Biodiversity Information Standards Architecture

Roger Hyam <roger@tdwg.org>

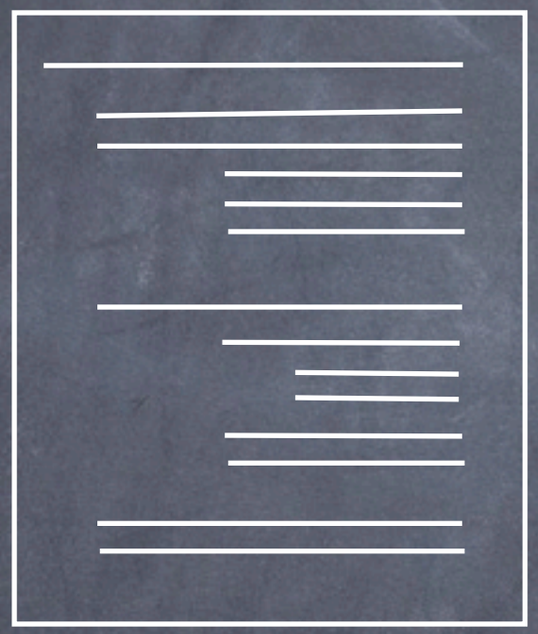
16 April 2007

Documents

Documents

TDWG Standards have been based on documents validated using XML Schema

Documents



Valid to XSD

Documents

- Valid to XSD
- Single Namespace

Documents

Valid to XSD

Single Namespace

Very good for transfer protocols

Documents

Valid to XSD

Single Namespace

Very good for transfer protocols

Sender & Receiver are same application

Universal Biodiversity Data Bus

The primary use case for standards is sharing data over the UBDB.

