

SWIB 2012: Workshop on Metadata Provenance

Part 3: Expressing Provenance using RDF

Agenda

Dublin Core

PROV Ontology

Comparison

Preliminaries

Several ontologies modelling provenance exist

Open Provenance Model (OPM)

PROV

Provenance Vocabulary

Provenir

...

--> Presentation limited to a simple and a more complex approach

Dublin Core

Vocabularies

Dublin Core Metadata Initiative (DCMI)

Element set

- 15 basic terms

- No defined ranges (--> arbitrary values possible)

Terms

- 55 granular terms (properties)

- Well defined ranges

Example

Namespace

Element set --> dc:

Terms --> dcterms: or dct:

```
ex:doc1 dct:title "A mapping from Dublin Core..." .  
ex:doc1 dct:creator ex:kai .  
ex:doc1 dct:created "2012-02-28" .  
ex:doc1 dct:publisher ex:w3c .  
ex:doc1 dct:issued "2012-02-29" .  
ex:doc1 dct:subject ex:dublincore .  
ex:doc1 dct:replaces ex:doc2 .  
ex:doc1 dct:format "HTML" .
```

Distinction

Some terms contain only information about the resource itself

But not how or when it was produced

---> Descriptive Terms

Some terms also contain information on the creation or derivation of the resource

--> Provenance Terms

Who?

Terms

Contributor

Creator

Publisher

RightsHolder

Range is dct:Agent

a resource that acts or has the power to act

Clearly influencing creation of a resource

RightsHolder is ownership --> provenance in works of art

When?

Terms

Available

Created

Date

DateAccepted

DateCopyrighted

DateSubmitted

Issued

Modified

Valid

When?

Ranges

- Date range

 - Available, valid

- Single date

 - All others

Dates are basic provenance information

- Availability and validity often inherent to the resource

- But: provenance related, if active change

How?

Terms

IsVersionOf, hasVersion

IsFormatOf , hasFormat

References, isReferencedBy

Replaces, isReplacedBy

Source

HasPart, isPartOf

accrualMethod

How?

Information on Derivation and Replacement

Information on relations to other resources

Information on processes involved in creation

dcterms:provenance

Definition

“statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.”

--> “classic” provenance of works of art

Summary

More than half of the DC terms deal with provenance related information

Who?

When?

How?

What?

Missing information

Where?

Why?

(only if replacement)

PROV Ontology

Introduction

W3C Provenance Working Group

“This ontology specification provides the foundation to implement provenance applications in different domains that can represent, exchange, and integrate provenance information generated in different systems and under different contexts.”

Not finalized

Final call working draft

Basic constructs

Entities

Resources or “things” to describe

Actions

Creation or changes of entities

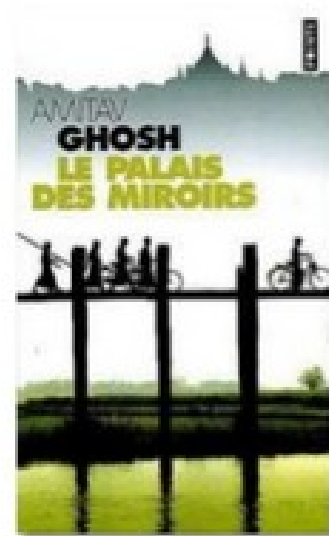
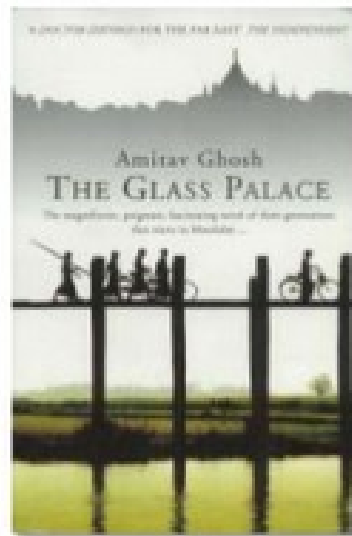
Agents

