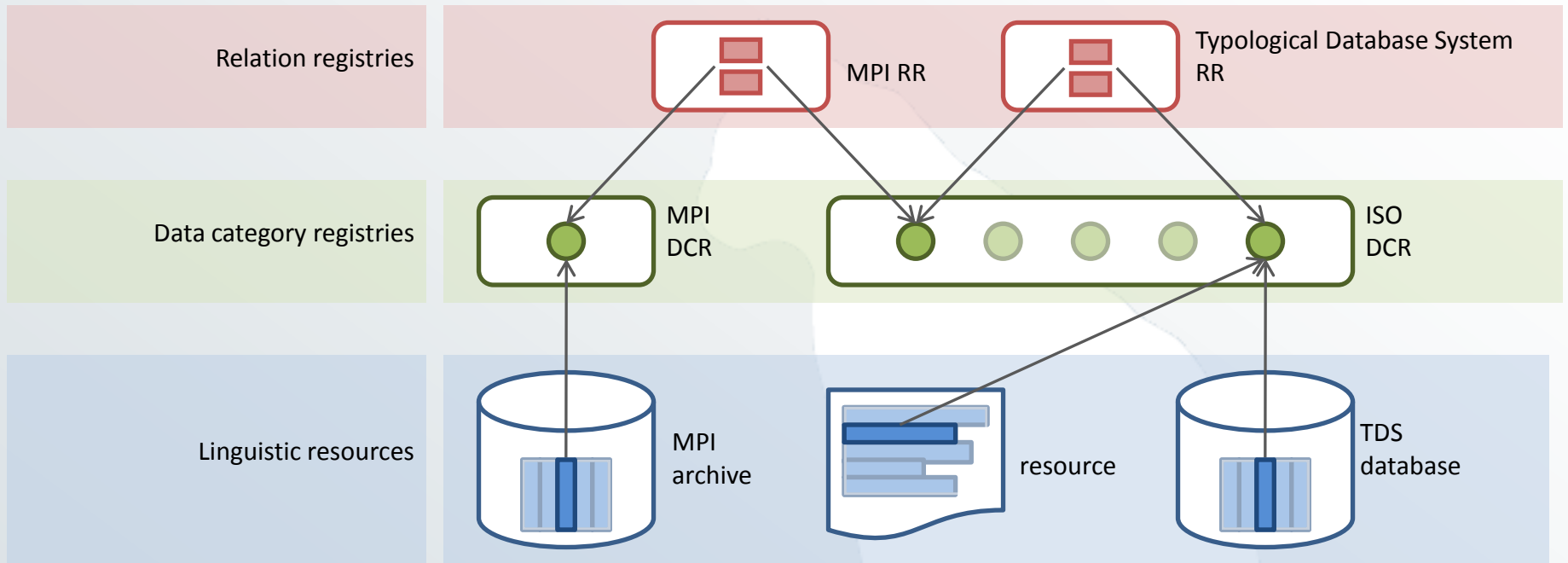


Beyond ISOcat



Vision



How to make semantics explicit?

- Associate data categories with your resources
 - using the PIDs
- Where to put the PIDs?
 - Preferably in a schema
 - Or in the resource itself (redundant)
 - Or in the metadata of the resource (less specific)

What is a schema?

- “comes from the Greek word "σχῆμα" (skhēma), which means *shape*, or more generally, *plan*.” (wikipedia)
- A collection of building blocks and rules on how to combine them into a valid resource
 - XML document:
 - DTD, XML Schema, Relax NG, ...
 - easy; see <http://www.isocat.org/12620/>
 - RDF graph
 - annotation property
 - easy; see <http://www.isocat.org/ns/dcr.rdf>
 - Text document:
 - A grammar
 - Extended Backus–Naur Form (EBNF)
 - ...
 - how to embed Data Category PIDs?
 - ...