Client 2

Data 2

Data 1

Universal Biodiversity Data Bus

Client 1

Heterogeneous data from multiple sources is used by different clients & services

Client 2

Data 2

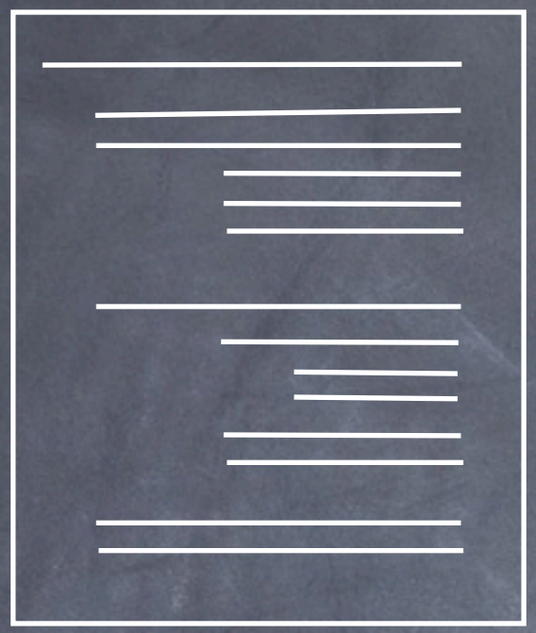
Data 1

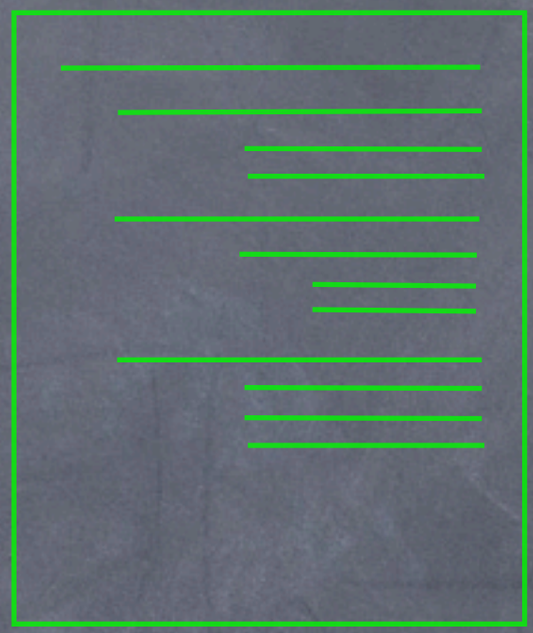
Universal Biodiversity Data Bus

Client 1

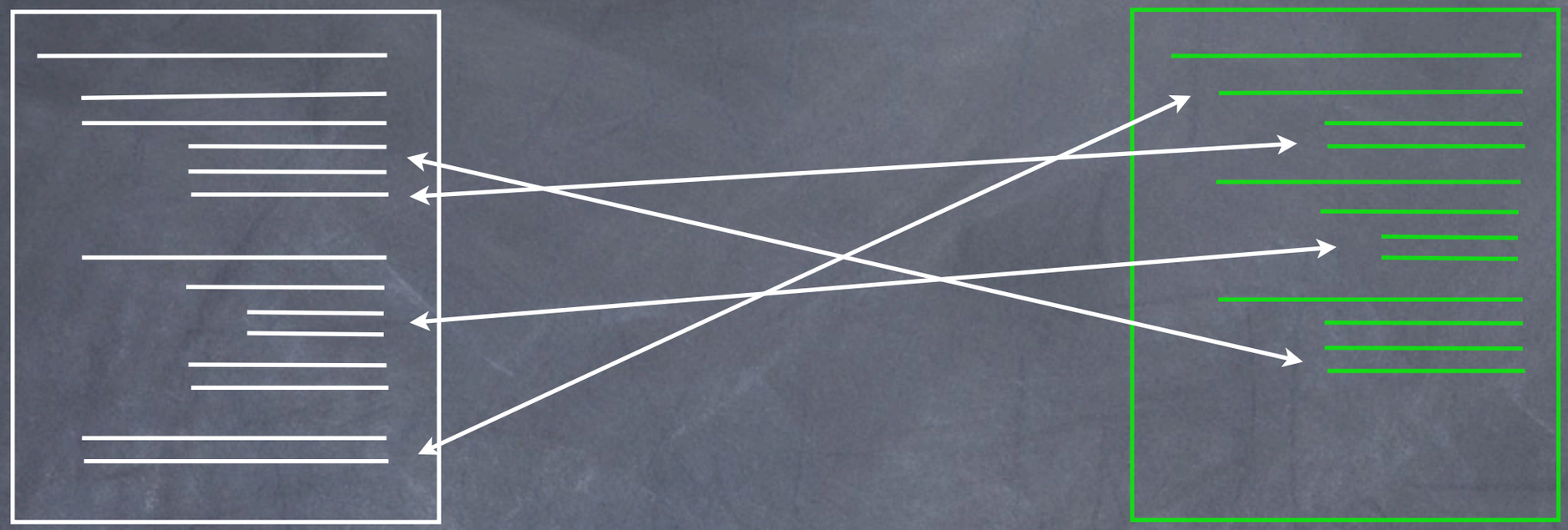
Service 1

Service 2



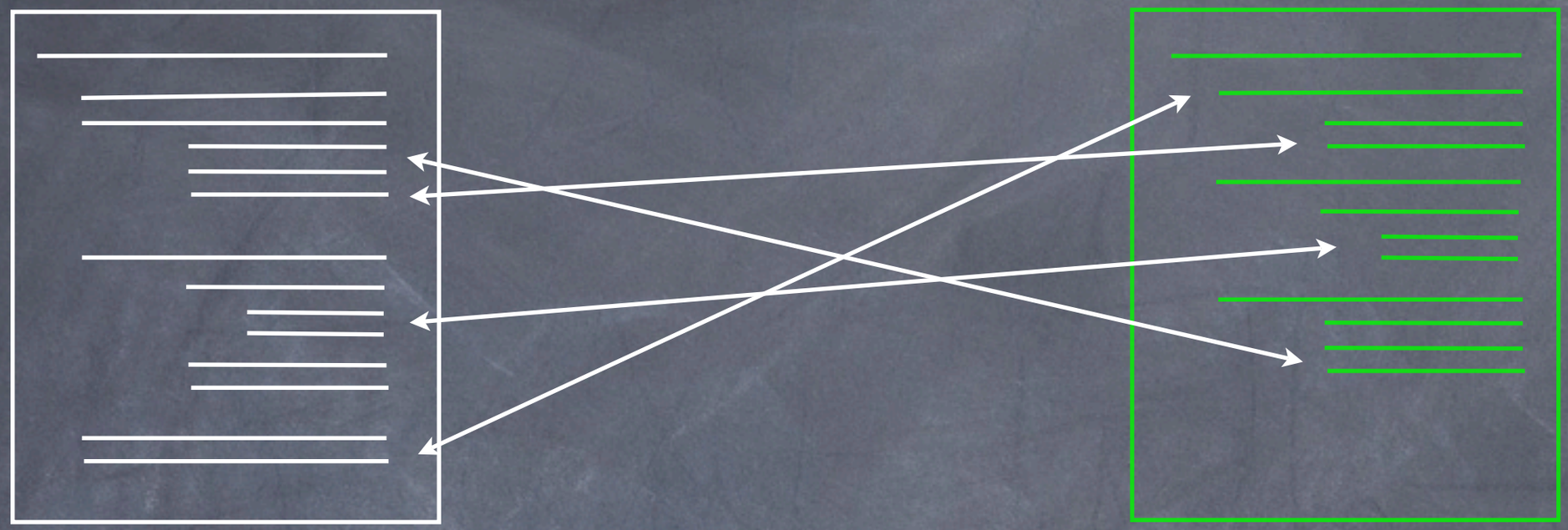


Clients & Services need to combine data from different sources to use it.

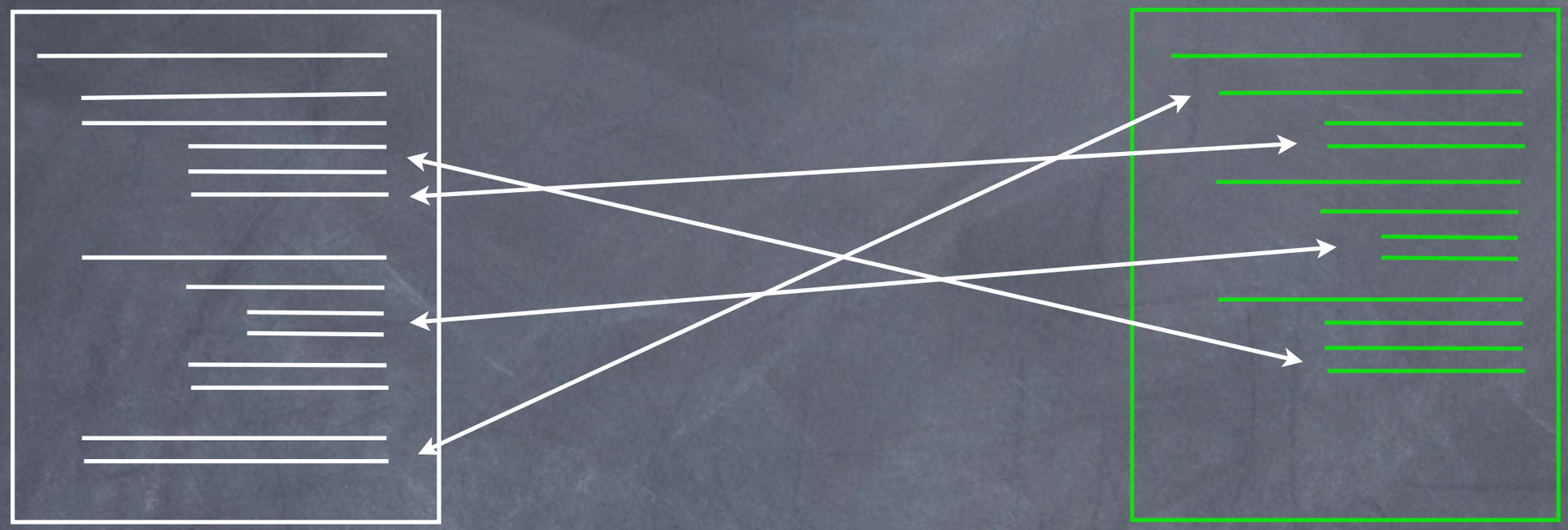


Clients & Services need to combine data from different sources to use it.

They need to related different parts of documents together.

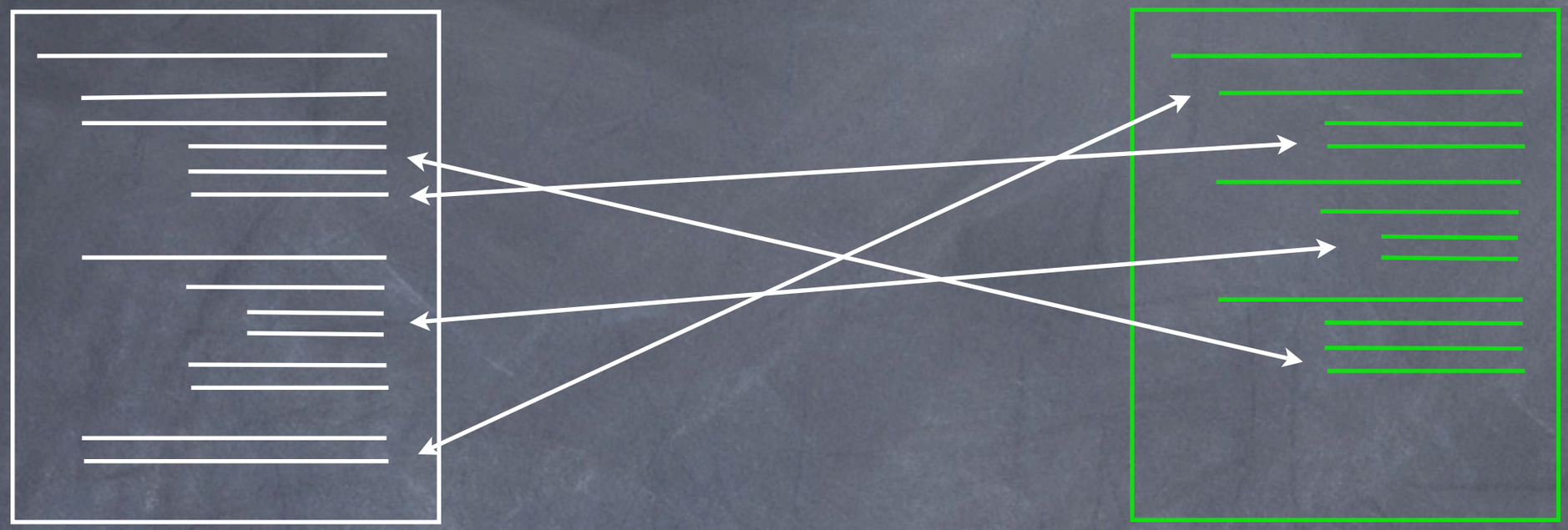


Mapping documents is difficult.



Mapping documents is difficult.

Combining data is difficult.



Mapping documents is difficult.

Combining data is difficult.

Primary use case is difficult.

Specific Epithet

Specific Epithet

This is an example from ABCD and DarwinCore

Specific Epithet

/DataSets/DataSet/Units/Unit/Identifications/
Identification/Result/TaxonIdentified/
ScientificName/NameAtomised/Botanical/FirstEpithet

Specific Epithet

/DataSets/DataSet/Units/Unit/Identifications/
Identification/Result/TaxonIdentified/
ScientificName/NameAtomised/Botanical/FirstEpithet

==

Specific Epithet

/DataSets/DataSet/Units/Unit/Identifications/
Identification/Result/TaxonIdentified/
ScientificName/NameAtomised/Botanical/FirstEpithet

==

/DataSets/DataSet/Units/Unit/Identifications/
Identification/Result/TaxonIdentified/
ScientificName/NameAtomised/Zoological/
SpeciesEpithet

Specific Epithet

/DataSets/DataSet/Units/Unit/Identifications/
Identification/Result/TaxonIdentified/
ScientificName/NameAtomised/Botanical/FirstEpithet

==

/DataSets/DataSet/Units/Unit/Identifications/
Identification/Result/TaxonIdentified/
ScientificName/NameAtomised/Zoological/
SpeciesEpithet

==

Specific Epithet

/DataSets/DataSet/Units/Unit/Identifications/
Identification/Result/TaxonIdentified/
ScientificName/NameAtomised/Botanical/FirstEpithet

==

/DataSets/DataSet/Units/Unit/Identifications/
Identification/Result/TaxonIdentified/
ScientificName/NameAtomised/Zoological/
SpeciesEpithet

==

/DarwinCoreRecord/SpecificEpithet

Specific Epithet

Such mappings have to be maintained manually for every document-document combination.