Those responsible for actions

Example

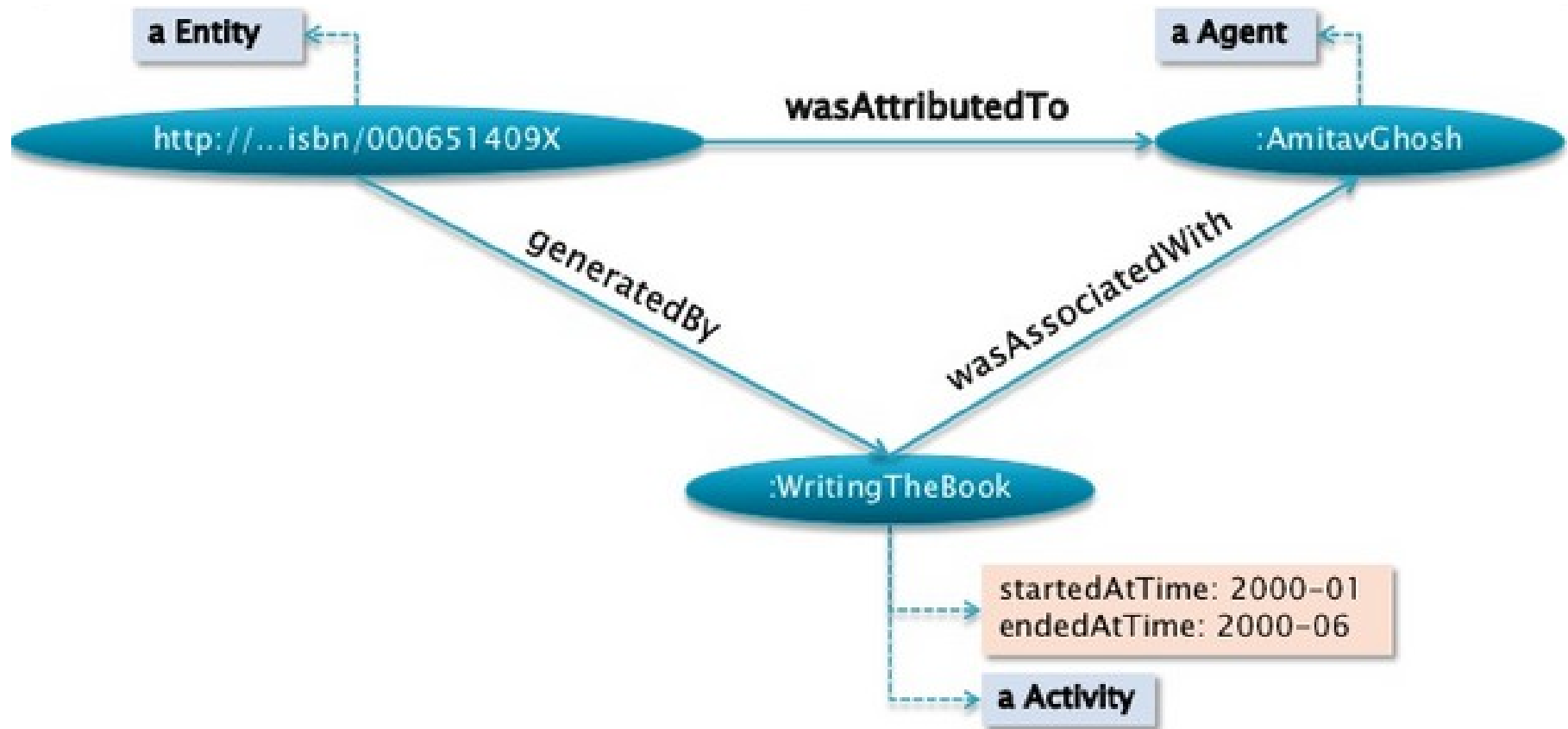
We have data on two books

- “The Glass Palace”, written by Amitav Ghosh
- “Le palais des miroirs”, the French translation, done by Christianne Besse, of the book of Amitav Ghosh
- we want to describe some very basic facts on the provenance of these



Source: Herman (2012)

Example



Source: Herman (2012)