XML resource

```
<Imf:lexicon xml:lang="jp" alphabet="ipa">  
  <Imf:entry>  
    <Imf:lemma>  
      <Imf:writtenForm>nihongo</...>  
      ...  
    </...>  
    ...  
  </...>  
  ...  
</...>
```

XML resource

```
<Imf:lexicon xml:lang="jp" alphabet="ipa">
  <Imf:entry>
    <Imf:lemma>
      <Imf:writtenForm
        dcr:datcat="http://www.isocat.org/datcat/...">
        nihongo
      </...>
      ...
    </...>
    ...
  </...>
  ...
</...>
```

XML Relax NG schema

```
<rng:attribute name="alphabet"  
  dcr:datcat="http://www.isocat.org/datcat/...">  
  <rng:value  
    dcr:datcat="http://www.isocat.org/datcat/...">  
    ipa  
  </...>  
  ...  
</...>
```

CGN/DCOI grammar with DC references

<http://lux13.mpi.nl/schemacat/schema/CGN>

(early alpha version)

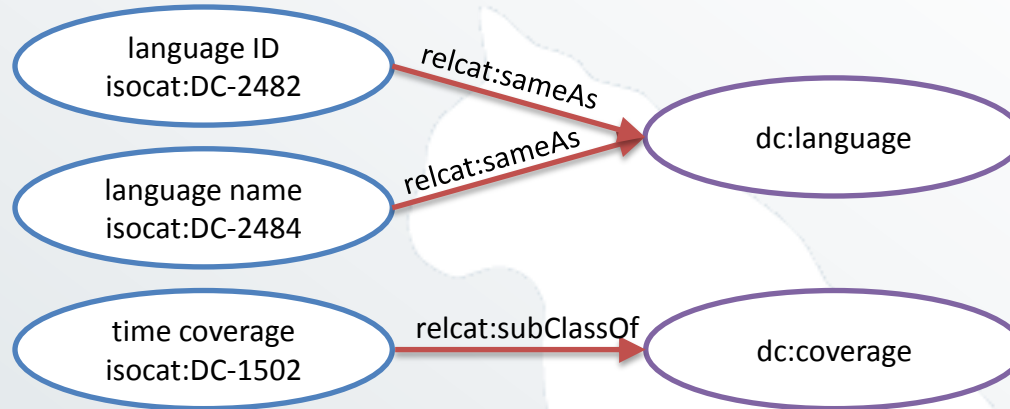
```
(* @dcr:datcat 'N' http://www.isocat.org/datcat/DC-4909 *)
...
tag = 'N',      '(', NTYPE,      ', ', GETAL,      ', ', GRAAD,      ', ',
GENUS,        ', ', NAAMVAL, ')\`
...
(* @dcr:datcat NTYPE http://www.isocat.org/datcat/DC-4908 *)
(* @dcr:datcat 'soortnaam' http://www.isocat.org/datcat/DC-4910 *)
(* @dcr:datcat 'eigennaam' http://www.isocat.org/datcat/DC-4911 *)
NTYPE = 'soortnaam'
      | 'eigennaam'
      ;
...
```


Multiple DCRs?

- Actually we don't need multiple DCRs to have overlapping subsets
 - Overlaps are created due to
 - Data categories are typed, and might not have the type you need
 - *POS field (closed DC) of the lexical entry "walk" gets the value 'verb' (simple DC)*
 - » *PoS = 'verb'*
 - *Verb (open DC) feature of a feature structure gets the value "walk"*
 - » *Verb = 'walk'*
 - External sets are imported just as they are
 - NKJP, GOLD, STTS, ...
 - Only some take the effort to also provide mappings
 - There might be very fine differences between your data category and an existing one, and the owner doesn't want to adapt
- Still we would like to know that these data categories are the same or almost the same!

