

CLARIN

Common Language Resources and Technology Infrastructure



MIMORE

Demonstratieproject CLARIN-NL

Folkert de Vriend

Folkert.de.Vriend@meertens.knaw.nl

Meertens Instituut, Amsterdam

26/08/2010

Overzicht

- I. Resources en onderzoeksgereedschappen in MIMORE
- II. Inbedding van resources en onderzoeksgereedschappen in CLARIN-architectuur

I: Resources en
onderzoeksgereedschappen in
MIMORE

Algemeen

- MIMORE = **M**icrocomparative **M**orphosyntax **R**esearch Tool
- Doel is de ontwikkeling van CLARIN-compatibel onderzoeksgereedschap waarmee taalkundigen (morfo-)syntactische verschijnselen uit de resources GTRP, DiDDD en DynaSAND gecombineerd kunnen interpreteren.

Resources in MIMORE

- GTRP (Goeman-Taeldeman-Van-Reenen-project)
 - Morfologische variatie op **woordniveau**: *boekje* vs. *boekske*
- DiDDD (Diversity in Dutch DP Design)
 - Syntactische variatie op het niveau van **zelfstandig-naamwoordgroepen**: *hem zijn boek* vs. *zijn boek*
- DynaSAND (Dynamische Syntactische Atlas van de Nederlandse Dialecten)
 - Syntactische variatie op **zinsniveau**. Bijv. variatie in woordvolgorde op het einde van bijzinnen: *(dat hij) moet kunnen zwemmen* vs. *zwemmen kunnen moet*.

Voorbeeld van een (morfo-)syntactisch vraagstuk

Zijn voornaamwoorden van de tweede persoon (*jouw, jullie*) misschien geen onanalyseerbare **woorden**, maar eerder **woordgroepen**, die bestaan uit twee of meer woorden of morfemen in een hiërarchische structuur (zoals bij zelfstandig-naamwoordgroepen)?

Hoe helpt MIMORE?

Met de MIMORE-gereedschappen wordt het mogelijk om bewijzen te zoeken voor een dergelijke aanname.

De variatie die voornaamwoorden van de tweede persoon vertonen **in isolatie** (GTRP) kan met MIMORE namelijk gemakkelijk vergeleken worden met de variatie die ze vertonen in verschillende **syntactische contexten**:

- binnen zinnen (DynaSAND)
- binnen zelfstandig-naamwoordgroepen (DiDDD)

Voorbeeld DiDDD: “jouw opa” versus “jullie opa”:

Vorstenbosch: <i>ouw opa</i>
Zoutleeuw: <i>oere bompa</i>
Axel: <i>joen opa</i>
...

Vorstenbosch: <i>hullie én opa</i>
Zoutleeuw: <i>oerlingze bompa</i>
Axel: <i>judder opa</i>
...