Advantages

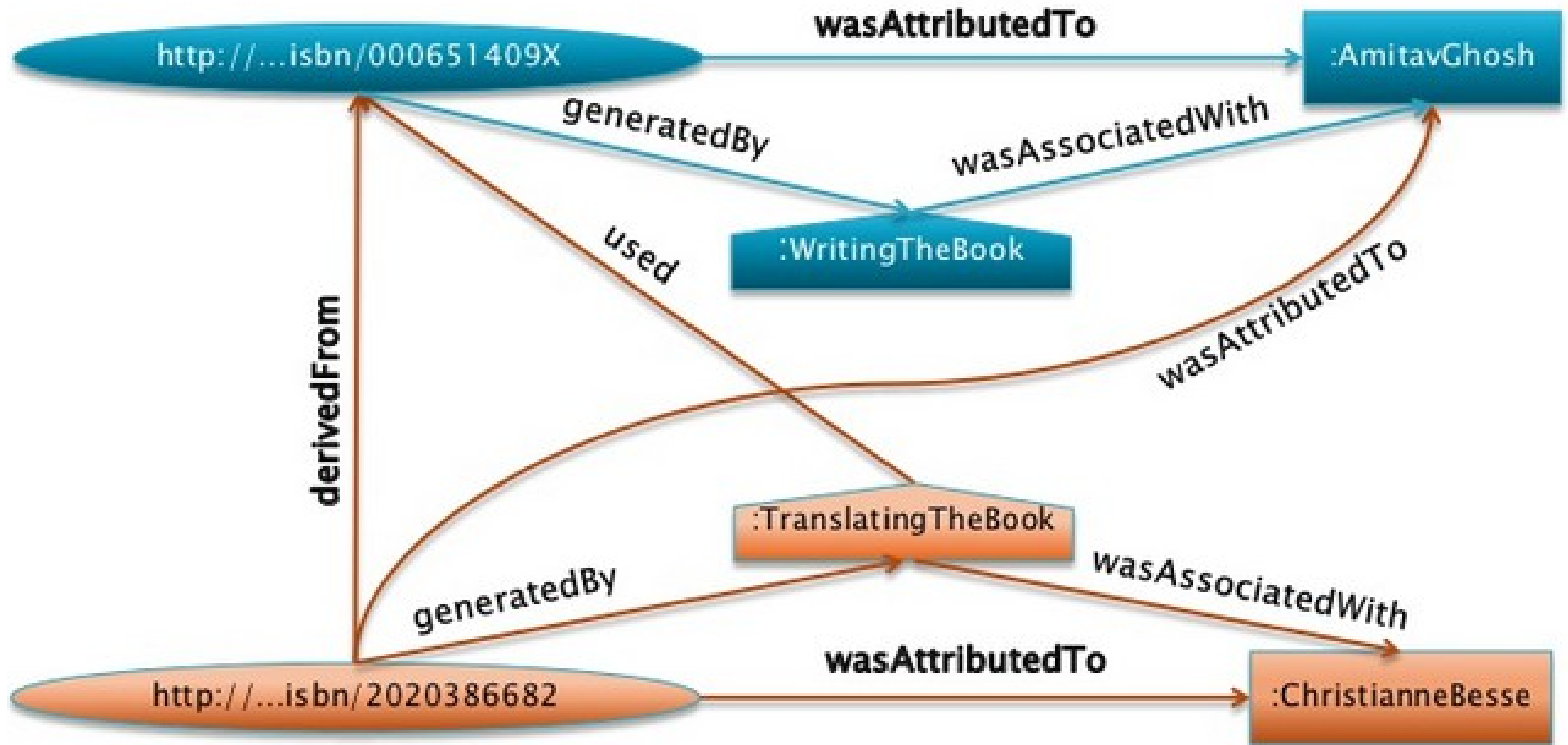
Abstract model

No limitation on types of action, entities or agents

All elements can be qualified

Relationships are *explicit*

Translation



Source: Herman (2012)

