

Cornetto-LFM-RDF Clarin-NL project

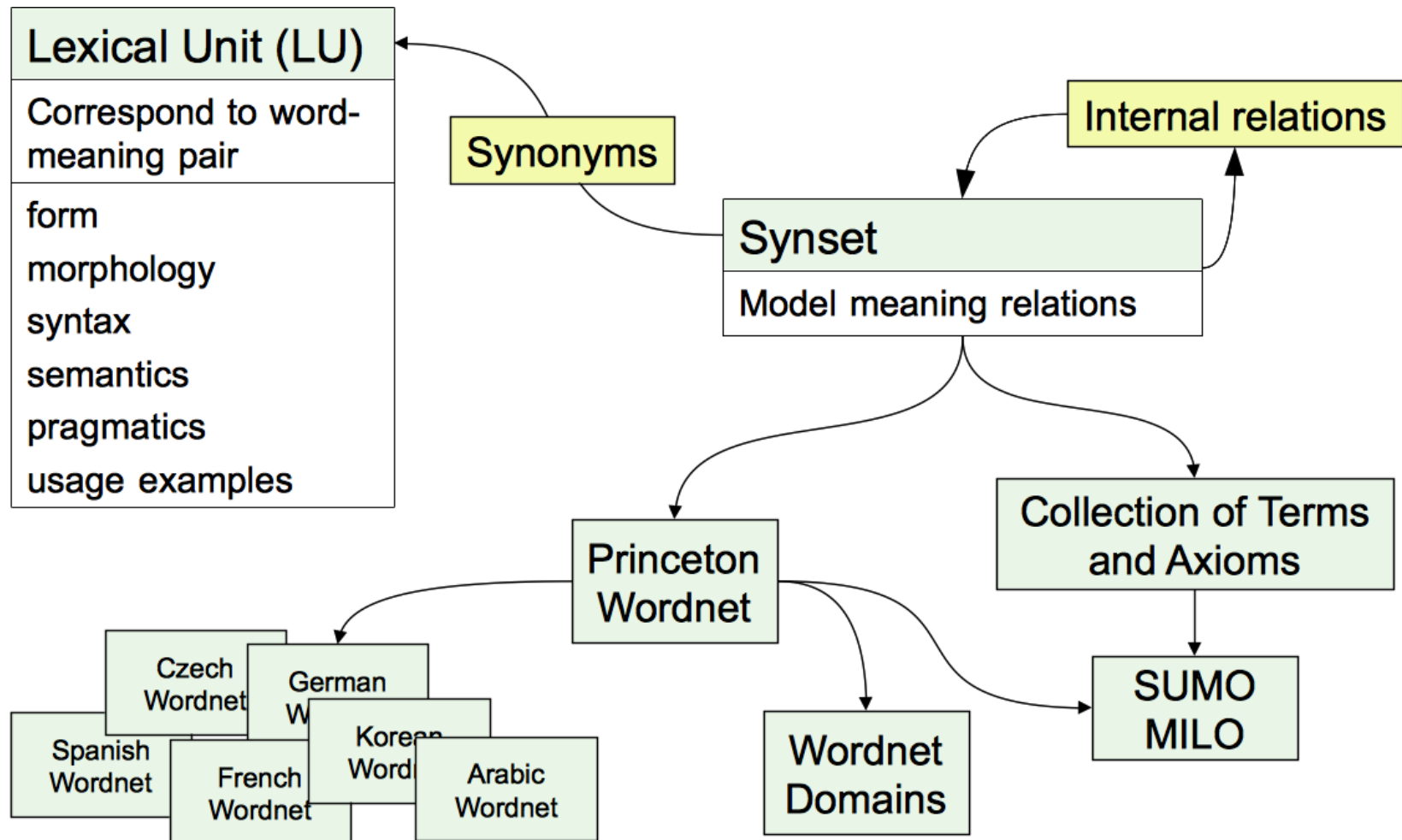
Piek Vossen, Faculteit der Letteren, VU
Jacco van Ossenbruggen, Faculteit Exacte Wetenschappen, VU
Remco van Veenendaal, Bob Boelhouwer, TST centrale