- specificEpithet

It helps to think of a specific epithet as the property of an object...

TaxonName

- genusPart
- specificEpithet
- authorship
- publishedInCitation

It helps to think of a specific epithet as the property of an object...

TaxonName

- genusPart
- specificEpithet
- authorship
- publishedInCitation

It helps to think of a specific epithet as the property of an object...

and for some properties to contain literals and some to contain objects

TaxonName

- genusPart
- specificEpithet
- authorship
- publishedInCitation

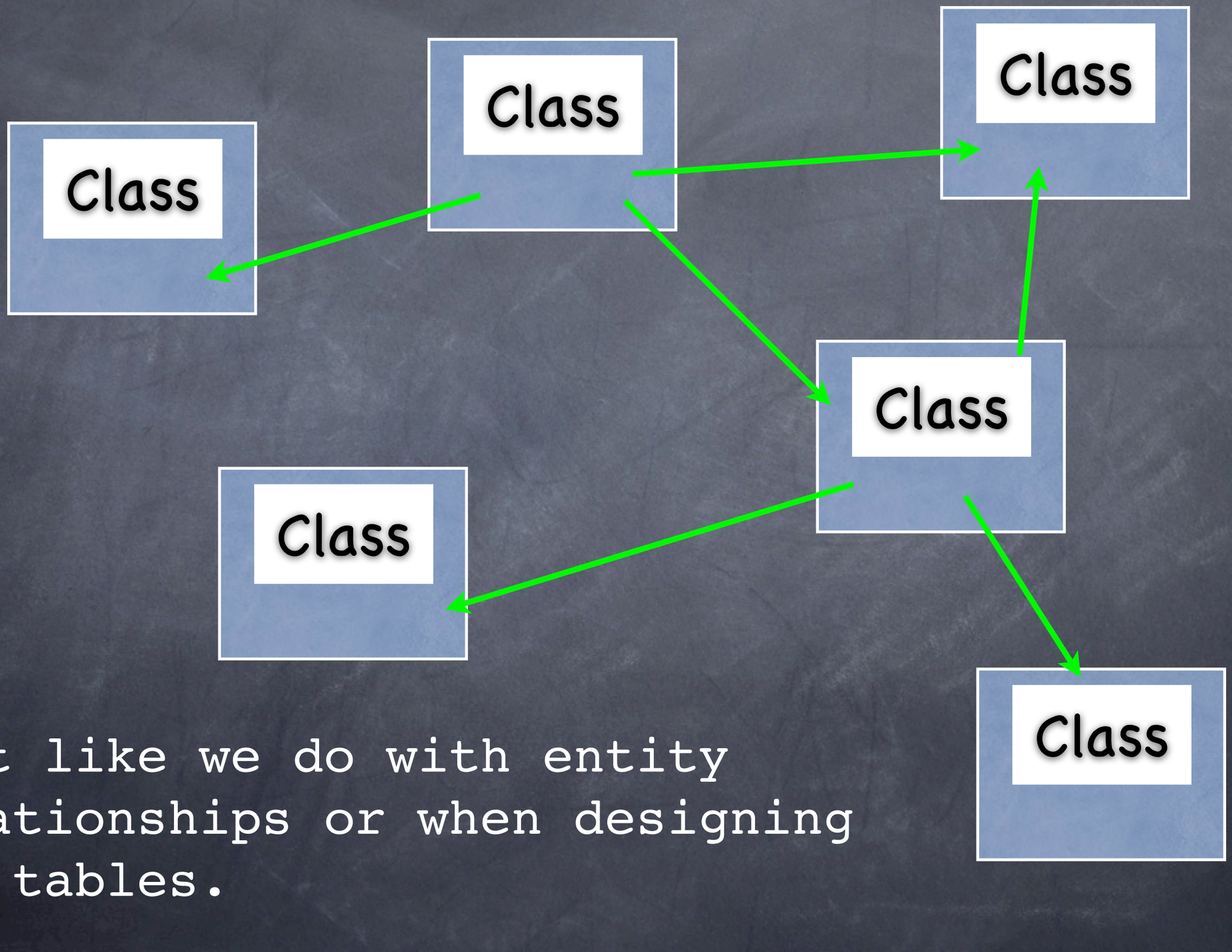
It helps to think of a specific epithet as the property of an object...

PublicationCitation

- title
- authorTeam
- isbn
- date

and for some properties to contain literals and some to contain objects

This way we can build a graph of objects



Just like we do with entity relationships or when designing SQL tables.

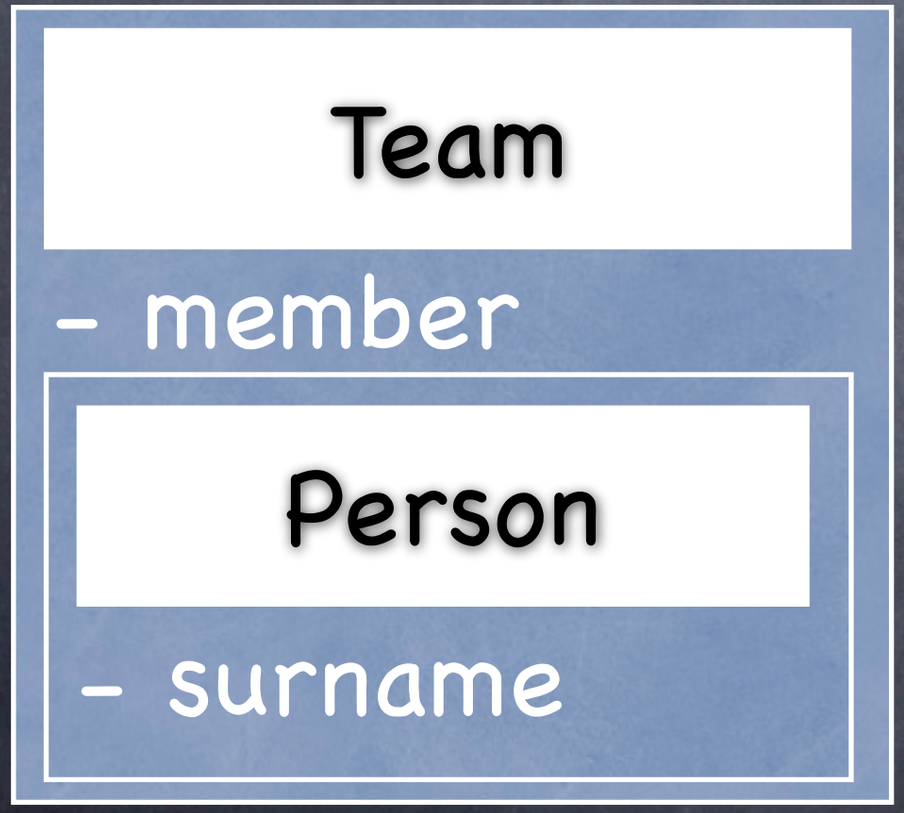
The semantics of data encoded according to the graph depends on perspective.

The semantics of data encoded according to the graph depends on perspective.

