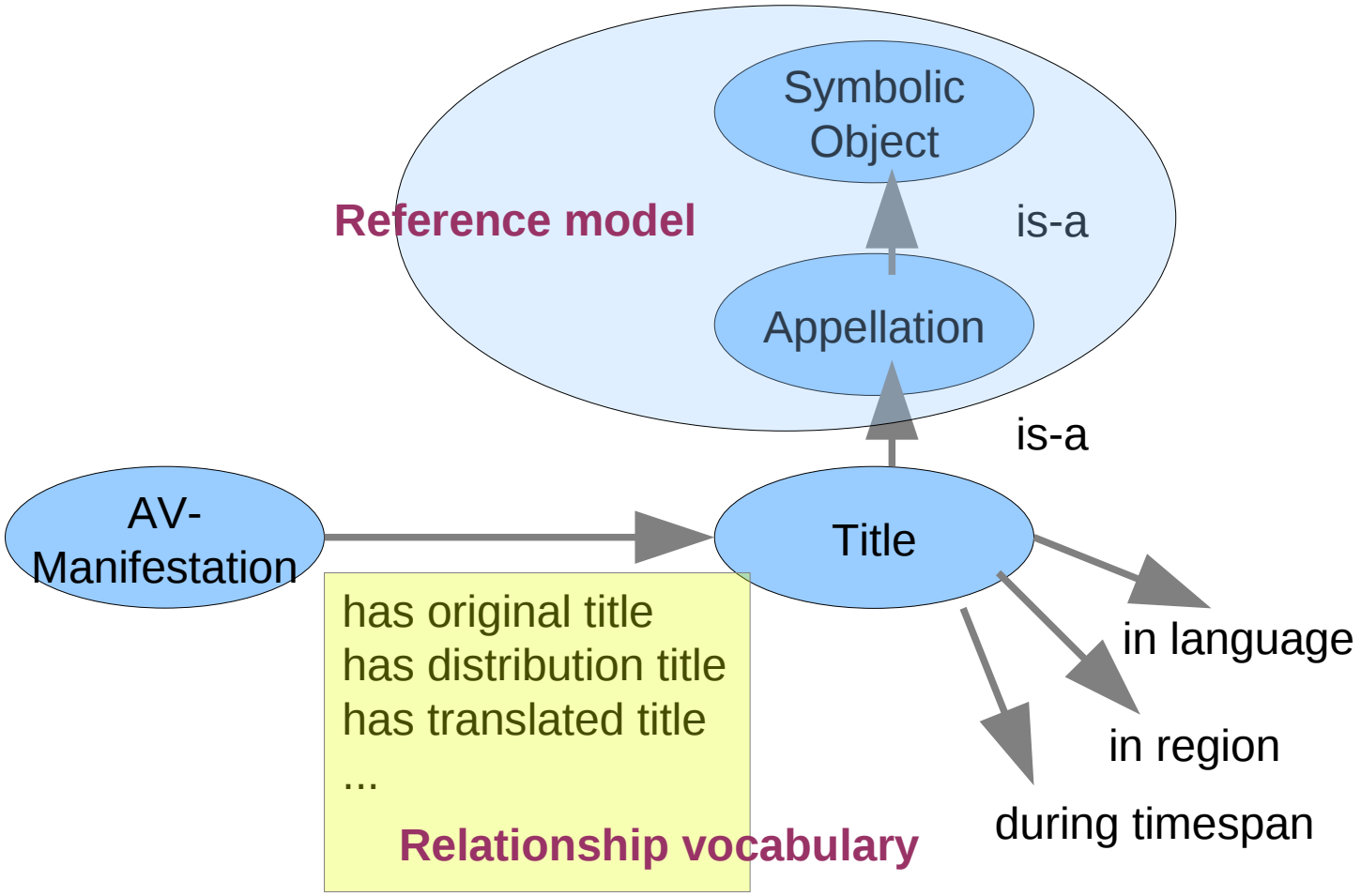
A case for relationship vocabularies

(From a legacy database input form):