PROV categories

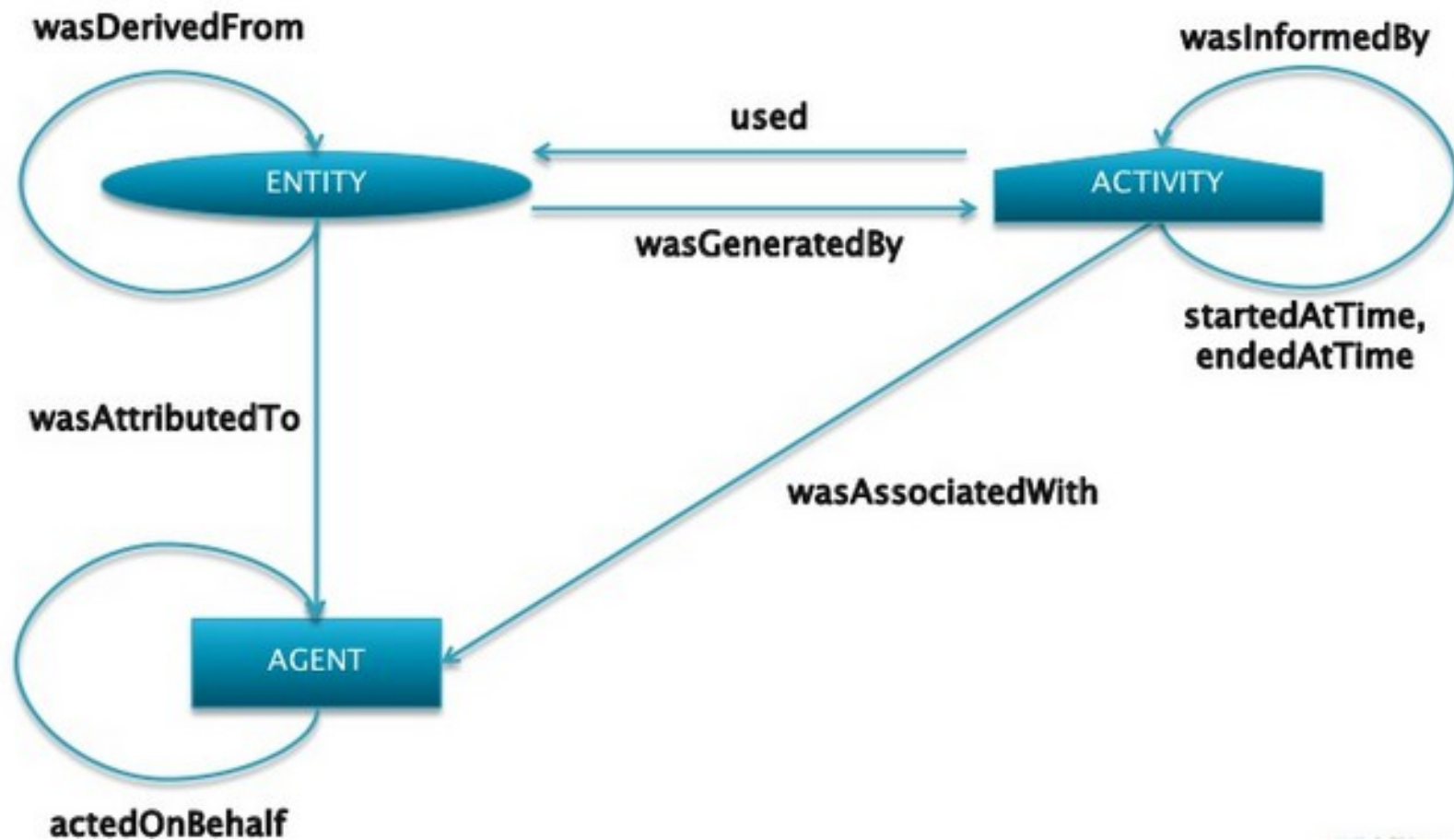
Different terms for different needs

Starting point classes and properties

Extended classes and properties

Qualified classes and properties

Starting point classes



Source: Herman (2012)