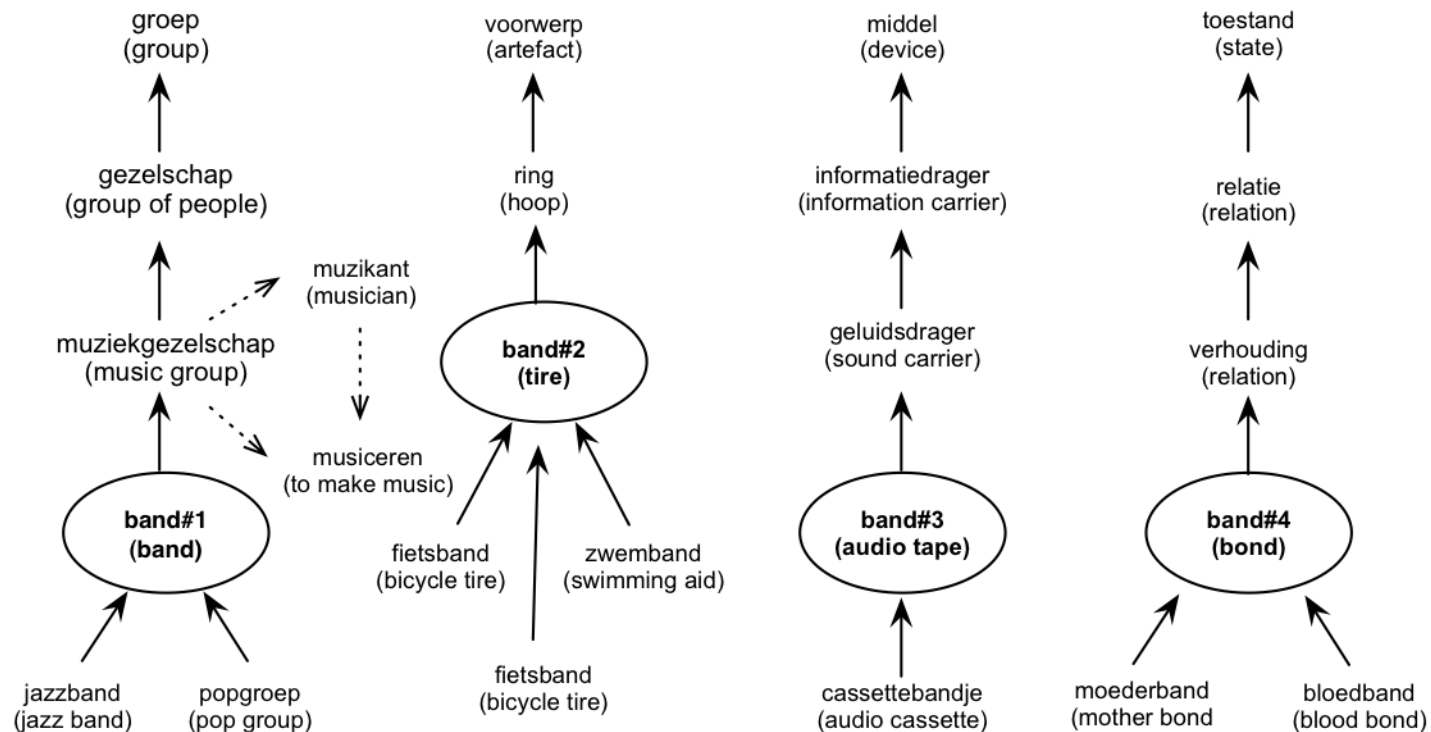
Goals

- Make the Cornetto database available in standardized format on a stable Clarin-server
 - Wordnet-LMF and LMF
 - Wordnet-RDF
- Provide an easy to use demonstrator to access the database

Cornetto data structures



Combinatorics	Combinatorics	Combinatorics	Combinatorics
in een band spelen to play in a band	de band oppompen to inflate a tire	de band starten to start a tape	een goede/sterke band a good/strong bond
een band oprichten to start a band	een band plakken to fix a tire	de band afspelen to play from a tape	de banden verbreken to break all bonds
SUMO: +, , MusicalGroup	SUMO: +, , Artifact	SUMO: +, , Device	SUMO: +, , Relation
WN-domain: music	WN-domain: transport	WN-domain: music	WN-domain: factotum



Cornetto viewer

The image displays three side-by-side windows from the Cornetto viewer, each showing different levels of detail for the word 'water'.