Original Title	
Translated Title	
Distribution Title	

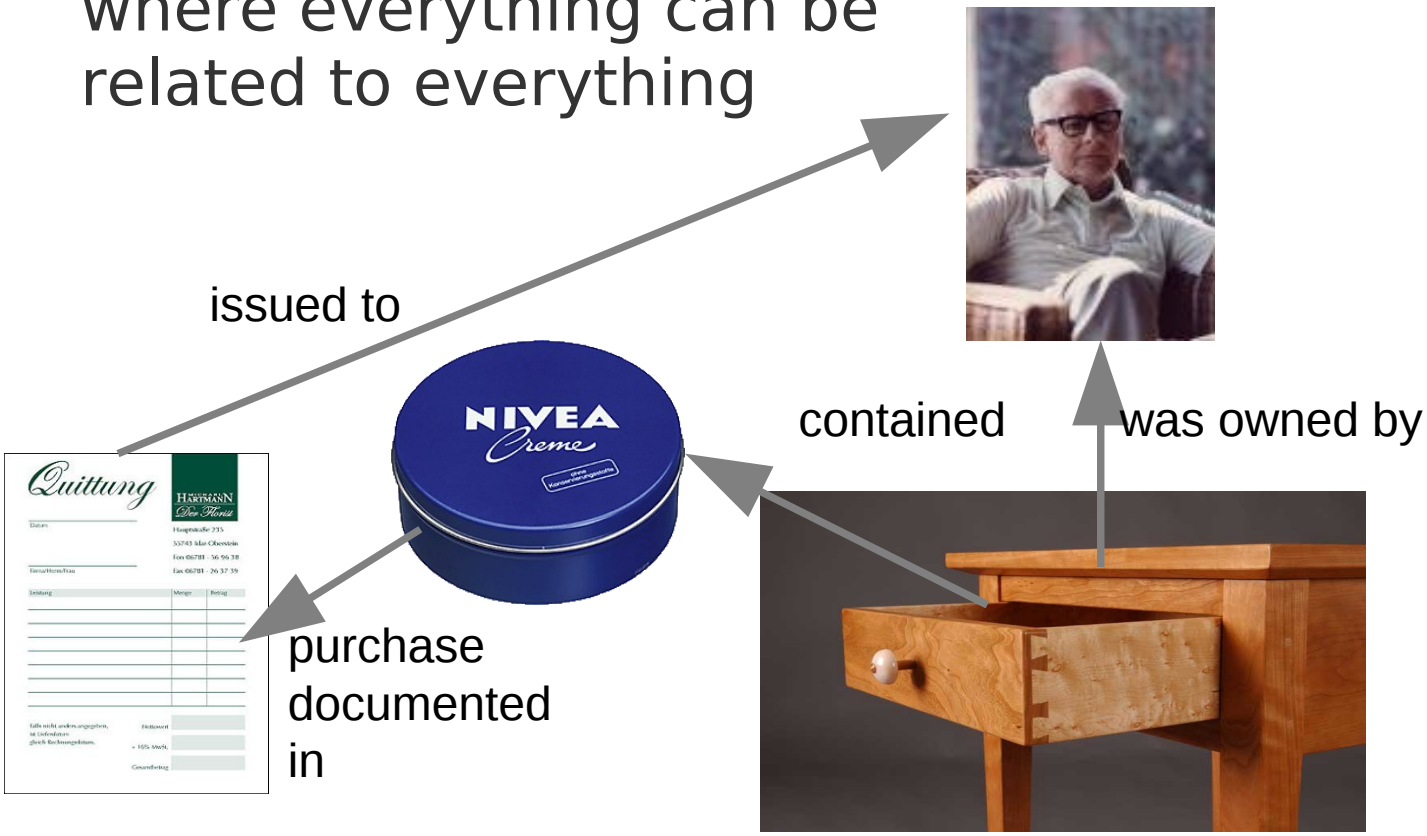
- What if there are several original titles (e.g. in a co-production)?
- Into what language is the second title translated?
- What country does the third title refer to?

