

# Bridging the LAPPS Grid and CLARIN

Erhard Hinrichs\*, Nancy Ide\*\*, James Pustejovsky†, Jan Hajič‡, Marie Hinrichs\*, Mohammad Fazleh Elahi\*, Keith Suderman\*\*, Marc Verhagen†, Kyeongmin Rim†, Pavel Straňák‡, Jozef Mišutka‡

\*University of Tübingen, \*\*Vassar College, †Brandeis University, ‡Charles University



CLARIN



LAPPS Grid



## User Authentication & Identification

### CLARIN

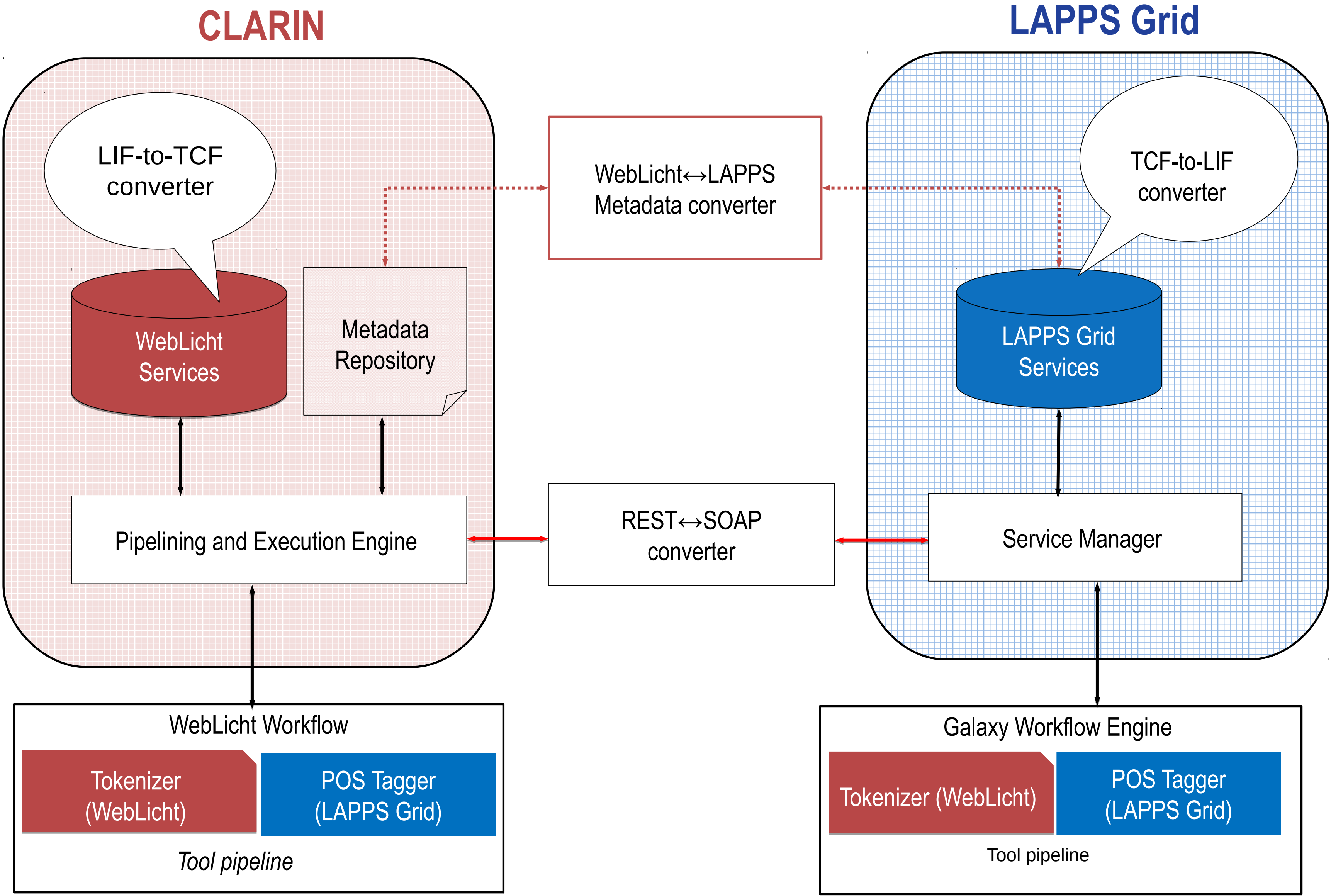
- For academic use only
- User is authenticated via identity providers (IdP's) belonging to EU national academic identity federations

- Establish a joint single sign-on user authentication and authorization mechanism
- ...
- ...