Extended classes

Subclasses of Agents

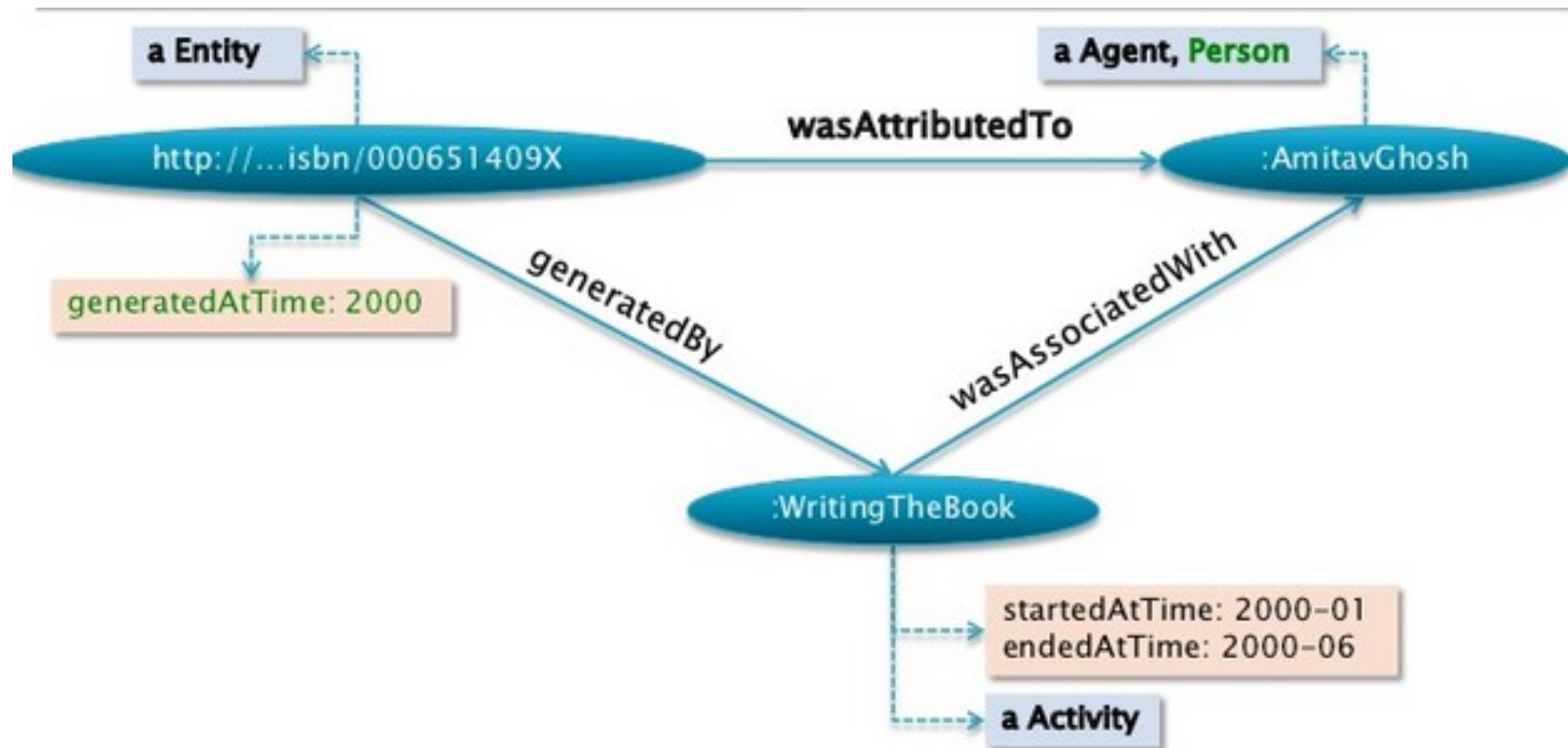
Additional properties

Versioning

Validity

Influence

Extended classes example