Putting titles into context

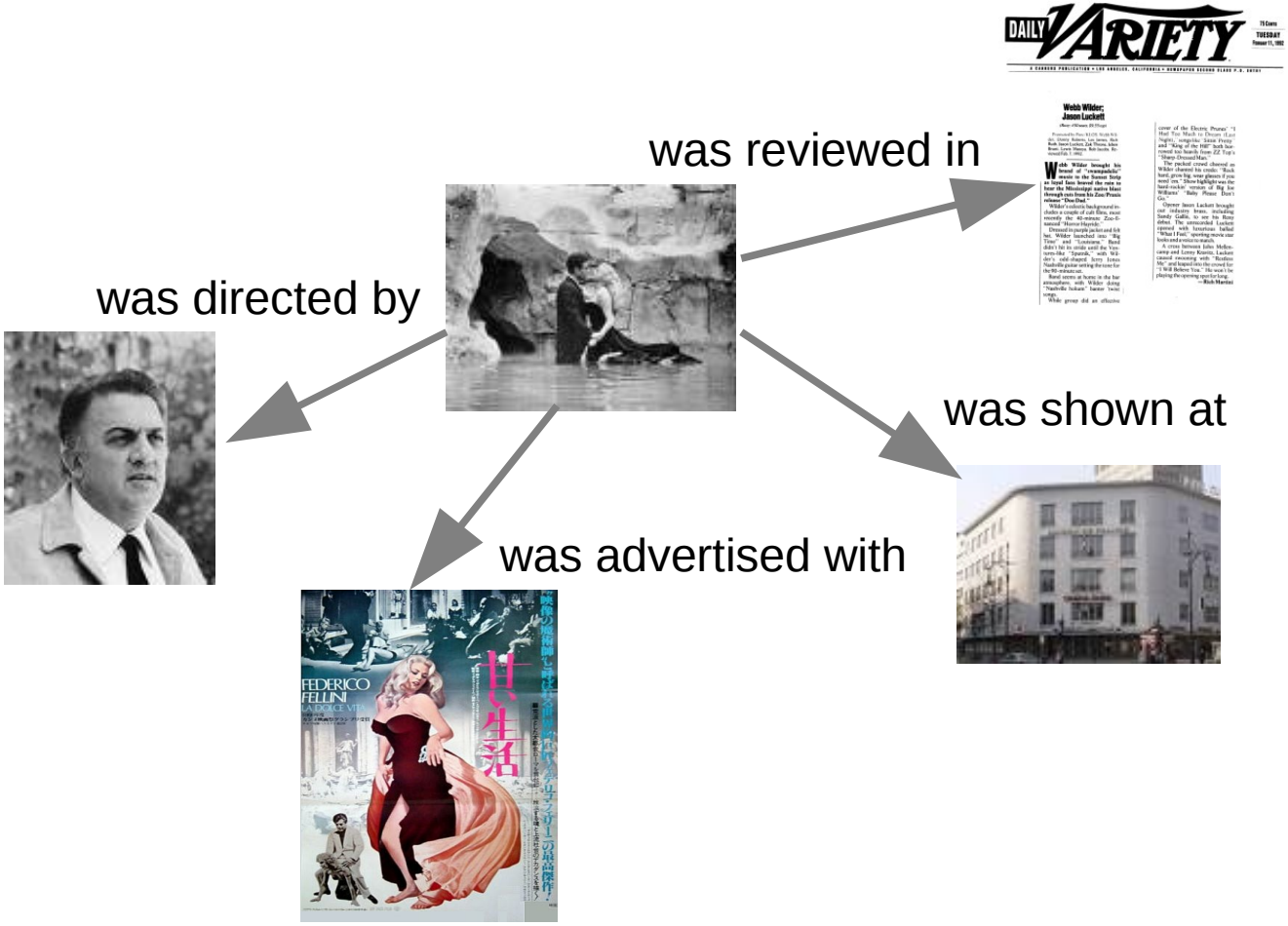


Relationship vocabularies in practice

- Museums have pioneered databases where everything can be related to everything

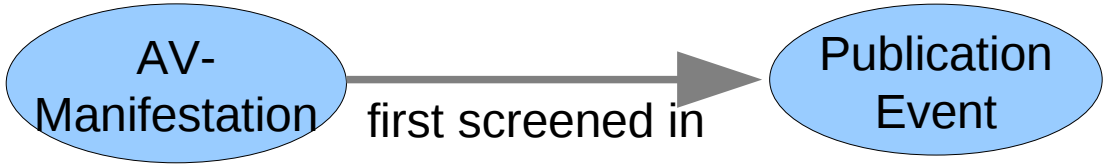


Filmography is about context, too



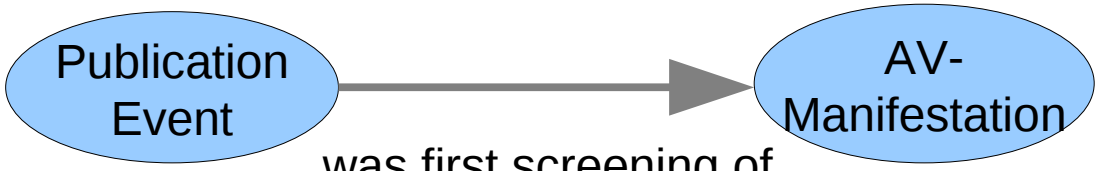
EFG
european film gateway

Expressing a relationship ...



Subject **Predicate** **Object**

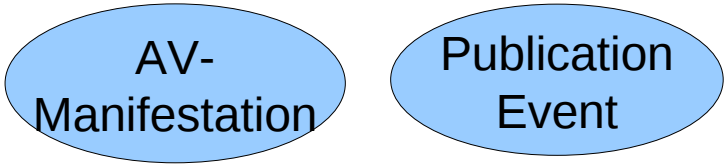
... and its inverse form



Subject **Predicate** **Object**

Defining relationships

first screened in



Property	**Domain**	**Range**
defines: **Predicate**	defines: **Subject**	defines: **Object**

- RDFS and OWL are the most common formal languages used for such definitions
- They are also useful for representing reference models

EFG
european film gateway

Does this work in practice?

- Yes, it does.
- DIF has introduced the concept in 2003 to avoid repeated changes to the data model.
- The DIF database now contains almost a million subject-predicate-object triples.
- To date, no modifications to the data model have become necessary.

EFG
european film gateway

Is this just database technology?

- No, it isn't.
- It can be used as a data modelling principle, but there is much more to it.
- Some call it the foundation of the „Semantic Web“.

EFG
european film gateway

