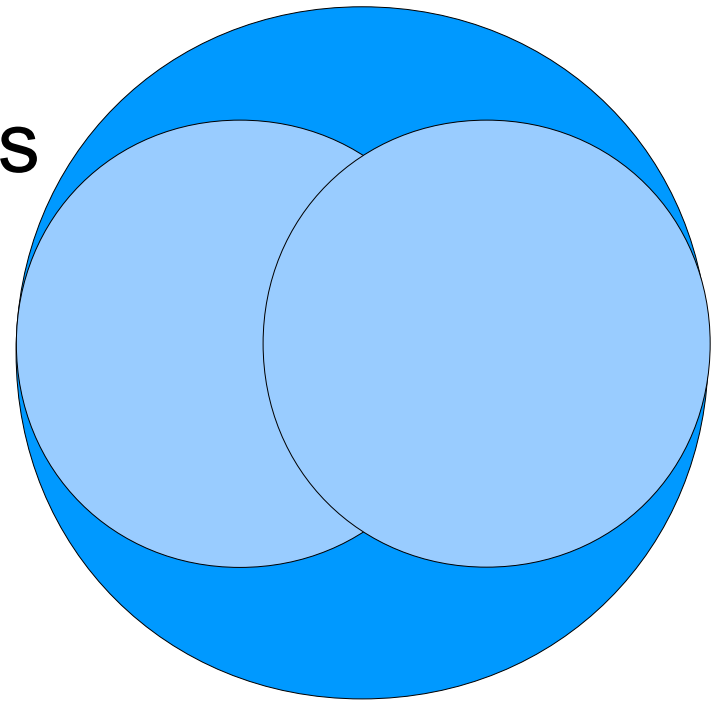


XHTML Dialects:

Managing variant perspectives
with upward expansion

Erik Hennum
2010-8-3



Objectives of the talk

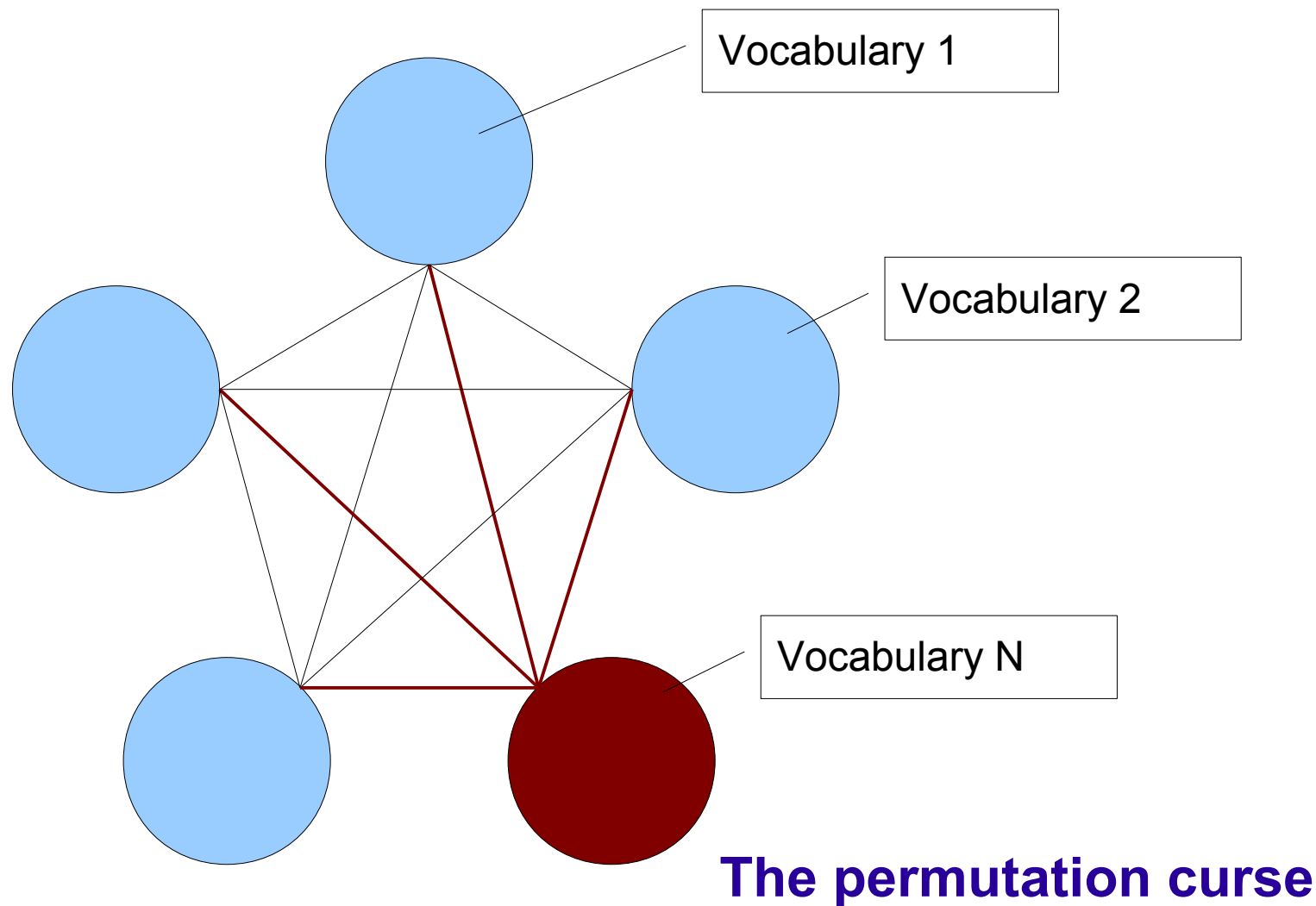
- **Propose enhancing XML for managing vocabulary relations**
 - Drawing on RDF for type relations
 - Particular focus on XHTML
- **Invite feedback**