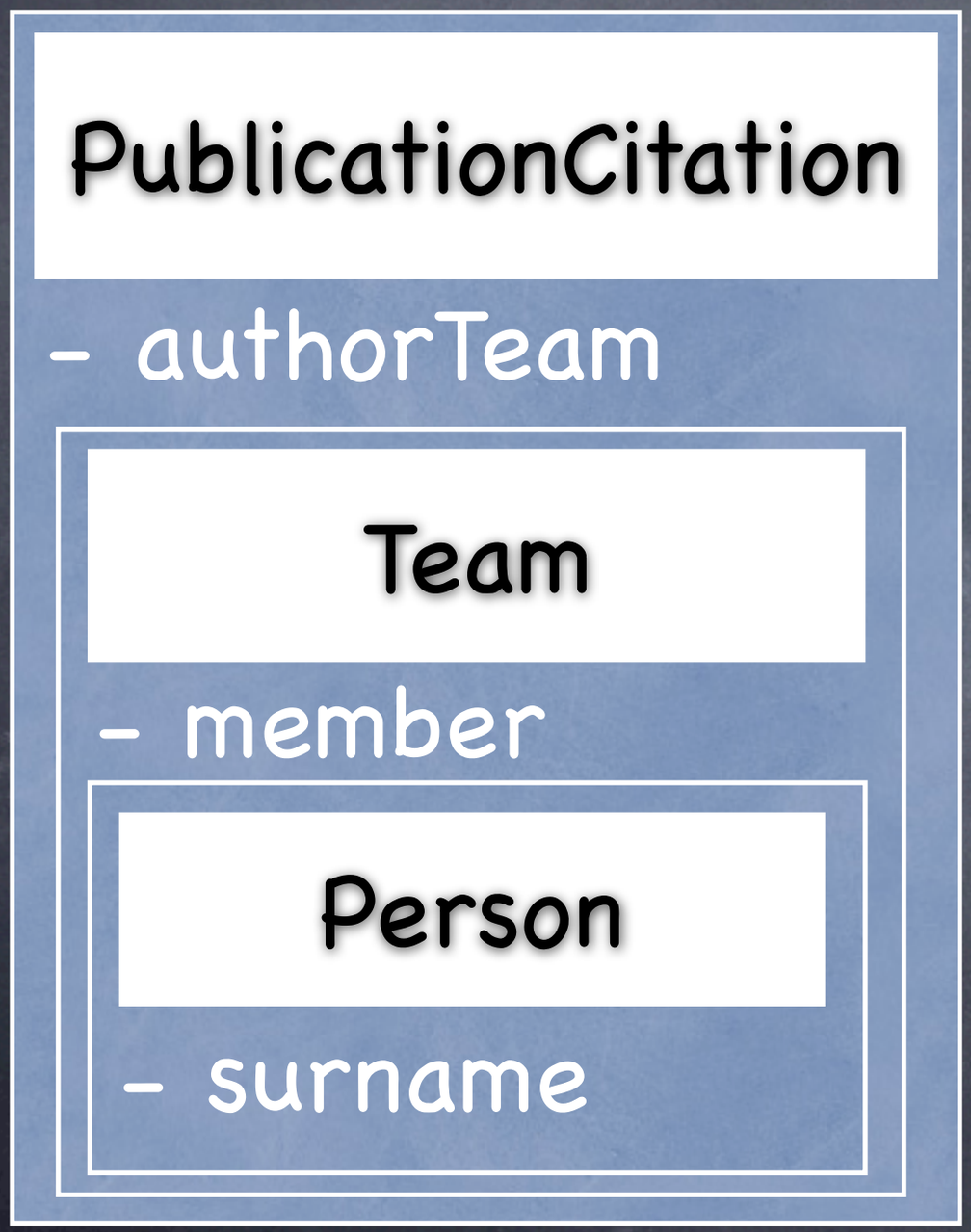
- surname

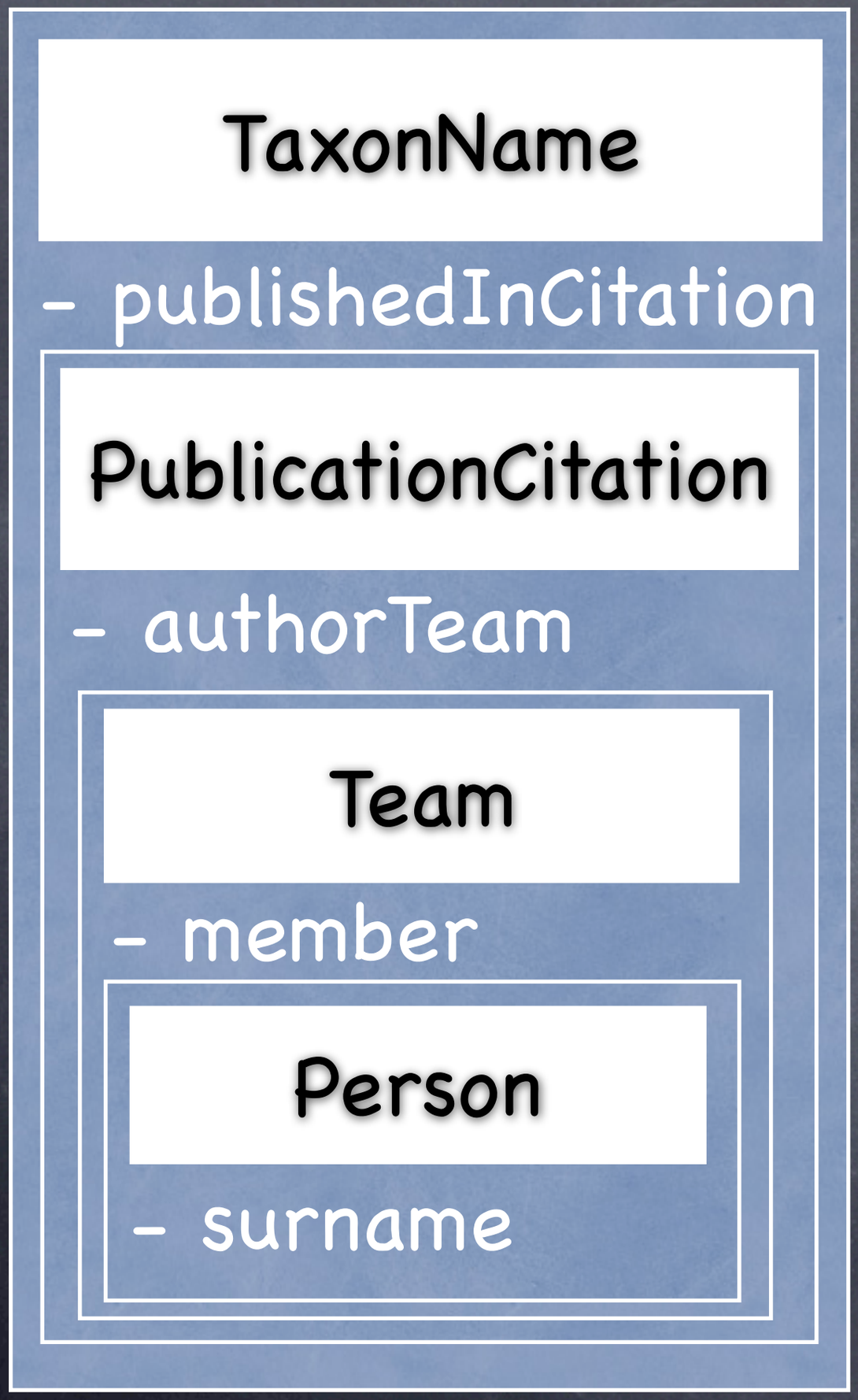
The semantics of data encoded according to the graph depends on perspective.

The semantics of data encoded according to the graph depends on perspective.