Relation Registry - RELcat

- The Relation Registry stores user specific sets of relations:



<http://lux13.mpi.nl/relcat/>

(alpha version)

‘RELcat: a Relation Registry for ISOCat Data Categories’

poster and demo at LREC 2012

Session: P40 - Knowledge and Ontologies; Poster Area (2)

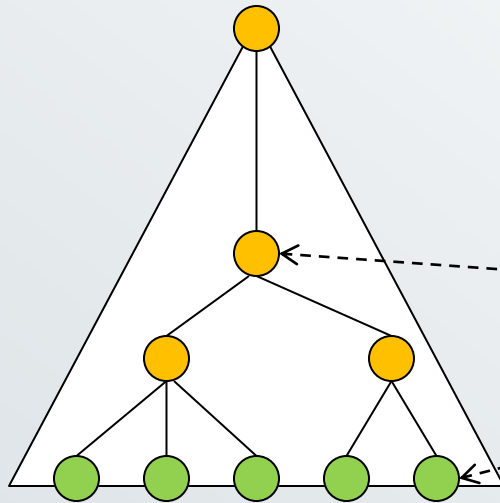
Friday 25 May, 11:45 – 13:25

Another new kitten: SCHEMAcat

- Resource schemata of any type should be stored somewhere persistently
 - Get a PID
- These schemata are preferably annotated with data categories
 - SCHEMAcat → ISOcat
- These data categories will then have (typed) relationships among each other
 - SCHEMAcat → RELcat
- Status: very early alpha, but some schemata are already available

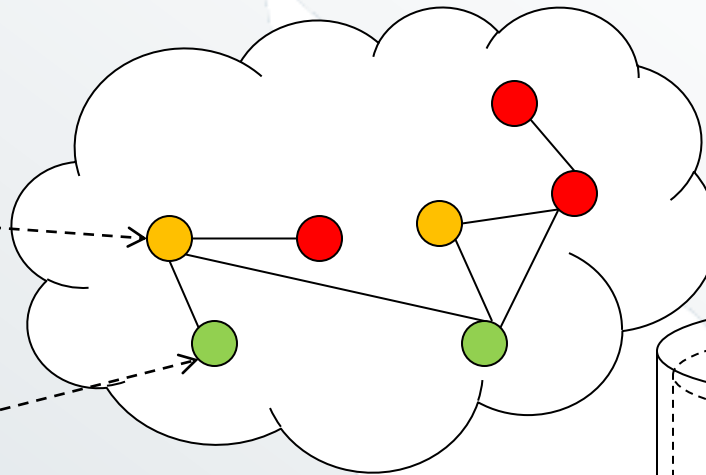
A whole litter!




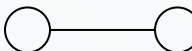
Linguistic resource (schema)

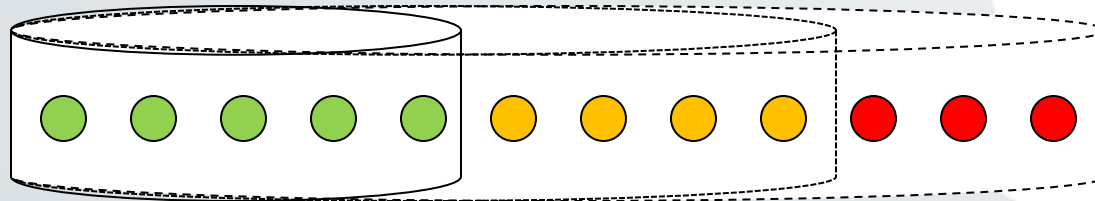


Schema Registry - SCHEMAcat

Linguistic knowledge base

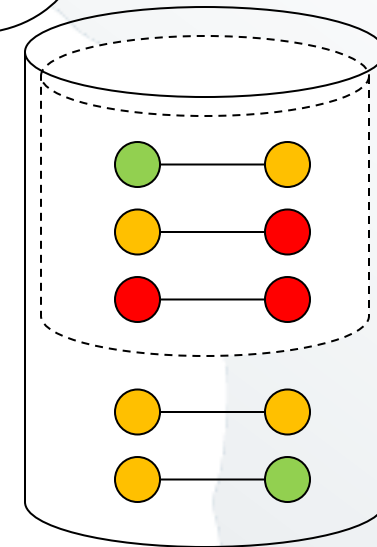


-  Data categories
-  Containers
-  Concepts
-  Relation



Data Category Registry - ISOcat

Concept Registry



Relation Registry - RELcat