Points

➤ **Standards and customizations**

- Vocabulary relations
- Theoretical alignments
- Dialects
- XHTML Dialects

Life without a standard vocabulary



Customizable standards – a double bind

This vocabulary captures expert consensus

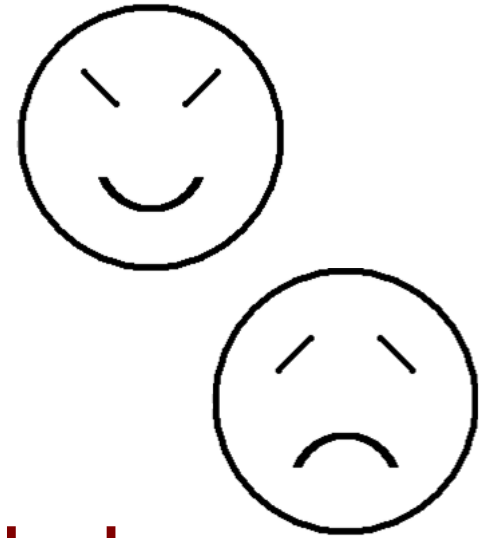
But you probably have special requirements

You can customize any aspect

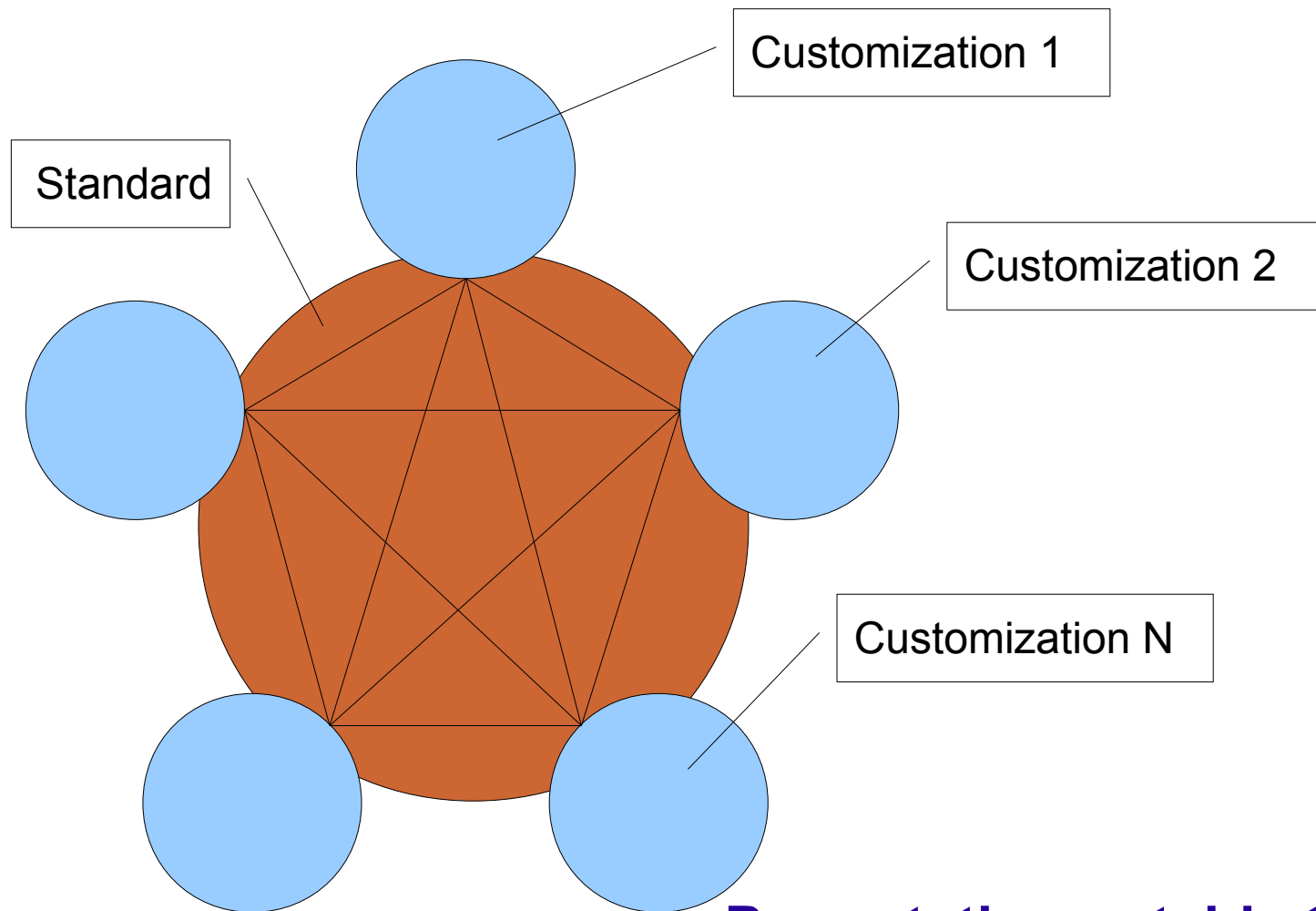
But if you do, your vocabulary won't be standard

So, the choice is all yours

Go with the standard or go it alone



Life with a customized standard vocabulary



Permutation outside the standard

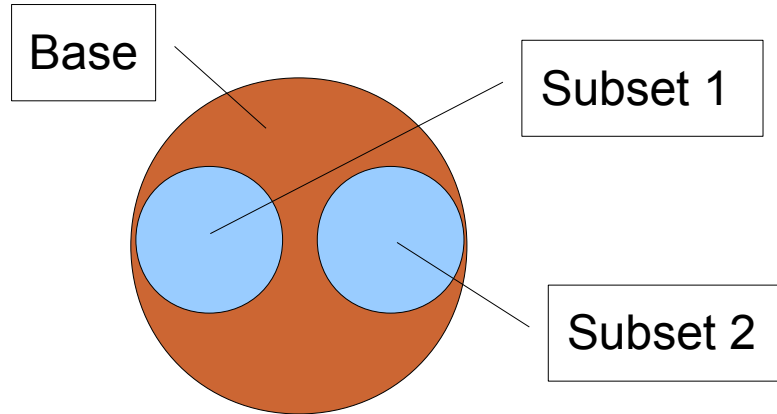
Can we do better?