### LAPPS Grid

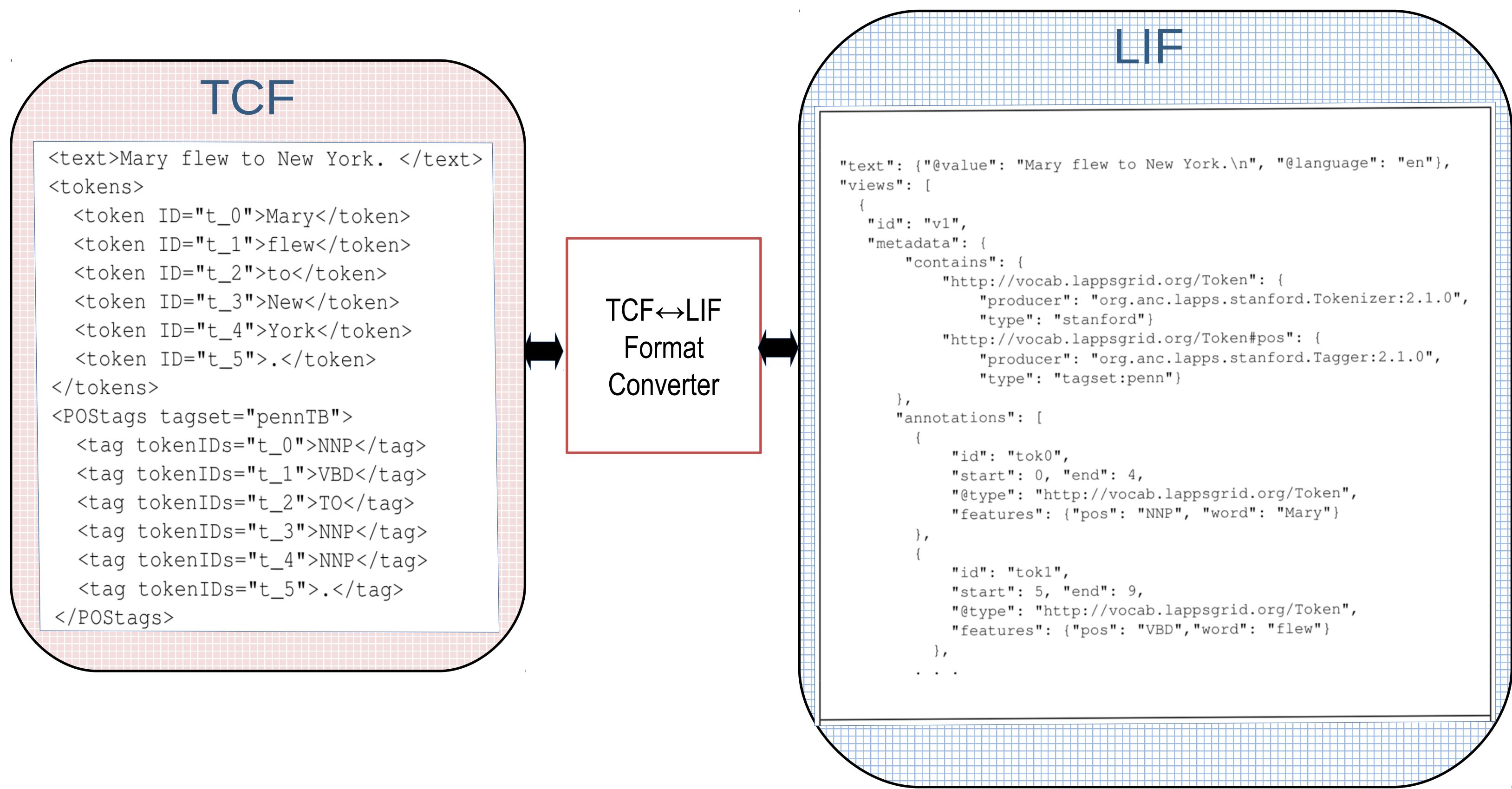
- Open for all users
- Users must register to get a workspace
- Click-through licenses for access to secured data

- WebLicht is used to create, execute, and visualize NLP processing pipelines
- Consume and produce data in Text Corpus Format(TCF)
- Webservices invoked using the RESTful API