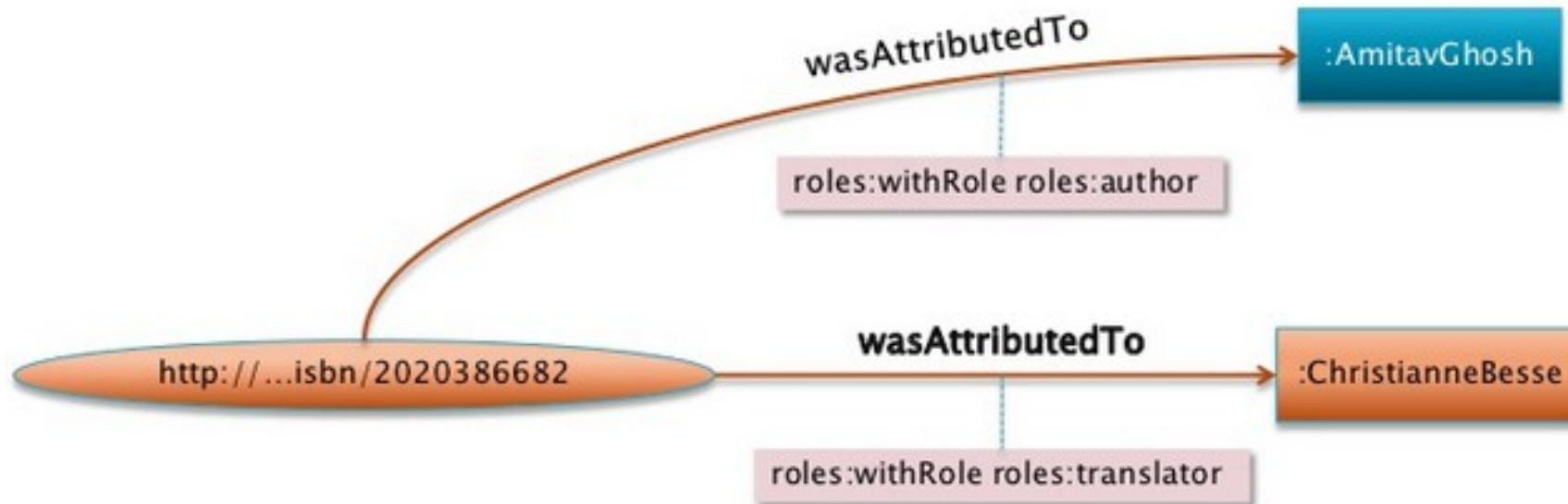


Source: Herman (2012)

Qualified classes

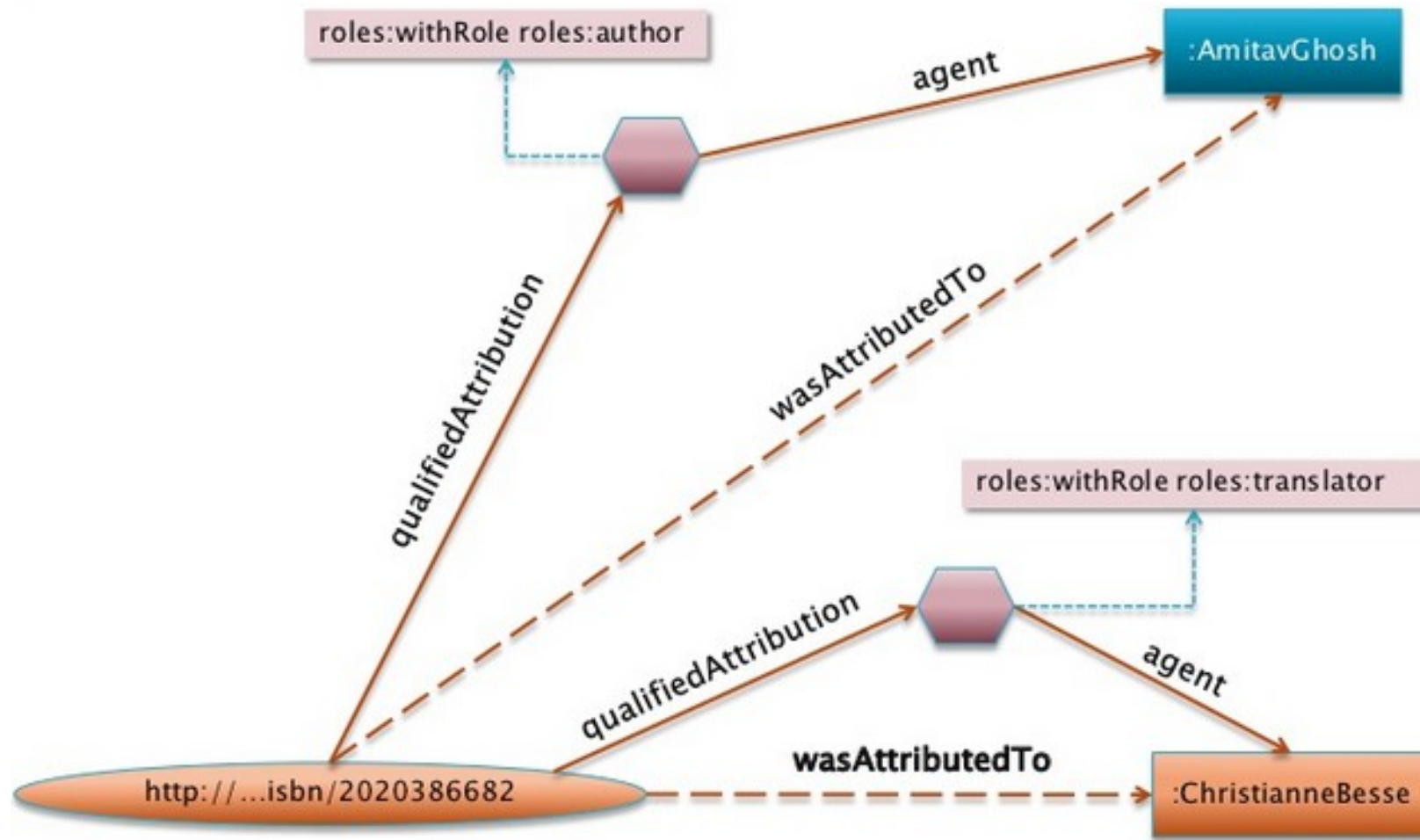
Allow additional information on relations