- **Instead of standards as legislative statutes, something more like common law?**
- **Improve vocabularies through decentralized refinement by adopters?**
- **Adapt to variant requirements?**
- **Tolerate multiple perspectives (represented in variant vocabularies) for the same problem domain?**

Points

- Standards and customizations
- **Vocabulary relations**
- Theoretical alignments
- Dialects
- XHTML Dialects

Subset relations between vocabularies



- **Interchange at the base level**
- **Sharing tools at the base level**
- **TEI pure customizations – distinguished from superset or overlap customizations**

Subset vocabularies by overloading elements

- **microformats**

- By definition, the structure is at least as restrictive
- The semantic is narrower than the overloaded element

```
<div class="geo">  
  <abbr class="latitude">N 37° 24.491</abbr>  
  <abbr class="longitude">W 122° 08.313</abbr></div>
```

- **Can scale to complex vocabularies**

- hAtom expresses Atom as a microformat

Overloading through restrictive substitution

- **Manifest an overload as an element (not an attribute value)**
 - Restriction + substitution groups in XML Schema, specialization in DITA

```
<geo>  
  <latitude>N 37° 24.491</latitude>  
  <longitude>W 122° 08.313</longitude></geo>
```

- **Works only if base element has loose content model**

Base with mixed, recursive content supports any structure

```
<GeoBase>(text | GeoBase)*</GeoBase>
```

Substituting a container of a text value for GeoBase

```
<GeoLocation>(text)*</GeoLocation>
```

Substituting a container / contained structure for GeoBase

```
<Geo>(Latitude, Longitude)</Geo>
```

Advantages of restrictive substitution

- **Often wrap content fragments with special containers**
- **Special content still processable as base content**
 - Ignore the special semantics and restrictions

```
<div>
  <abbr>N 37° 24.4
  <abbr>W 122° 08.
</div>
```

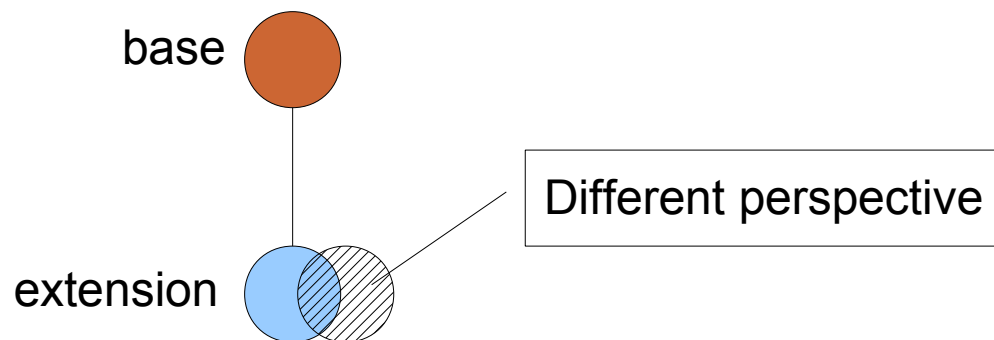
```
<geo>
  <latitude>N 37° 24.491</latitude>
  <longitude>W 122° 08.313</longitude>
</geo>
```

- **Unknown constructs preserved and ignored**
 - No need to read a new doctype to ignore the unknown parts
 - Added properties hidden in a processing instruction

```
<div>
  <?dx:prop url="h
  <?dx:prop max="5
</div>
```