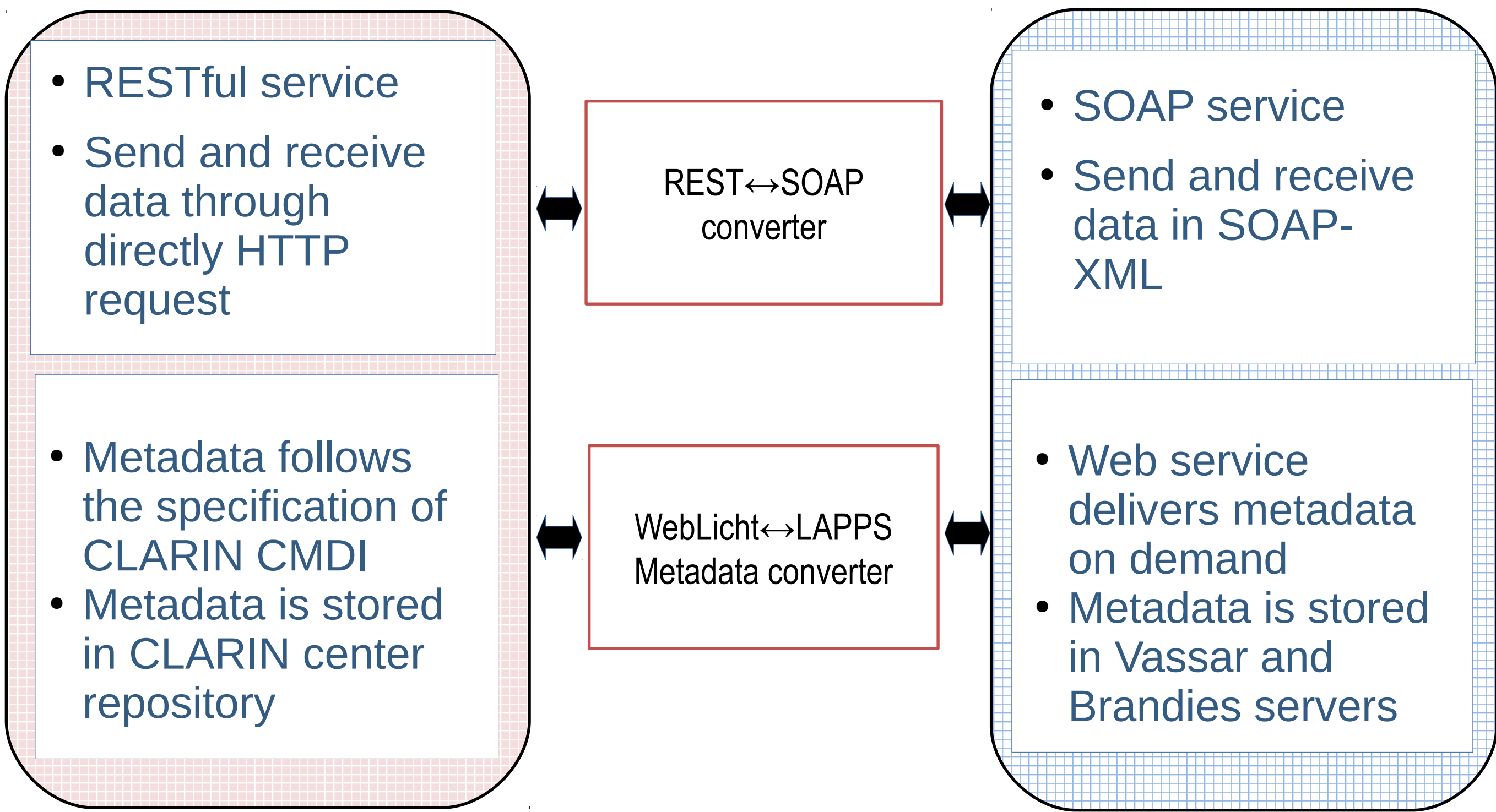


- Language tools and resources are developed in Galaxy framework
- Consume and produce data in LAPPSS Interchange Format(LIF)
- Webservices invoked using the SOAP protocol

## Syntactic Interoperability



## Communication Protocol & Metadata Conformance



## Semantic Interoperability

TCF layer	WSEV equivalent
token	Simple mapping to Token
sentence	Simple mapping to Sentence
lemma	Map to Token#lemma
POSTag	Map to Token#pos
parsing	Simple mapping to PhraseStructure and Constituent. Requires conversion of the XML tree structure to an explicit directed graph.
depparsing	DependencyStructure, Dependency
namedEntities	Simple mapping to NamedEntity
references	Map to Coreference. Add features to correspond to TCF's heads, type, rel, etc.
textstructure	Paragraph and Sentence. TCF's page element has no equivalent in WSEV.
synonymy	No equivalent
matches	No equivalent
wordSplittings	No equivalent
geo	No equivalent
phonetics	No equivalent
orthography	No equivalent
wsd	No equivalent

TCF Element	WSEV equivalent
parsing#tagset	PhraseStructure@categorySet
parse	PhraseStructure
parse#ID	PhraseStructure#id
constituent	Constituent
constituent#ID	Constituent#id
constituent#cat	Constituent#category
constituent#edge	No mapping
constituent#tokenIDs	Constituent#children
cref	No mapping
cref#constID	No mapping
cref#edge	No mapping

TCF layer	WSEV equivalent
token	Simple mapping to Token
sentence	Simple mapping to Sentence
lemma	Map to Token#lemma
POSTag	Map to Token#pos
parsing	Simple mapping to PhraseStructure and Constituent. Requires conversion of the XML tree structure to an explicit directed graph.
depparsing	DependencyStructure, Dependency
namedEntities	Simple mapping to NamedEntity
references	Map to Coreference. Add features to correspond to TCF's heads, type, rel, etc.
textstructure	Paragraph and Sentence. TCF's page element has no equivalent in WSEV.
synonymy	No equivalent
matches	No equivalent
wordSplittings	No equivalent
geo	No equivalent
phonetics	No equivalent
orthography	No equivalent
wsd	No equivalent