

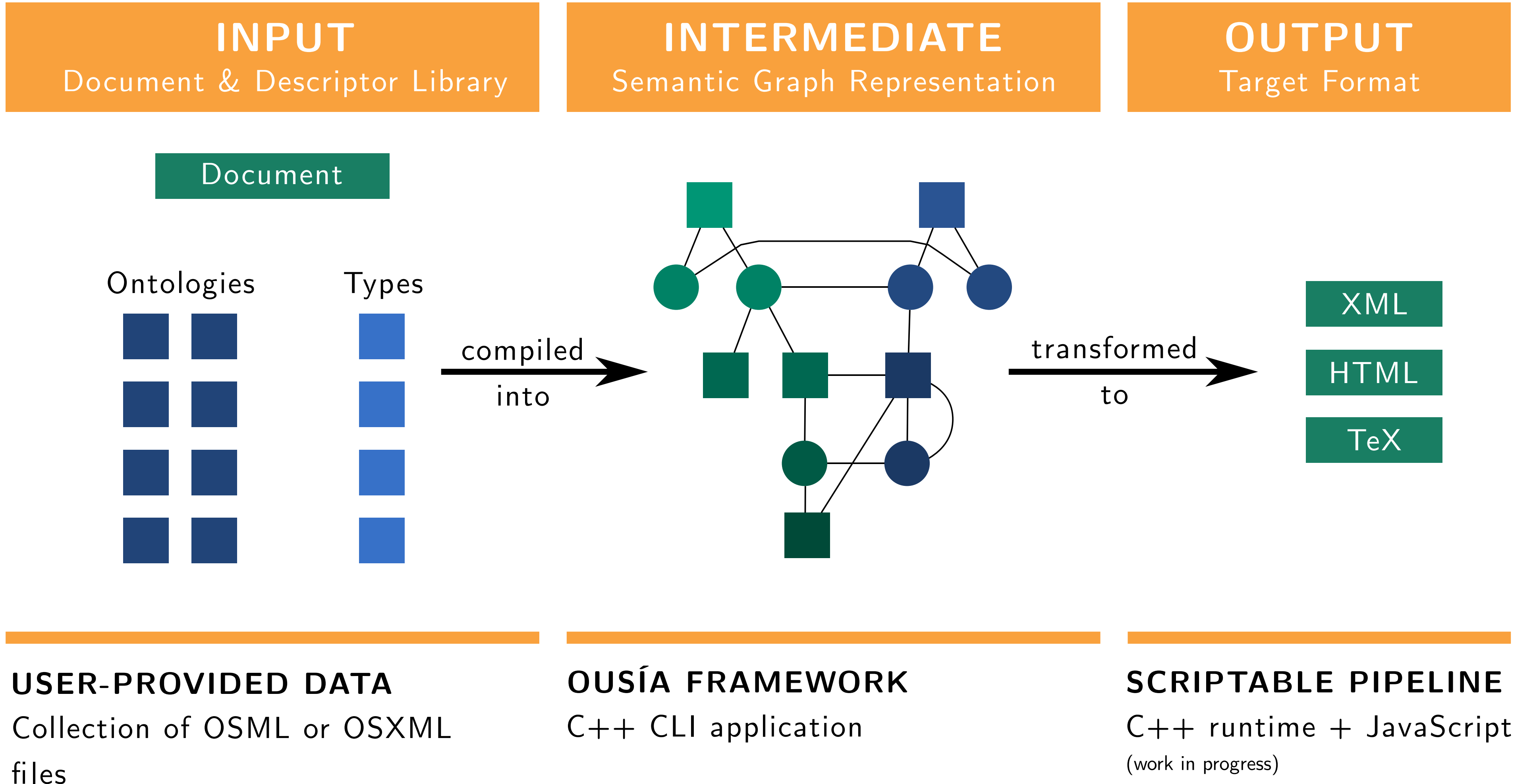
# Ousía

## Extensible Semantic Markup Framework

<http://ousia-framework.org/>

### MOTIVATION

- **Semantic markup enables a variety of interesting applications.** [1]
- **Problem:** Automatic extraction of semantics from plain text is difficult! [2, 3]
- **Solution:** Provide semantic information during document creation
- Benefits authors by capturing the intent and allows greater focus on content rather than stylistic details
- **Observation:** Current markup languages...
  - do not separate style and semantics consistently (e.g. LaTeX),
  - are not user-extensible (e.g. HTML) and
  - do not incorporate both, the concept of "structure" and "annotation" (e.g. XML).