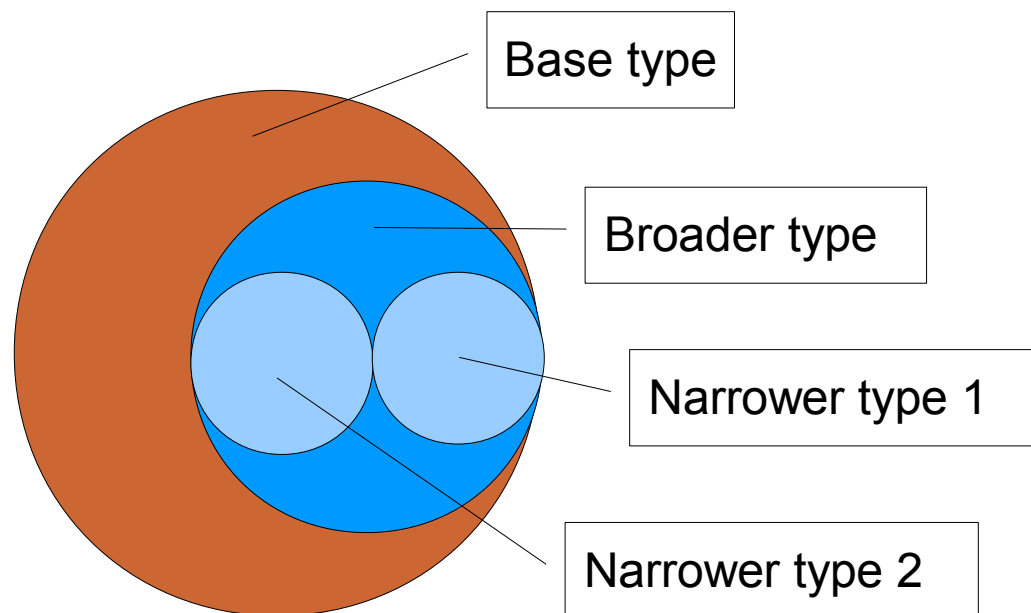
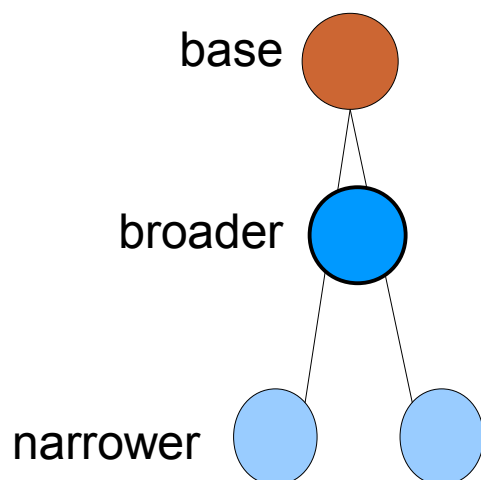
```
<feed
  url="http://slashdot.org/index.rss"
  max="5"/>
```

Limitations of restrictive substitution



- **Variant perspectives on the same thing**
 - Poster child – some users wanted section in DITA task
- **After restriction, no longer loose**
 - Special cases often require more properties

Upward expansion



- **Define new broader vocabulary as minimal superset**
- **Inserted between base and existing narrower vocabulary**
 - For example, broader DITA task with section
- **Allows multiple perspectives with commonality**

Scenarios for upward expansion

- **Adding a locationName property to an existing Geo type**

```
<Geo> (Latitude, Longitude) </Geo>
```

```
<GeoAdd> (LocationName?, Latitude, Longitude) </GeoAdd>
```

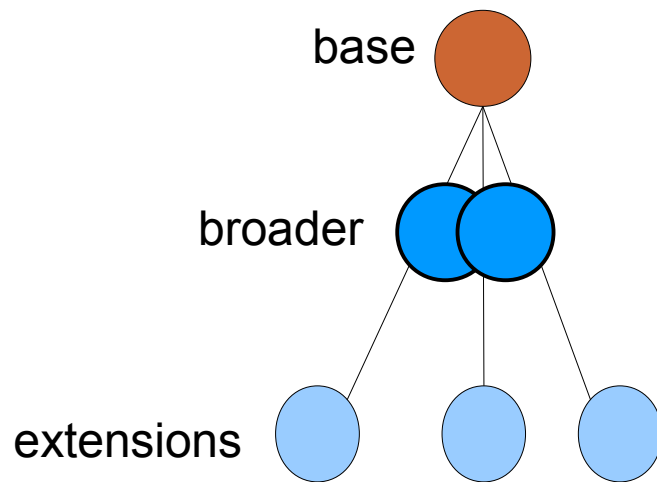
- **Defining interoperability for existing types**

```
<Geo> (Latitude, Longitude) </Geo>
```

```
<Location> (LocLong, Timezone) </Location>
```

```
<GeoInter> (Latitude?, Longitude, Timezone?) </GeoInter>
```