Gereedschappen in ontwikkeling

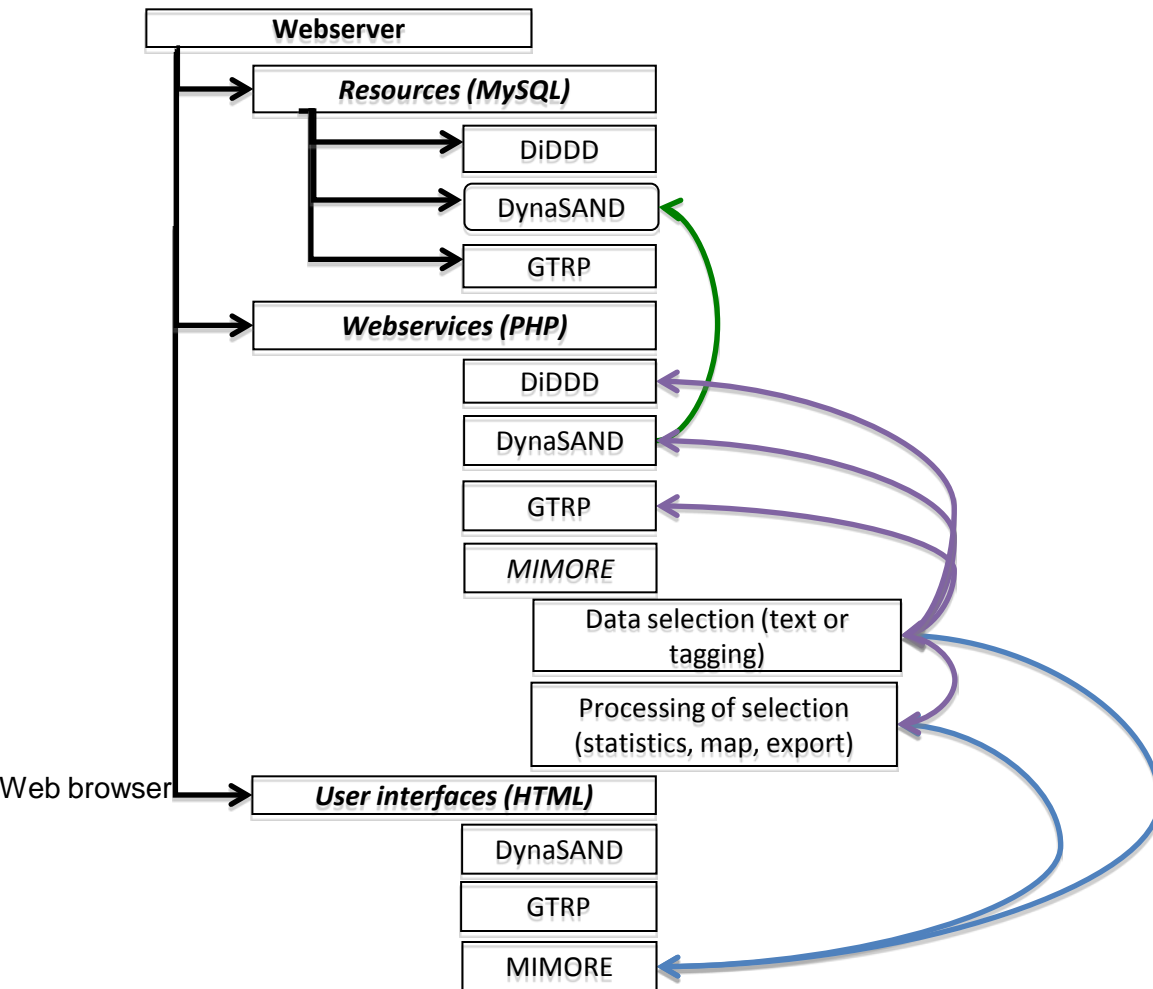
Gecombineerd zoeken

- Op basis van tekst in glossen
- Op basis van woordsoorten

Verwerking van zoekresultaat

- Kaart
- Statistiek
- Export

Onderliggende architectuur



Implementatie webservices and UI

- Zoeken: [testomgeving](#)
- Verwerking zoekresultaten: to do.