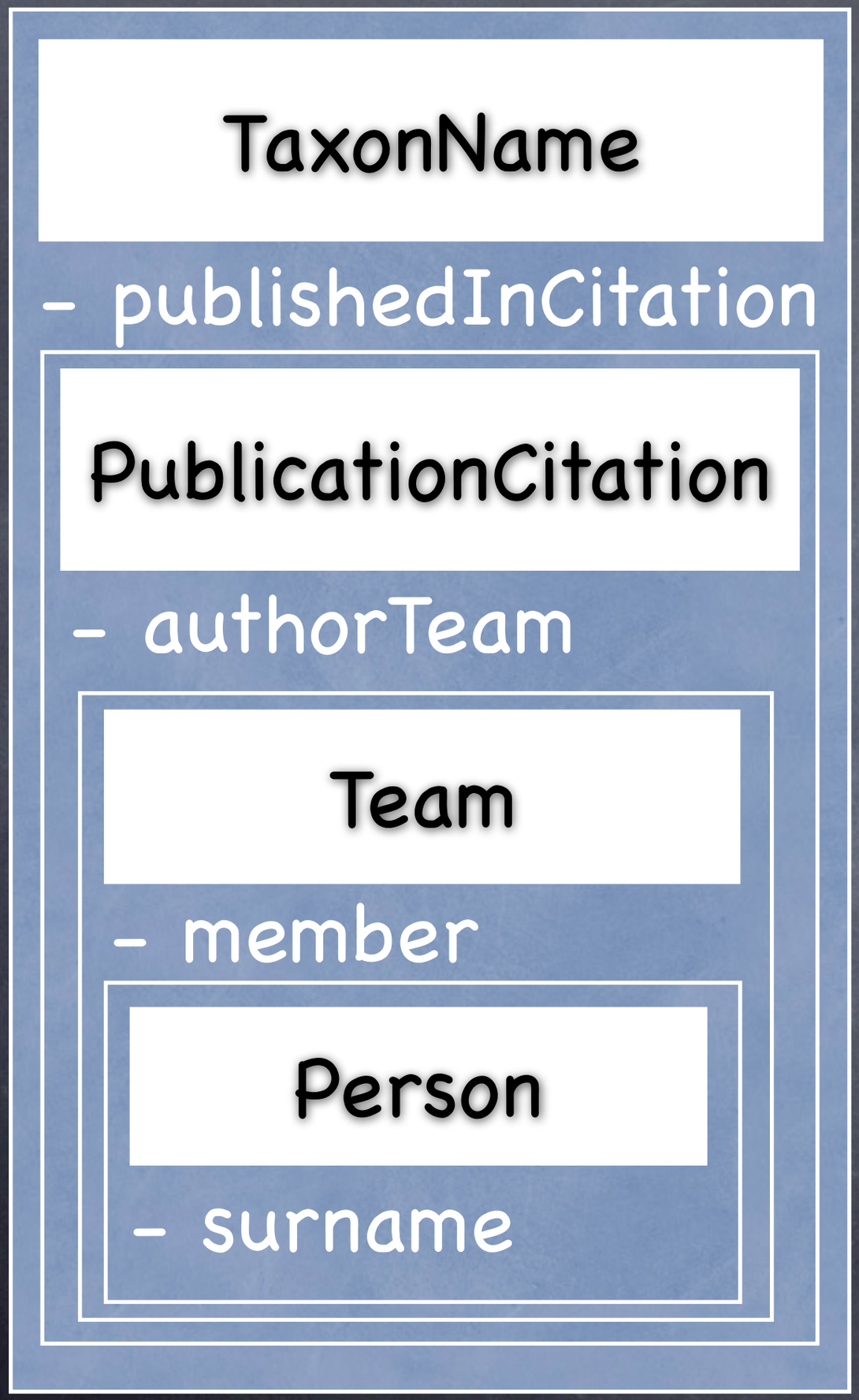


The semantics of data encoded according to the graph depends on perspective.



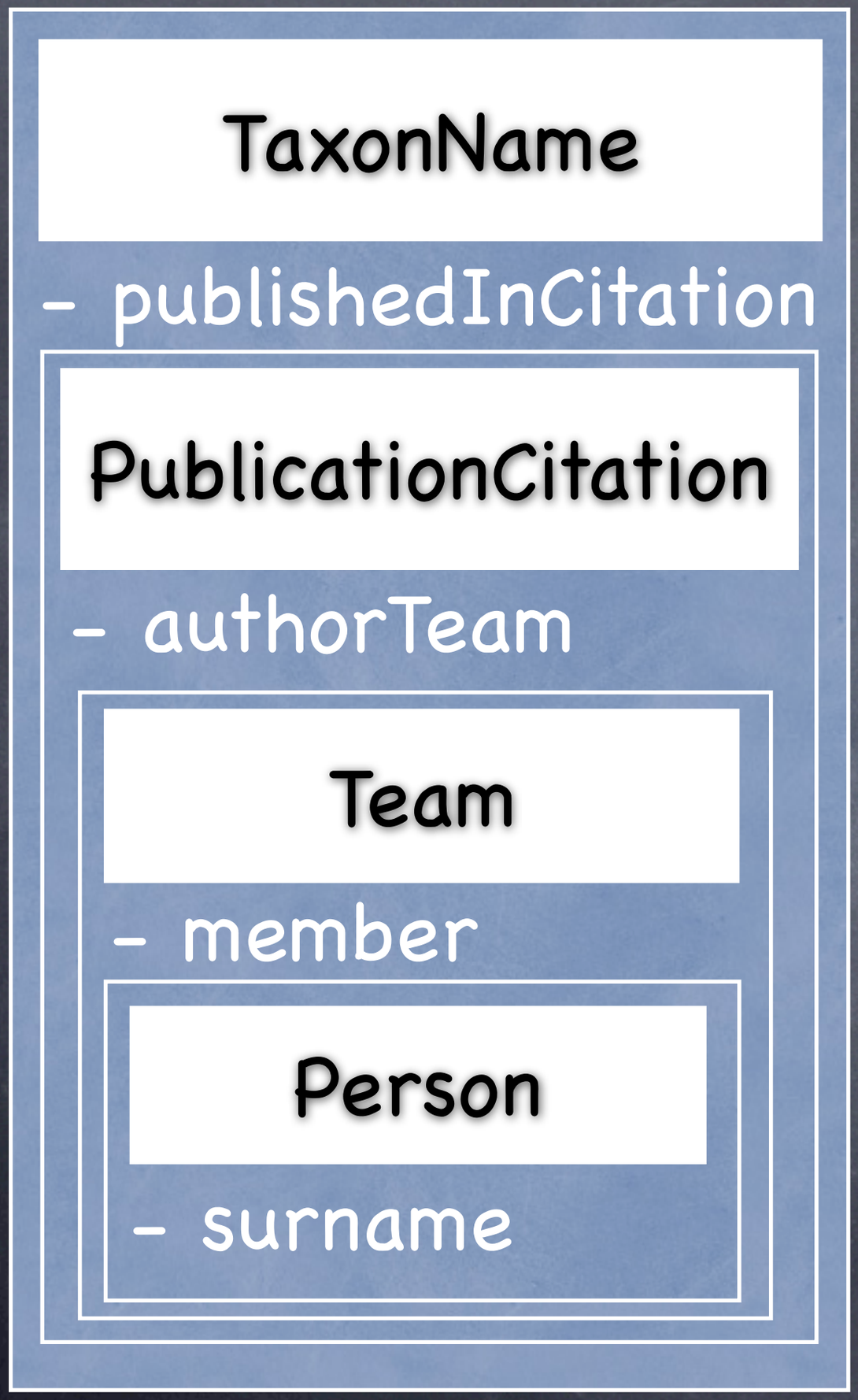


The semantics of data encoded according to the graph depends on perspective.



The semantics of data encoded according to the graph depends on perspective.

This is the surname of the author of a taxon name.



The semantics of data encoded according to the graph depends on perspective.

This is the surname of the author of a taxon name.

Viewing the data this way it looks like a document ...