Multiple upward expansions for the same type

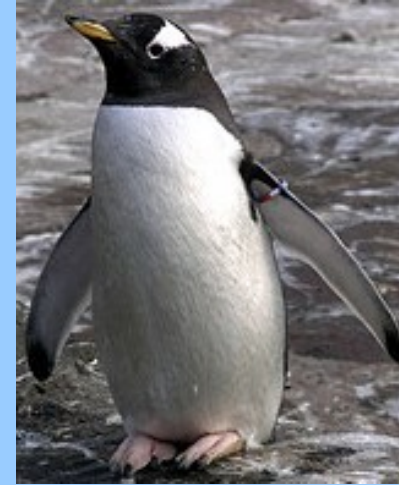


- **Different minimal supersets for different purposes**
- **Different supersets created through concurrent work**

Points

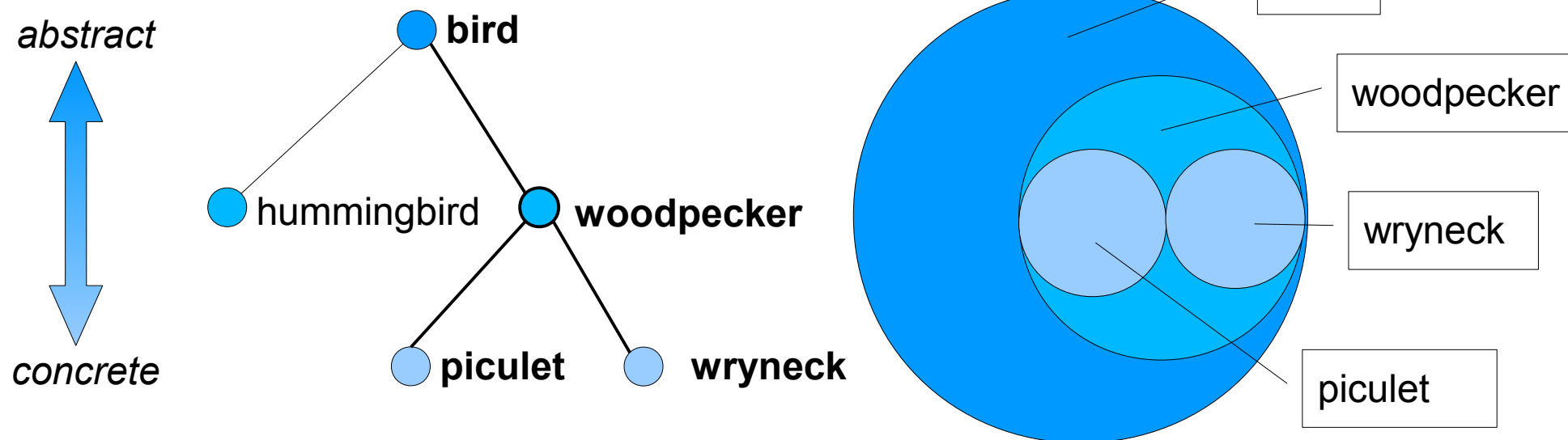
- Standards and customizations
- Vocabulary relations
- **Theoretical alignments**
- Dialects
- XHTML Dialects

Variant perspectives and family resemblance



- **The *bird* category is not defined by a single set of properties**
 - Members of the *bird* category have overlapping properties
- **Properties differ depending on the kind of bird being defined**
 - Flowers consumed, pecking depth, swim speed

Perspectives and prototype theory



- **Recognizable categories have a prototypical image**
 - Abstract categories (*bird*) group prototype categories
 - Specific categories (*piculet*) analyze prototype categories
- **Explains the double bind for customized standards**
 - A vocabulary can't accommodate all of the prototypes

Points

- Standards and customizations
- Vocabulary relations
- Theoretical alignments
- **Dialects**
- XHTML Dialects