Benodigde bewerkingen van resources

POS-tagging

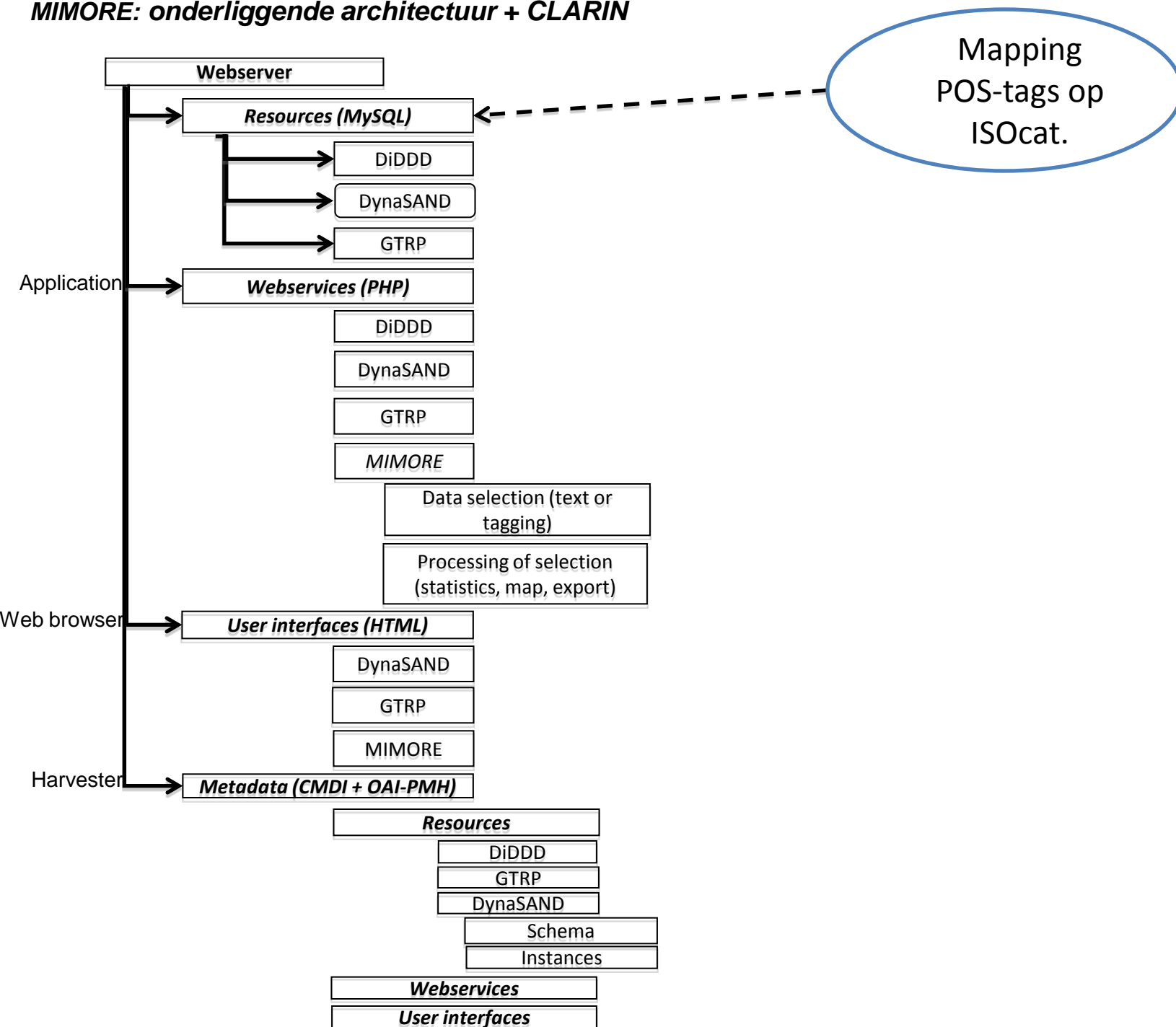
- DynaSAND is inmiddels geheel voorzien van POS-tagging.
- GTRP en DiDDD worden automatisch van POS-tags voorzien en vervolgens handmatig gecorrigeerd.
- Voor alle drie is dezelfde set gebruikt.