TaxonName

- publishedInCitation

PublicationCitation

- authorTeam

Team

- member

Person

- surname

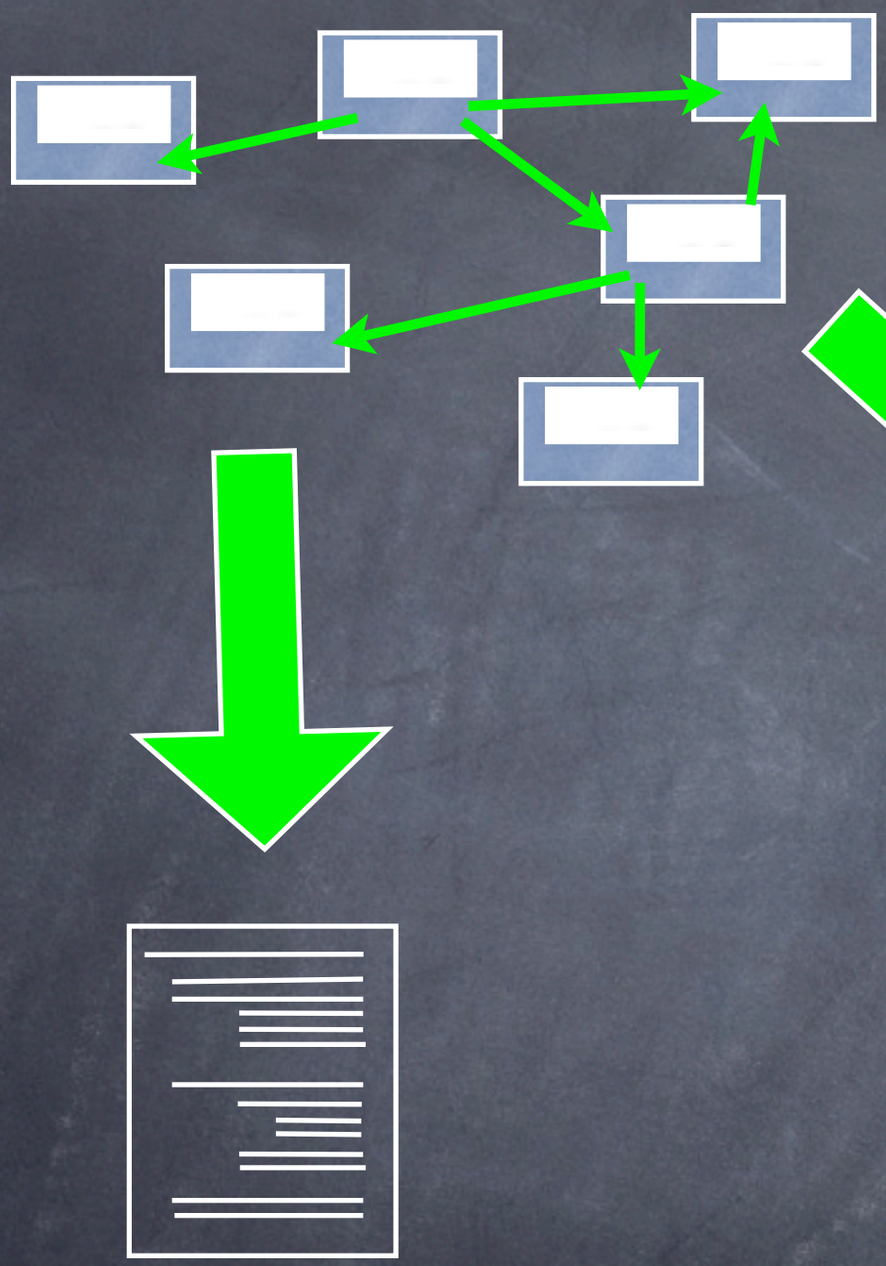
```
<TaxonName rdf:about="123">
  <publishedIn>
    <PulicationCitation>
      <authorTeam>
        <Team>
          <member>
            <Person>
              <surname>
                ...
              </surname>
            </Person>
          </member>
        </Team
      </authorTeam>
    </PulicationCitation>
  </publishedIn>
</TaxonName>
```

Modeling data as objects does not prevent the construction of documents.

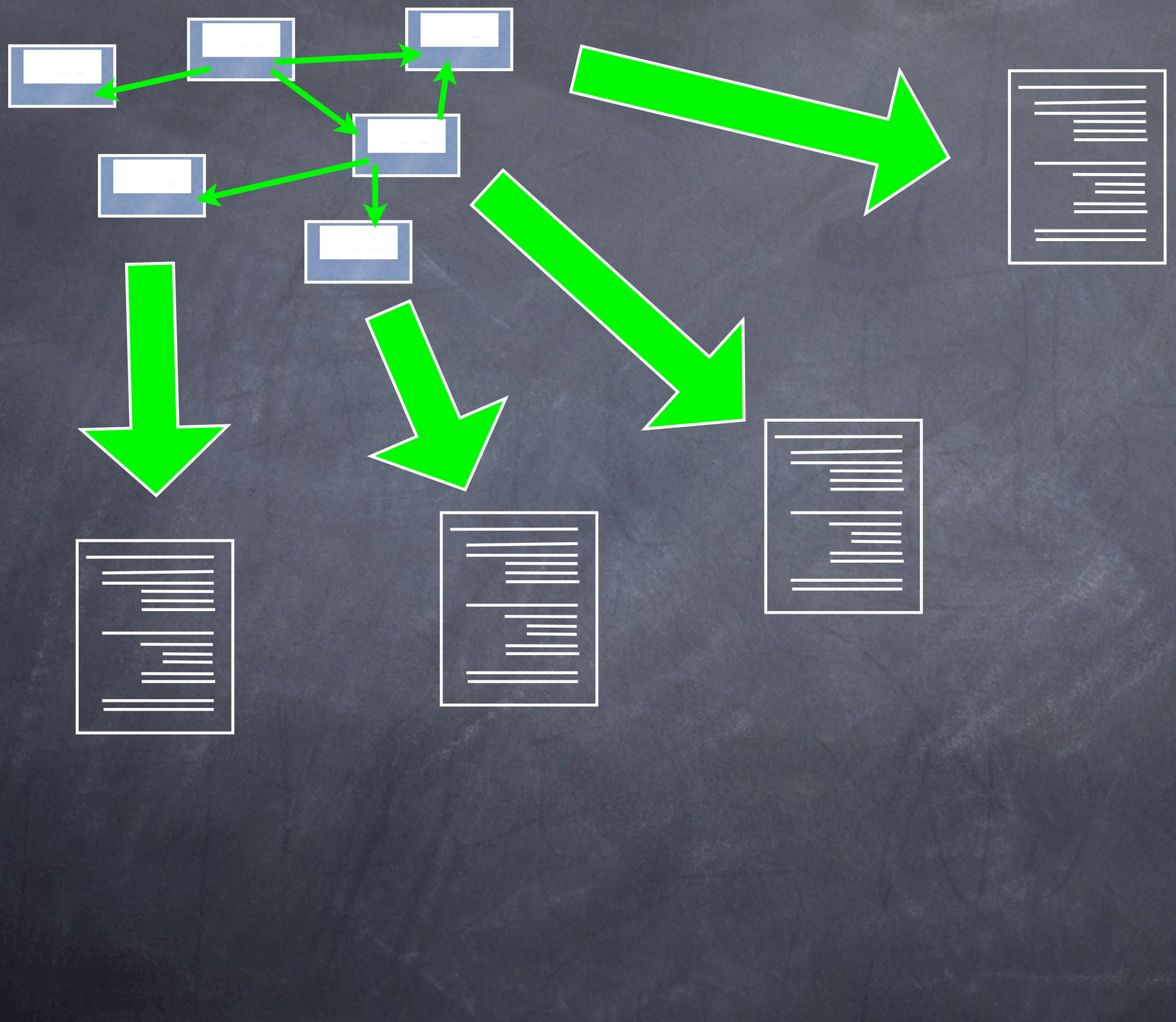
```
<TaxonName rdf:about="123">
  <publishedIn>
    <PulicationCitation>
      <authorTeam>
        <Team>
          <member>
            <Person>
              <surname>
                ...
              </surname>
            </Person>
          </member>
        </Team
      </authorTeam>
    </PulicationCitation>
  </publishedIn>
</TaxonName>
```