“qualified” statements possible



Source: Herman (2012)

Qualification structure



Source: Herman (2012)

Summary

PROV is a complex abstract model

- Initially harder to use than straightforward dcterms:

- But is able to express complex relationships

PROV models provenance by describing actions

- Activities that affected the state of a resource

- Agents that are responsible for the action

PROV can easily model the whole lifecycle of a resource

Comparison

Idea

Determine

Overlap between dcterms and PROV

Concepts from dcterms that cannot be expressed in PROV

And vice versa

Practical application

Map between dcterms and prov

Create PROV data from the multitude of available dcterms information

Create dcterms data from PROV information for legacy applications

Properties

Dublin Core

Very distinct roles, implicit and part of the semantics

Limited number of terms

PROV

Explicit modelling of roles

Subclassing

Qualified classes

1:1 mappings

Basic classes

Direct relations between entity and agent

dct:references	rdfs:subPropertyOf	prov:wasDerivedFrom
dct:creator	rdfs:subPropertyOf	prov:wasAttributedTo
dct:rightsHolder	rdfs:subPropertyOf	prov:wasAttributedTo
dct:date	rdfs:subPropertyOf	prov:generatedAtTime
dct:Agent	owl:equivalentClass	prov:Agent
prov:hadOriginalSource	rdfs:subPropertyOf	dct:source
prov:wasRevisionOf	rdfs:subPropertyOf	dct:isVersionOf

Source: Eckert (2012)

PROV specialisations

Dcterm concepts that PROV lacks

Differentiation between creator and contributor

dcprov:CreationActivity	rdfs:subClassOf	prov:Activity, dcprov:ContributionActivity
dcprov:ContributionActivity	rdfs:subClassOf	prov:Activity
dcprov:CreatorRole	rdfs:subClassOf	prov:Role, dcprov:ContributorRole
dcprov:ContributorRole	subClassOf	prov:Role

Source: Eckert (2012)