Oplijning

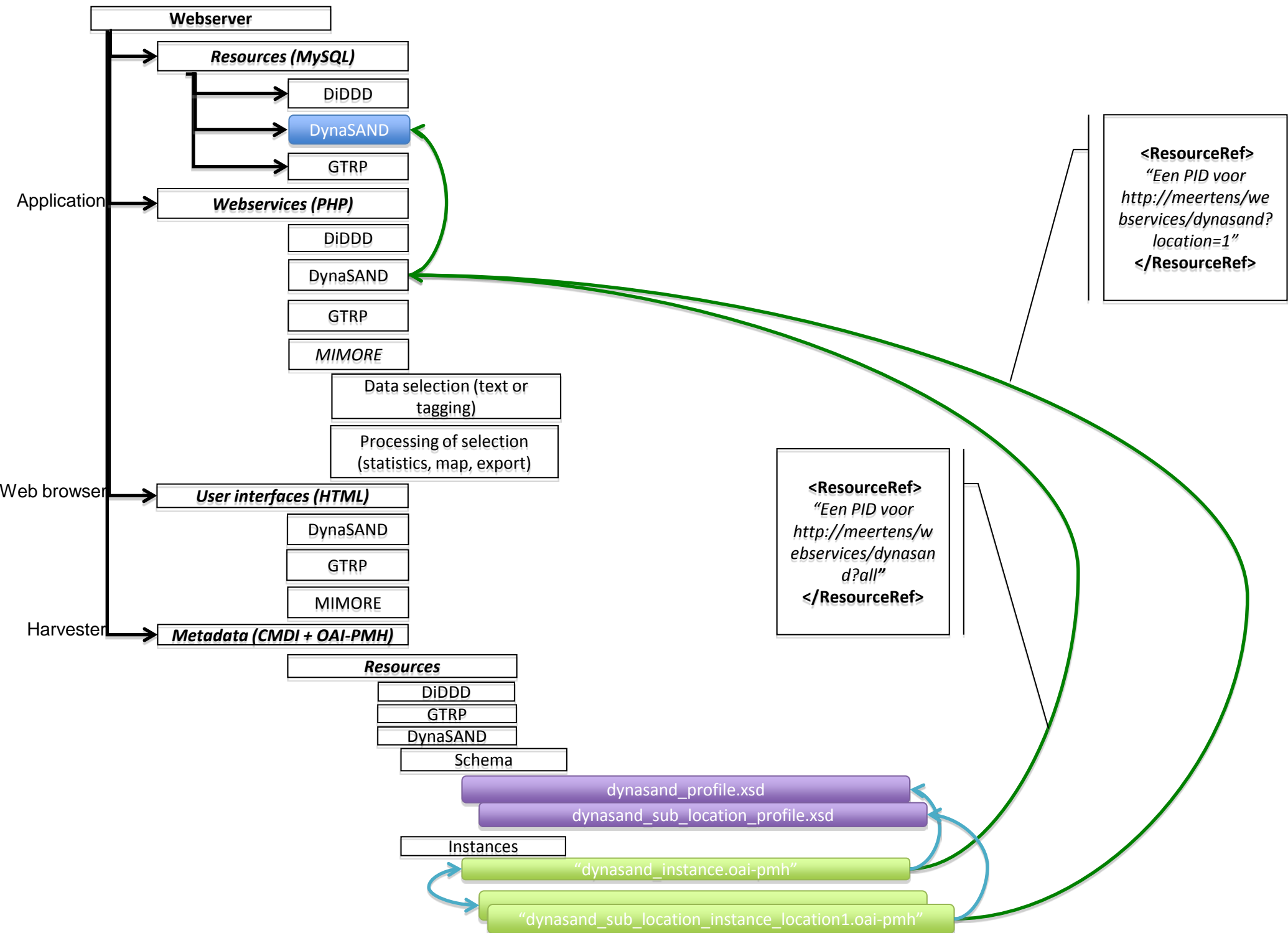
- De opnames van de DynaSAND zijn al opgelijnd.
- Voor DiDDD semiautomatische procedure met PRAAT
- We bekijken nog wat haalbaar is voor GTRP.