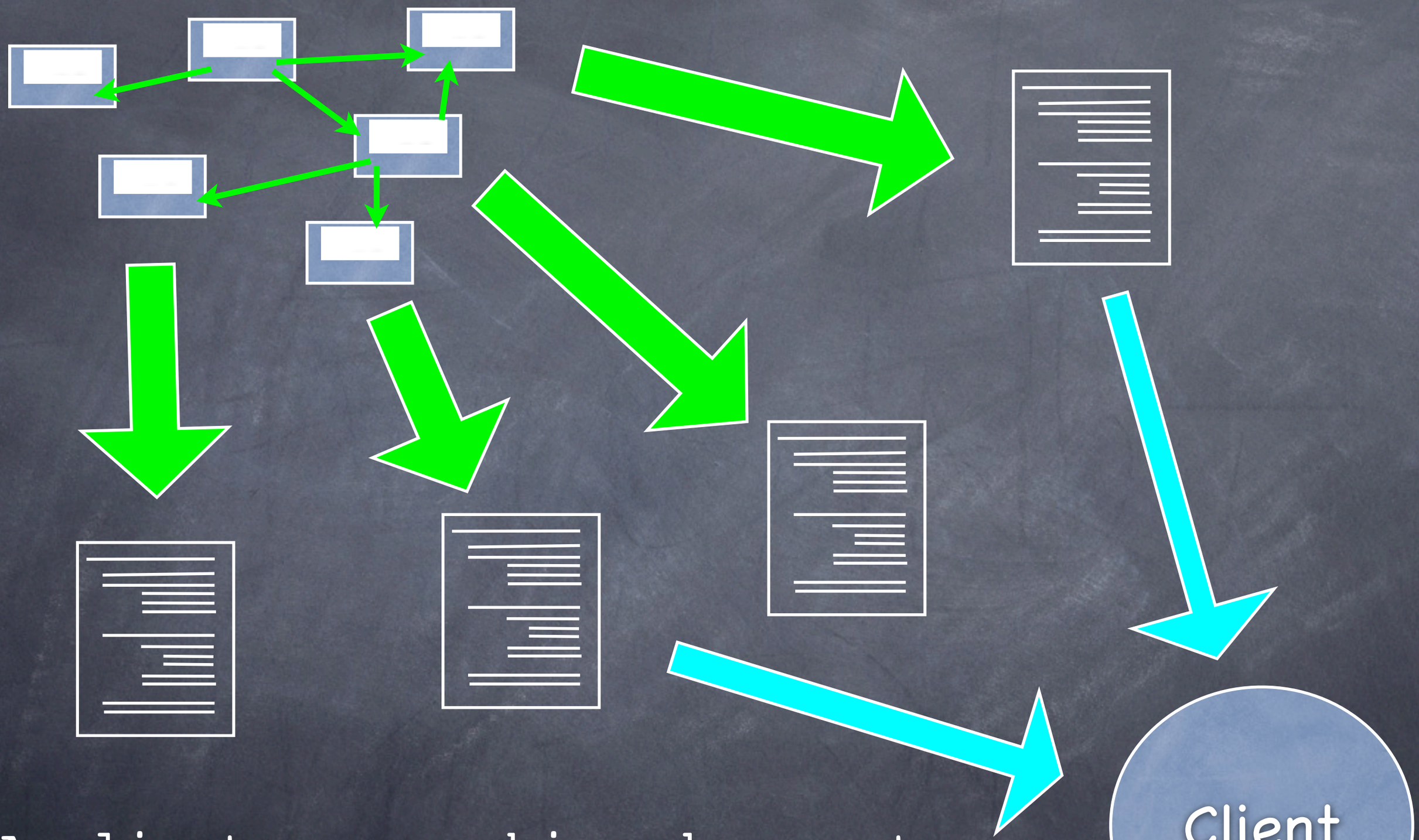
Object Orientated Documents Give Us The

Best of Both Worlds



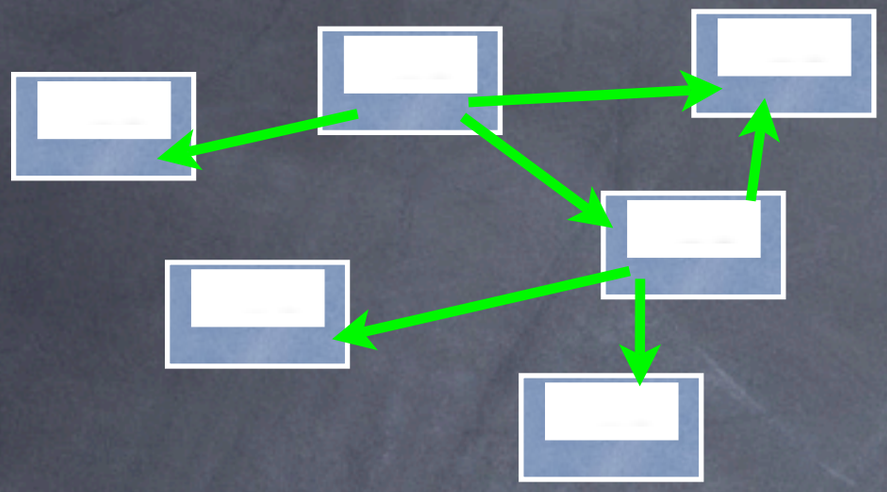
We can construct documents from the perspective of different base classes in the model.





A client can combine documents
from different perspectives
because it understands the model

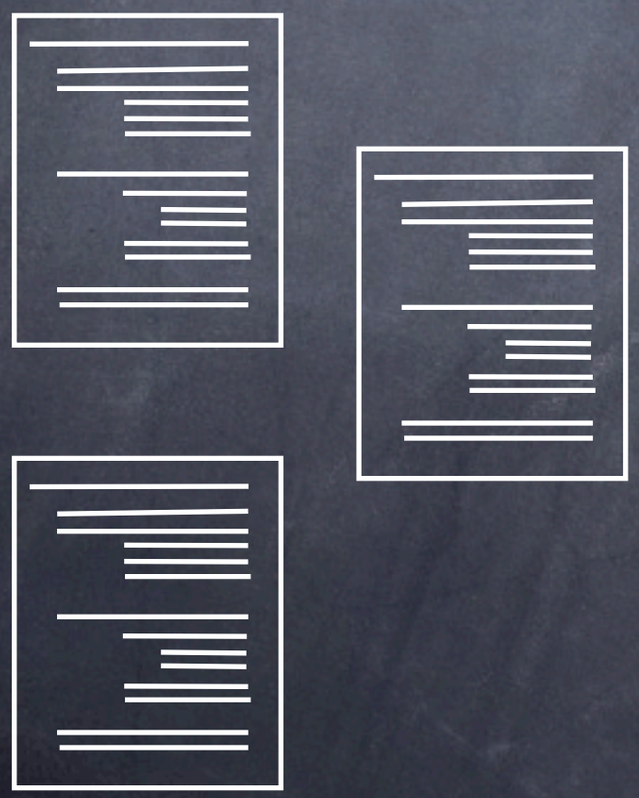
We can use exstablished technologies to work with object orientated documents.