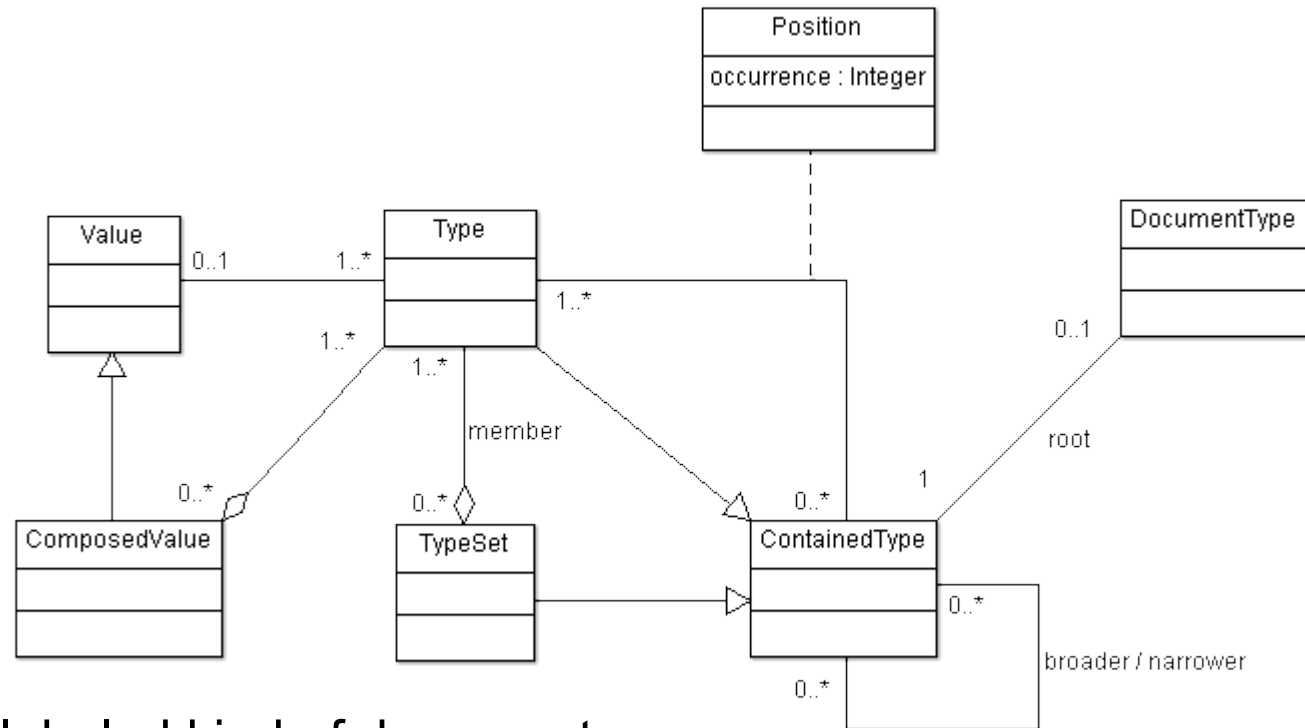
What's needed for variant vocabularies?

- **Extensions of a common base vocabulary**
 - The base vocabulary grounds interpretation of extensions
 - Restrictive substitution and upward expansion for the base
- **Processable representation of vocabularies**
 - Broader-narrower or equivalent relations between types
- **Conversion from narrower to broader vocabularies**

XML Dialects

- **RDF representation of model definitions**
 - Core focus on aligning vocabularies
 - *Not representation of instances*
- **Separation of model and XML instances**
 - Process instances against the model
- **Comparison based on formalized content models**
 - A value (including mixed content) or structures
 - Structures as positions with unique types and cardinality
 - Groups as named types or type sets (choices)

Metamodel



DocumentType – A labeled kind of document.

Type – A labeled kind of value or structure.

Position – An association between a containing type and a contained type.

ContainedType – An abstract class for a type or type set.

TypeSet – An abstract equivalent to a union of types.

Value – A kind of text flow labeled by a type.

ComposedValue – A text flow with parts labeled by types.