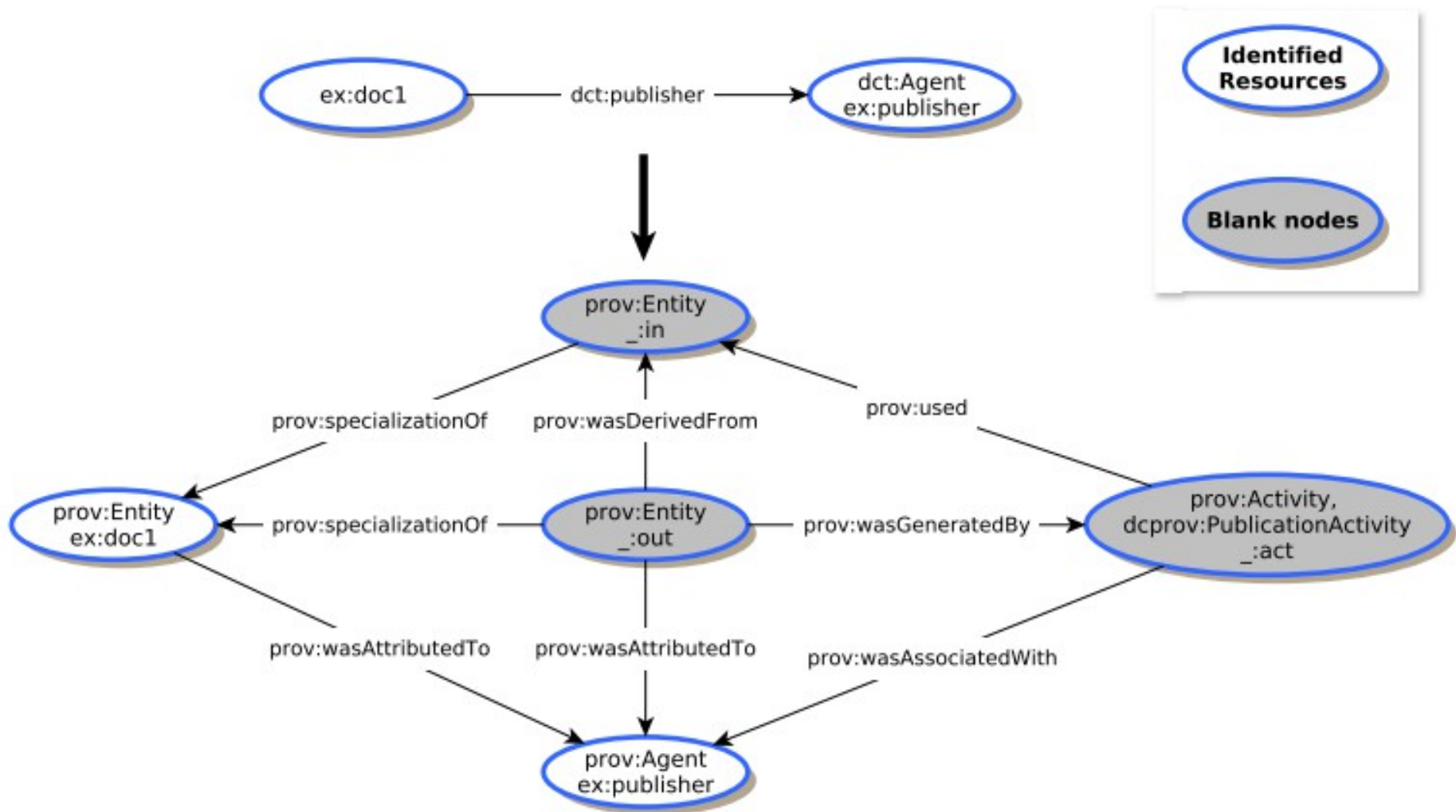
Complex Mappings

Creating PROV statements from dct:creator

```
CONSTRUCT {  
  ?doc a prov:Entity .  
      prov:wasAttributedTo ?ag .  
  _:out a prov:Entity .  
      prov:specializationOf ?doc .  
  ?ag a prov:Agent .  
  _:act a prov:Activity, dcprov:CreationActivity ;  
      prov:wasAssociatedWith ?ag ;  
      prov:qualifiedAssociation _:assoc .  
  _:assoc a prov:Association ;  
      prov:agent ?ag ;  
      prov:hadRole dcprov:CreatorRole .  
  _:out prov:wasGeneratedBy _:act ;  
      prov:wasAttributedTo ?ag .  
} WHERE {  
  ?doc dct:creator ?ag .  
}
```

Source: Eckert (2012)

Example



Source: Eckert (2012)

DC ↔ PROV

Mapping will be part of the coming PROV update

An editorial draft is available — your input is welcome!

<https://dvcs.w3.org/hg/prov/raw-file/default/dc-note/Overview.html>

Conclusion

Distinct use cases

Dublin Core as a simple provenance vocabulary with explicit roles

PROV as an expressive abstract model

Modelling complex provenance chains and relations

Mapping possible

Much provenance information is within Dublin Core metadata

Semantic Web applications can make sense of both models, if they understand at least one.

References

W3C Provenance Working Group. (2012c). PROV-O: The PROV Ontology (T. Lebo, S. Sahoo, & D. McGuinness, Eds.). W3C.

<http://www.w3.org/TR/> (Working Draft)

DCMI Usage Board. (2010a). DCMI Metadata Terms. Dublin Core Metadata Initiative.

<http://dublincore.org/documents/2010/10/11/dcmi-terms/>

Ivan Herman. (2012). The W3C PROV vocabulary. (Presentation slides), examples used with permission

http://de.slideshare.net/ivan_herman/the-w3c-prov-vocabulary