UML to design/discuss

OWL to specify

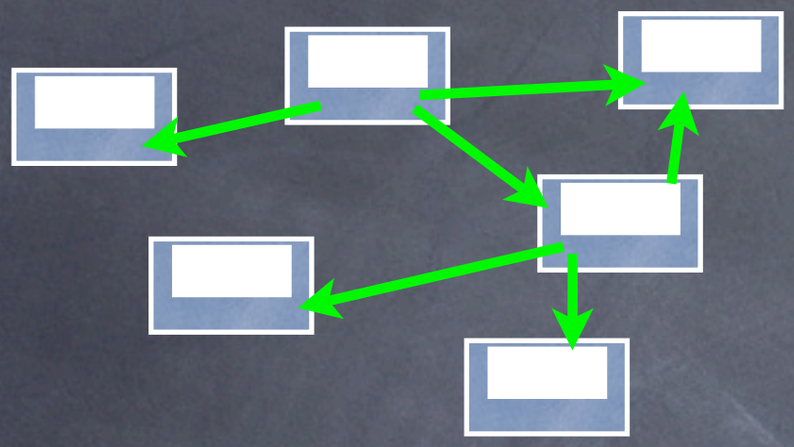
RDF/S as instance data



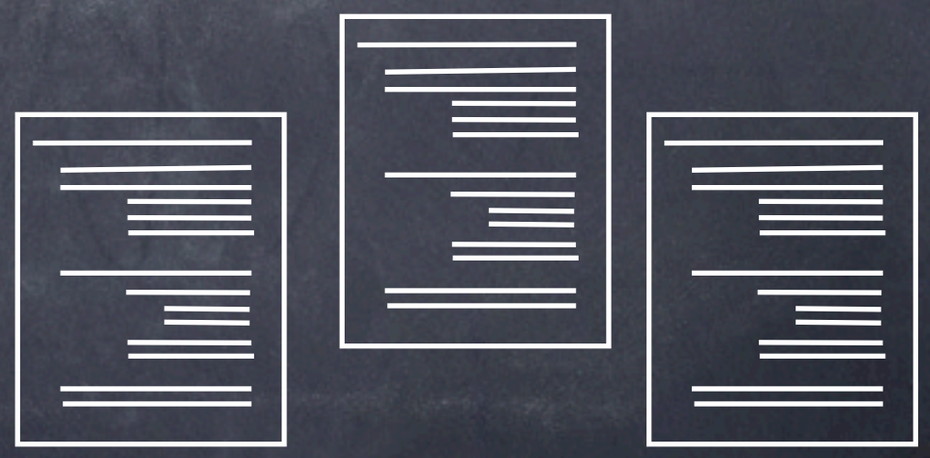
XSD to validate

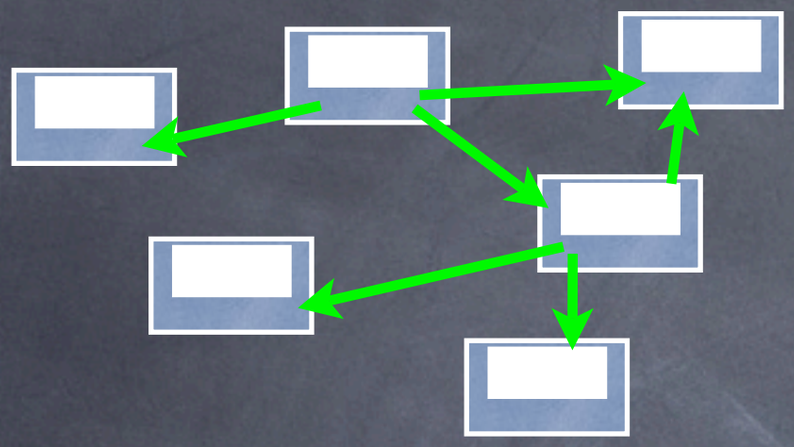
XSD for TAPIR output

Software implementers have a choice. They either treat data as XML documents or as pure RDF.



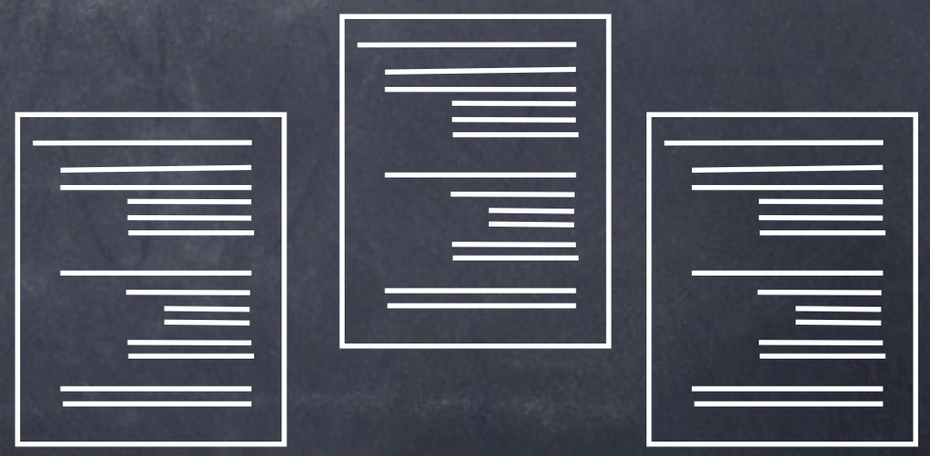
	Consume as RDF	Consume as XML
Produce as RDF	No Limitations	Problematic
Produce as XML	No Limitations	Minor Limitations



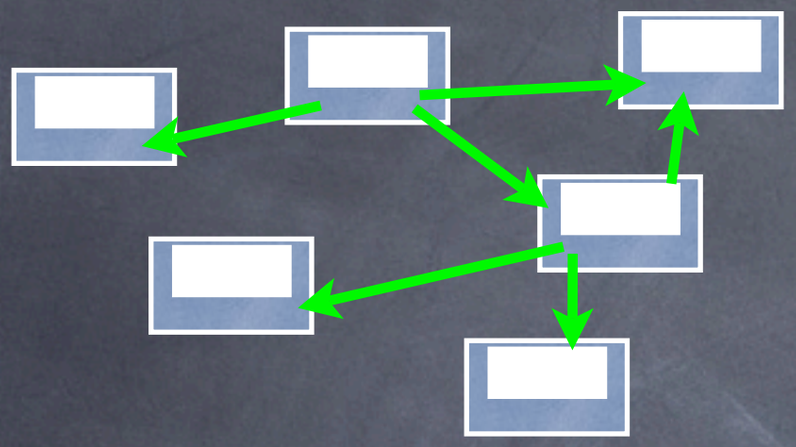


	Consume as RDF	Consume as XML
Produce as RDF	No Limitations	Problematic
Produce as XML	No Limitations	Minor Limitations

This is TAPIR

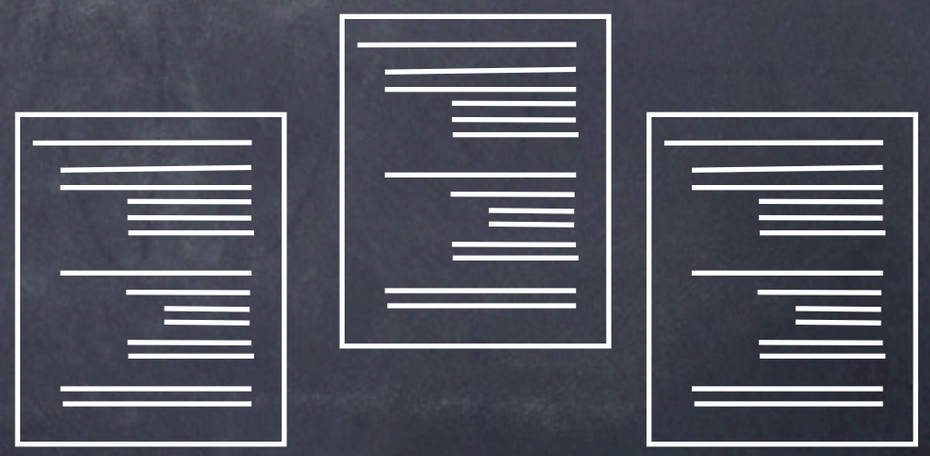


This is a Semantic Web Client



	Consume as RDF	Consume as XML
Produce as RDF	No Limitations	Problematic
Produce as XML	No Limitations	Minor Limitations

This is TAPIR



The existing XML Schemas and the TDWG Ontology can be combined into a single architecture using TAPIR...

XML Schemas

XML
Schemas

TDWG
Ontology

XML
Schemas

Informs

TDWG
Ontology

TAPIR Query Filters

TAPIR Output Models

TAPIR Concepts

TAPIR Concepts

Avowed Structures

XML Schemas

Informs

TDWG Ontology

TAPIR Query Filters

TAPIR Output Models

TAPIR Concepts

TAPIR Concepts

Avowed Structures

LSID Vocabularies

XML Schemas

Informs

TDWG Ontology

TAPIR Query Filters

TAPIR Output Models

TAPIR Concepts

TAPIR Concepts

Avowed Structures

LSID Vocabularies

XML Schemas

Informs

TDWG Ontology

Semantic Web

What to do?

- Visit the TAG wiki for latest news.
  http://wiki.tdwg.org/twiki/bin/view/TAG/WebHome
- Look at the LSID Vocabularies - they are the basic units for the ontology.
  http://wiki.tdwg.org/twiki/bin/view/TAG/LsidVocs
- Questions to Roger Hyam <roger@tdwg.org>
- Questions to TAG List <tdwg-tag@lists.tdwg.org>