II: Inbedding van resources en onderzoeksgereedschappen in CLARIN-architectuur

MIMORE: onderliggende architectuur + CLARIN

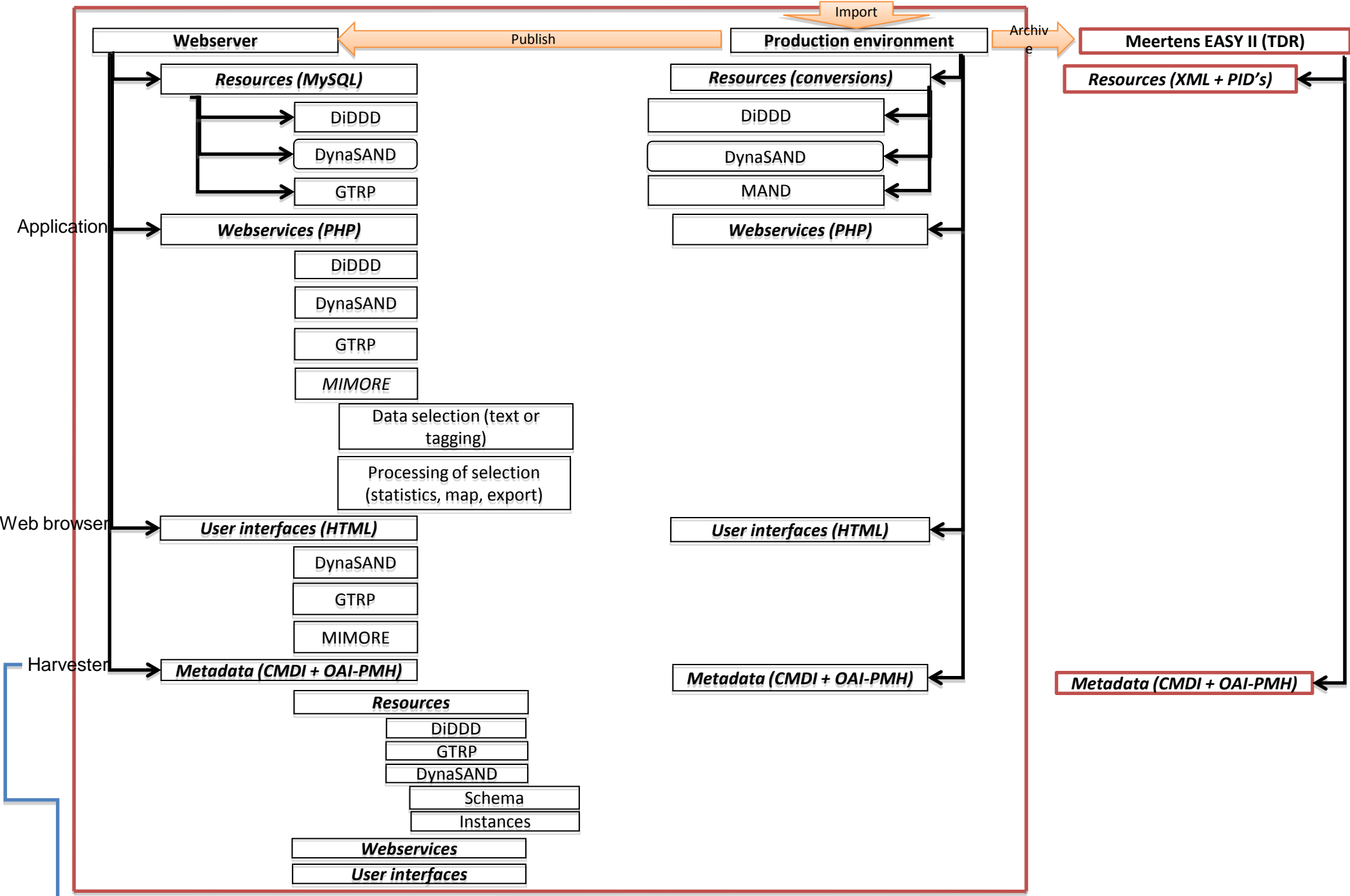


MIMORE: metadata voor resource DynaSAND



MIMORE: inbedding in breder CLARIN-context

Source: Researchers editing resources and metadata



CLARIN metadata repository

Metadata search service

Metadata browsing service

Tijd en mens

- Looptijd project: 1 april tot en met 31 december 2010
- Betrokkenen:
 - Sjef Barbiers (Projectcoördinator, Meertens Instituut)
 - Norbert Corver (Universiteit Utrecht, UiL-OTS)
 - Jan Pieter Kunst (Meertens Instituut)
 - Matthijs Brouwer (Meertens Instituut)
 - Franca Wesseling (Meertens Instituut)
 - Folkert de Vriend (Meertens Instituut)

Dank