Cornetto Synset: This window shows a list of synsets for 'water'. The selected synset is `d_n-20623 [noun] water:1, majem:1 SCHEIK`. Below the list, there are tabs for 'Concept', 'Relations', 'Tree', 'Revtree', 'Edit', and 'Xml'. The 'Concept' tab is active, showing the following information:

- C_SY_ID:** `d_n-20623`
- Synonyms:** `majem:1/d_n-166370, water:1/r_n-42415`
- PoS specific:** `NOUN_HET`
- Definition:** `kleur- en smaakloze vloeistof waarvan de moleculen uit 2 atomen waterstof en 1 atoom zuurstof bestaan`
- Differentiae:** `0`
- VLIS domains:** `SCHEIK`
- SUMO:** `(+, , Water)`
- Relations:**
 - `-->> [HAS_HYPERONYM] vloeistof:1/d_n-20474`
 - `<<-- [HAS_HYPONYM] regenwater:1/d_n-19186`

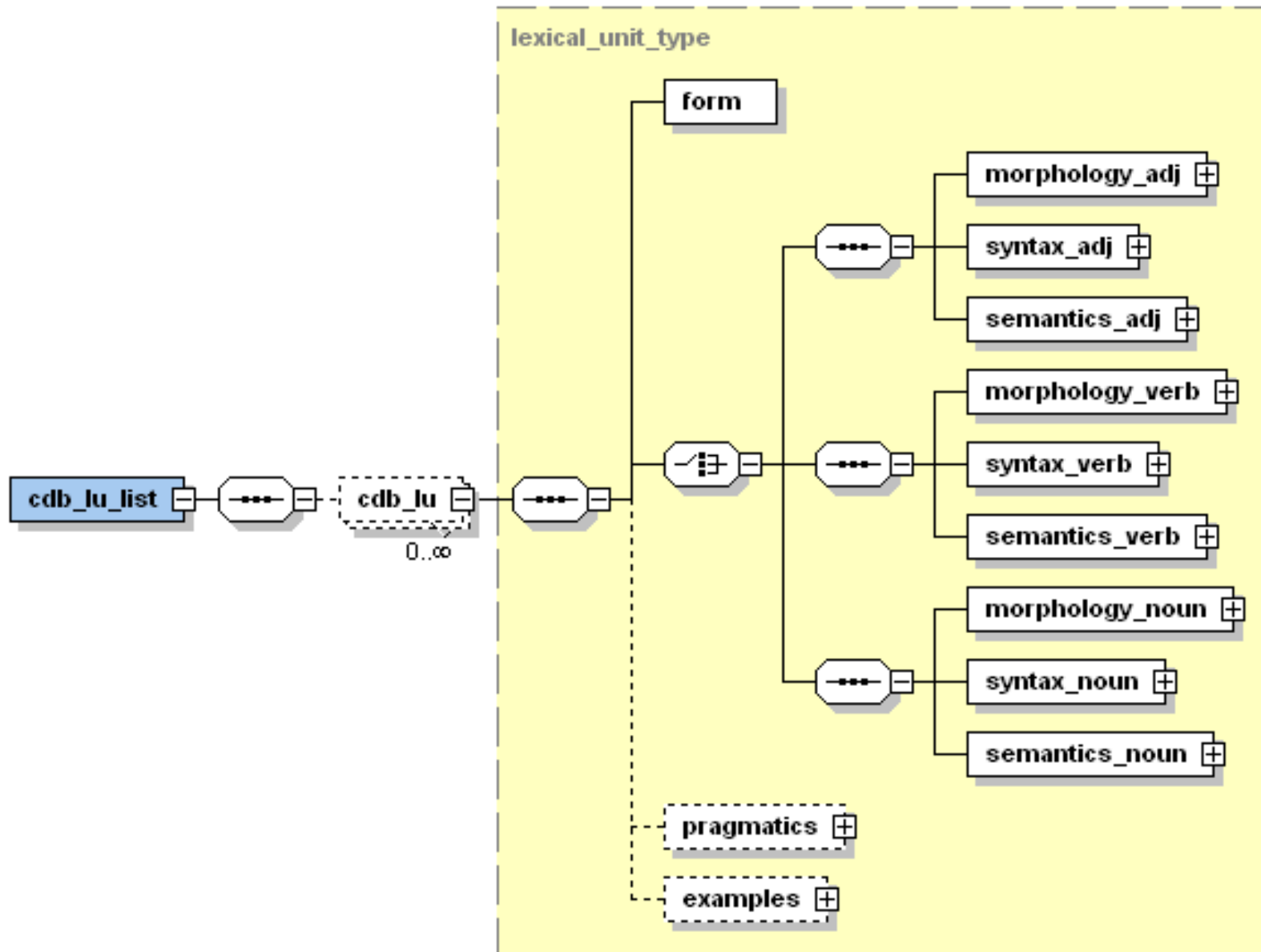
Cornetto LU: This window shows a list of lexical units for 'water'. The selected LU is `r_n-42415 noun, water:1, vocht`. Below the list, there are tabs for 'Form', 'Examples', 'Edit', and 'Xml'. The 'Form' tab is active, showing the following information:

- C_LU_ID:** `r_n-42415`
- C_SEQ_NR:** `1`
- [noun] water**
- Morphology:** `type: simpmorph`
- Syntax:** `gender: n article: het`
- Semantics:**
 - `reference: common`
 - `countability: uncount`
 - `type: substance`
 - `subclass: natuurlijke stof`
 - `resume: vocht`
 - `definition: XVD: die helder en`

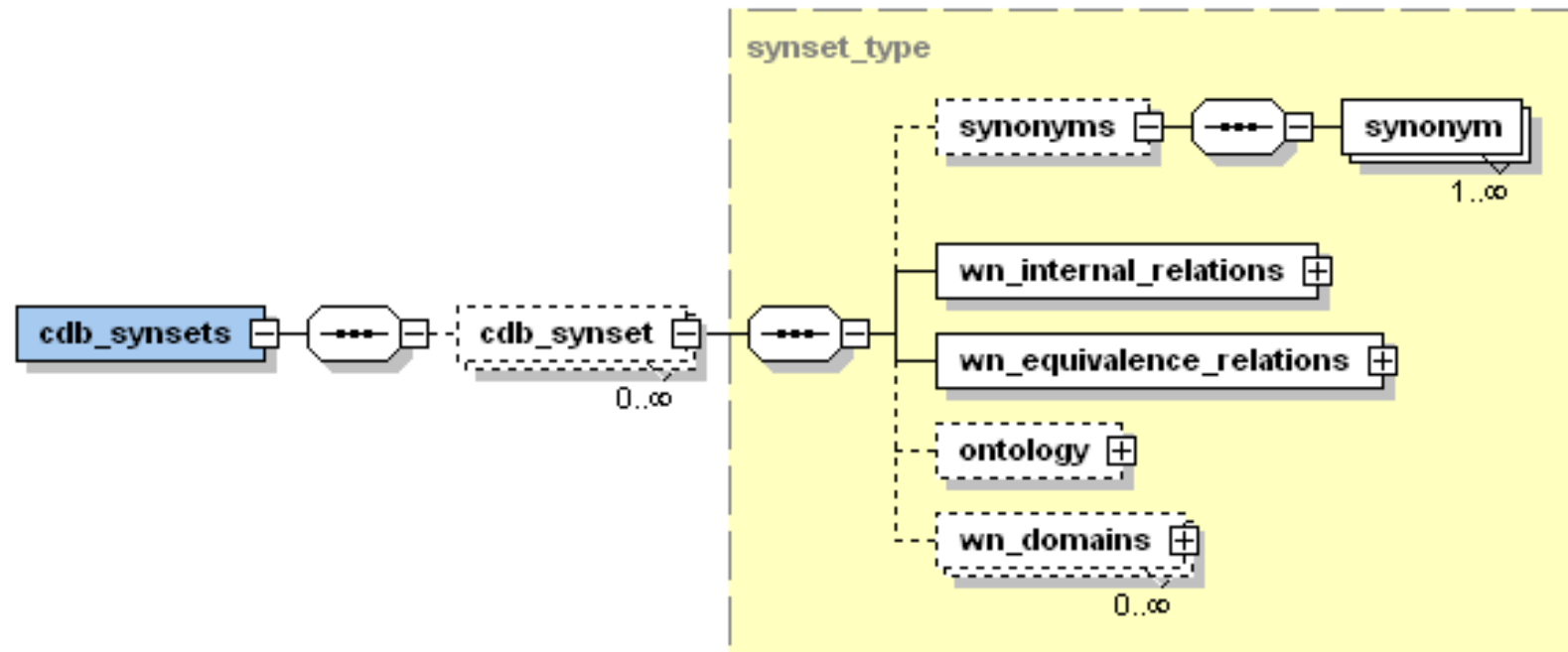
Cornetto ID: This window shows the internal ID structure for 'water'. It includes a search bar and a list of IDs. The selected ID is `water:1, NOUN, LU=r_n-42415, SY=d_n-20623, MC`. Below the list, there are tabs for 'Preview', 'Edit', and 'Xml'. The 'Preview' tab is active, showing the following information:

- CID_ID:** `156793`
- SEQ_NR:** `1`
- Form:** `water`
- PoS:** `NOUN`
- Score:**
- Status:** `manual`
- LU ID:** `r_n-42415`
- Synset ID:** `d_n-20623`

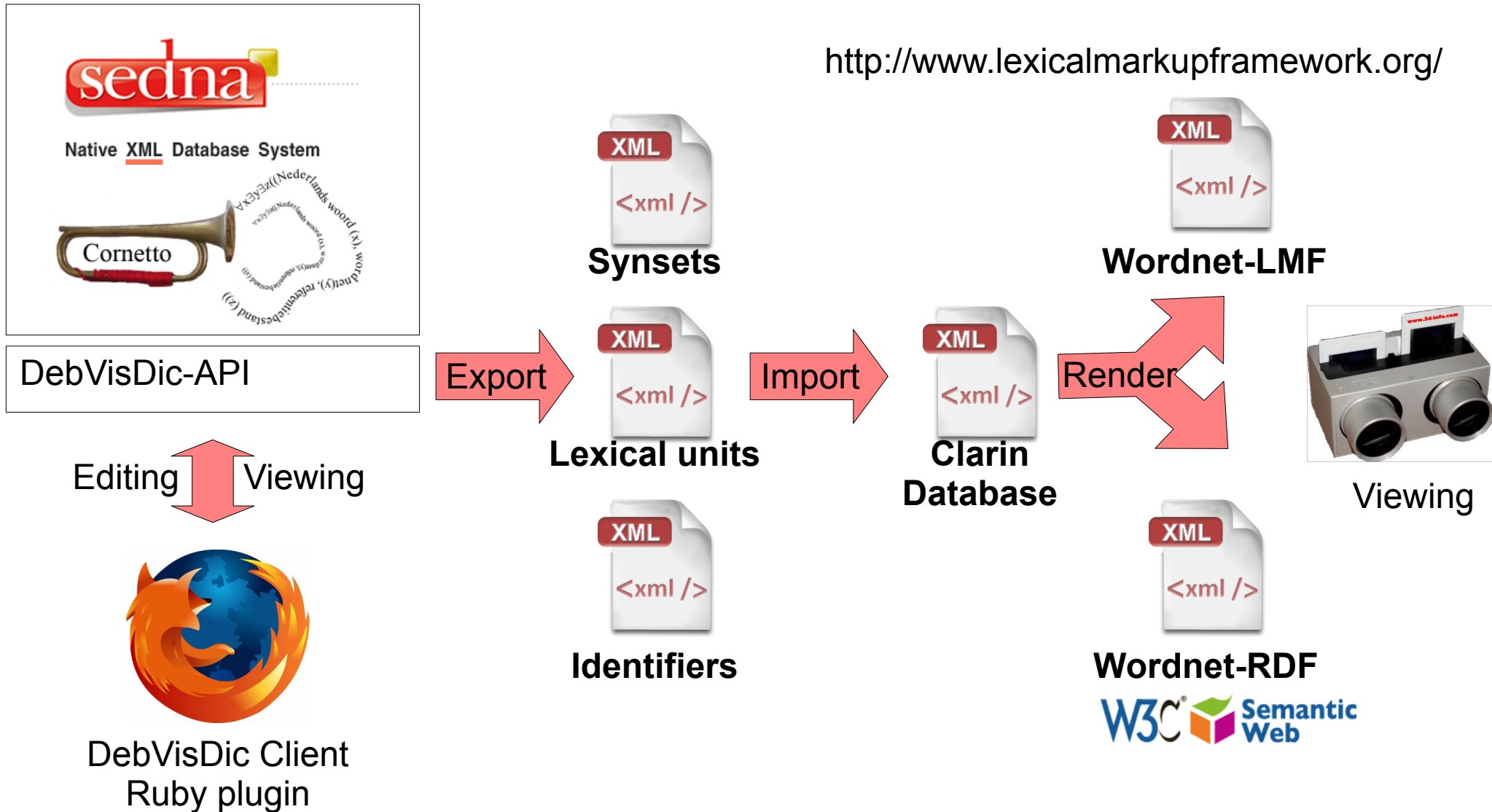
Lexical Unit Schema



Synset schema



Cornetto-LMF-RDF



Id	Target date	Type	Description	Who
D1	M6, Oct 2012	Document	Metadata mappings LMF, WordNet-LMF, SKOS and RDF (XSLT convertors), proposal for ISOCAT extension for semantic resources	VUA-LET-CL, VUA-SCI-SSS, INL-SSS
D2	M7 Nov 2012	Metadata	Metadata of the resource (CMDI or IMDI format, XSLT convertors)	INL-SP
D3	M3 Jul 2012	Document	Demonstration scenario for non-technical humanities researchers	VUA-LET-CL
D4	M7 Nov 2012	Document	Design of the demonstrator: database, API and web-client	INL-SP
D5	M8 Dec 2012	Milestone	Metadata available on accessible INL server	INL-SP
D6	M10 Feb 2012	Data and software	Curated resource available in XML and as back-end database on the INL server with API	INL-SP
D7	M11 Mar 2012	Milestone	Metadata successfully harvested	INL-SSS
D8	M12 Apr 2012	Data	Web-client	INL-SP
D9	M12 Apr 2012	Document	Documentation of curated resource	VUA-LET-CL
D10	M12 Apr 2012	Document	Documentation of the demonstrator (database, API and webclient)	INL-SP
D11	M12 Apr 2012	Document	Requirements and desiderata for the CLARIN infrastructure, justified by the findings of the project	VUA-LET-CL

Sheet1

		VLET	VSCI	TST	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
D1	Metadata mappings LMF, WordNet-LMF, SKOS and RDF & XSLT convertors	5	2	2												
D2	Metadata of the resource (CMDI or IMDI format) & XSLT convertors			1												
D3	Demonstration scenario for non-technical humanities researchers	1														
D4	Design of the demonstrator: database, API and web-client	1		1												
D5	Metadata available on accessible INL server			0,5												
D6	Curated resource available in XML and as back-end database on the INL server with API			2												
D7	Metadata successfully harvested			0,5												
D8	Web-client			2												
D9	Documentation of curated resource	1														
D10	Documentation of the demonstrator (database, API and webclient)			1												
D10	Requirements and desiderata for the CLARIN infrastructure, justified by the findings of the project	2														
D11	TOTAL PERSON MONTHS	10	2	10												

First Phase

- Project start date: 1st of May 2012
- Kick off meeting in May:
 - Persistent identifiers for concepts
 - Lexical relations (word-to-word)
 - Conceptual relations (wordnet-like)
 - Lexical unit data Cornetto in LMF and RDF
 - LMF and ontologies: LexInfo and Wordnet-LMF
 - IPR
 - End users and user scenarios