From local database records to distributed resources

A library

was reviewed in

A film archive

was directed by

A biography collection

was shown at

A geographical information system

was advertised with

An art museum

EFG
european film gateway

Lots of hard work ahead ...

- Consolidate everything that can be named in different ways into shared vocabularies,
- and keep them updatable.
- Assign these vocabularies to the relevant parts of the EFG data model.

Example: Identical statements in different forms

An element value:

<AspectRatio ID="157">**1.33**</AspectRatio>

<Aspect>**4:3**</Aspect>

<Ratio>**volbeeld**</Ratio>

A relationship name:

430 $n **Originální název**

<TitleType>**Original Title**</TitleType>

<TitelTyp>**Originaltitel**</TitelTyp>

Example: Identifying relationship names in various forms

as XML attribute value:

<role xl:href="/Code/key/ROLE//KLI" xl:title="**Klipp**"/>

as MARC subfield value

430 $n **Originální název**

as XML element name:

<**DIRECTOR**>...</DIRECTOR>

EFG
european film gateway

Aligning EFG vocabularies with external vocabularies

- Candidates:
  - LoC Moving Image Genre-Form Guide
  - FIAF Glossary
  - EBU P/META Concept Schemes
  - MPEG-7 Concept Schemes
  - SMPTE Metadata Registry
  - LoC Thesaurus of Graphic Materials
  - … and various others