RDF representation of a model

- **Traversable and queryable relations between types**

```
<dx:StructureType rdf:about=".../geo">
  <dx:hasBase rdf:resource=".../xhtml/div"/>
  <dx:partOf rdf:resource=".../FlowMix-99679"/>
  <dx:memberOf rdf:resource=".../BlkStructClass-99679"/>
  <dx:containerFor rdf:nodeID=".../geo-longitude"/>
  ...
</dx:StructureType>

<dx:FlowType rdf:about=".../longitude">
  <dx:hasBase rdf:resource=".../xhtml/abbr"/>
  <dx:hasValue rdf:resource=".../xhtml/datatypes/Text"/>
  <dx:containedBy rdf:nodeID=".../geo--longitude"/>
</dx:FlowType>
```


Operations on the model

- **Generating a XML Schema, Relax NG, or DTD**
- **Converting instances to broader document types**
- **Validating inclusions in broader contexts**
- **Role-based matching of instances against types**
- **Analysis of alignment between document types**
 - Assert commonality
 - Identify divergence
 - Generating minimum broader vocabulary for interoperability

Mapping an XML instance to the model

- **Markup as an interface – surface form of the model**
- **Mapping of attribute or element name to type**
- **Selection of attribute or element form for simple types**
- **Optional implicit container elements for flattening**
 - Content must be unique in all contexts
 - Nested groups in the schema

Serializing as a broader vocabulary

- **Special namespaced attributes for original type as CURIE**

```
<div dx:dialect="geo:geo">  
  <abbr dx:dialect="geo:latitude">N 37° 24.491</abbr>  
  <abbr dx:dialect="geo:longitude">W 122° 08.313</abbr>  
</div>
```

- **Processing instructions for added properties**

```
<div dx:dialect="ex:feed">  
  <?dx:prop url="http://slashdot.org/index.rss"?>  
  <?dx:prop max="5"?>  
</div>
```

- **Where base has a role attribute, use modified CURIE**

```
<div class="geo--geo">  
  <abbr class="geo--latitude">N 37° 24.491</abbr>  
  <abbr class="geo--longitude">W 122° 08.313</abbr>  
</div>
```

Points

- Standards and customizations
 - Vocabulary relations
 - Theoretical alignments
 - Dialects
- **XHTML Dialects**

Dialects with XHTML as the base

- **Not perfect but – because of div and span – good enough**
- **Seed designs with familiar vocabulary**
 - Surface <p> and where needed
- **Easy interpretation of extensions**
 - Familiar base elements indicate the kind of content
- **Rendering by converting to the base instead of transforms**
 - Base HTML in browsers, ePub, . . .
- **Broadening as needed as hypermedia for REST**
 - Content negotiation for validating special semantics

Wiki interchange through XHTML dialects

```
{rss:  
  url=http://slashdot.org/index.rss |  
  max=5}
```

Confluence

```
<div class="conf--rss">  
  <?dx:prop url="http://slashdot.org/index.rss"?>  
  <?dx:prop max="5"?></div>
```

```
<div class="doku--rss">  
  <?dx:prop feed=http://slashdot.org/index.rss?>  
  <?dx:prop number=5?></div>
```

DokuWiki

```
{{rss>  
  http://slashdot.org/index.rss  
  5  }}
```

Work in progress (1 of 2)

- **An HTML dialect to define dialect types**

```
<StructureType about="geo:geo">  
  <base ref="html:div"/>  
  <substructure ref="geo:latitude" range="zeroToOne"/>  
  <substructure ref="geo:longitude" range="zeroToOne"/>  
</StructureType>
```

```
<FlowType about="geo:latitude">  
  <base ref="html:abbr"/>  
  <value ref="xhtml:Text"/>  
</FlowType>
```

...

- **Converts to RDF graph**

Work in progress (2 of 2)

- **An HTML dialect for design by delta**

```
<NarrowerDocumentType about="ex:DocumentType">
  <extends ref="html:StrictDocument-1.0"/>
  <vocabulary namespace="http://.../example/" />
  <deletes container="*" contained="html:span"/>
  <appends container="*" contained="html:div"
    ref="geo:geo"/>
</NarrowerDocumentType>
```

...

- **Generates types implied by deltas and default mapping to an XML vocabulary**

Summary

- **Model a problem domain with a cluster of variant vocabularies**
 - Standard seeds the solution space
 - Restrictive substitution to narrow existing types
 - Upward expansion to broaden existing types
- **Interoperability across variant vocabularies**
 - Commonality for different perspectives
- **Benefits and practicality of basing dialects on XHTML**
 - Ease of interpretation
 - Rendering by automated conversion to base

. . . reactions?