### ONTOLOGY (OSML)

```

\ontology#artcl
  \struct#article[root=true]
    \field#title[subtree=true]
      \childRef[ref=primitive]
    \field
      \childRef[ref=section]
      \childRef[ref=paragraph]
  \annotation#emph

\struct#section
  \field#title[subtree=true]
    \childRef[ref=primitive]
  \field
    \childRef[ref=paragraph]

\struct#paragraph[transparent=true]
  \field
    \childRef[ref=primitive]

\struct#primitive

\struct#text[isa=primitive, transparent=true]
  \primitive[type=string]
    
```

### DOCUMENT (OSML)

```

\import[ontology]{artcl}

\begin{article}
  \title{Separation of presentation
    and content}

  <\emph Separation of presentation
    and content> is a design
    philosophy, and a methodology
    applied in the context of various
    publishing technology disciplines.

  \section{Intended Meaning}
    The underlying concept is to make
    a distinction between the actual
    <\emph meaning> of a document,
    and how this meaning is <\emph
    presented> to its readers.

  \section#sec_mr{Machine Readability}
    The general aim of separation of
    presentation and content is machine
    readability, making it possible for
    machines to detect meaning.

\end{article}
    
```

### OUTPUT (XML)

```

<document>
  <import rel="ontology" src="artcl"/>
  <artcl:article>
    <title><artcl:text>Separation of
      presentation and content
    </artcl:text></title>
    <artcl:paragraph><artcl:text>
      <a:start:artcl:emph/>Separation of
      presentation and content
      <a:end:artcl:emph/> is a design [...]
    </artcl:text></artcl:paragraph>
    <artcl:section>
      <title><artcl:text>Intended Meaning
      </artcl:text></title>
      <artcl:paragraph><artcl:text>The
      underlying concept is to make [...]
      </artcl:text></artcl:paragraph>
    </artcl:section>
    <artcl:section name="sec_mr">
      <title><artcl:text>Machine Readability
      </artcl:text></title>
      <artcl:paragraph><artcl:text>The
      general aim of separation of
      presentation [...]
      </artcl:text></artcl:paragraph>
    </artcl:section>
  </artcl:article>
</document>
    
```

Description

#### Fields

- Specify permitted **children** in the document tree
- Can be **primitive** → only primitive content of a certain type allowed

#### Classes

- Annotation and structure classes define the document vocabulary.
- Classes have **fields**.
- Classes have **attributes**.

#### Transparency

- Structures marked as transparent can be **deduced automatically**.
- Greatly reduces the effort needed to write documents

#### Annotations

- An **annotation** is a special structure outside the document tree.
- Range set by **start** and **end** anchor
- **Overlap** allowed

Annotations    ..... Correspondences

Example text adapted from [http://en.wikipedia.org/wiki/Separation\\_of\\_presentation\\_and\\_content](http://en.wikipedia.org/wiki/Separation_of_presentation_and_content)

### FORMATS

- Two native markup formats, **OSML** and **OSXML**:
- OSML is a LaTeX-esque language with high expressiveness targeted at end users.
- OSXML is an XML language used for automatic processing and as output format.
- Both formats are interchangeable.

### APPLICATIONS

- Scientific communication
- Semantic databases
- Semantic search
- Text mining
- Content management systems
- Export documents to a multitude of formats and devices
- Document validation (e.g. forms)
- Facilitating unified design principles (e.g. corporate design)
- Building block of a decentralized and dynamic semantic web

Literature

- [1] Shadbolt, N., Berners-Lee, T. and Hall, W. 2006. The Semantic Web Revisited. IEEE Intelligent Systems 21(3): 96-101
- [2] Russell, S. and Norvig, P. 2010. Artificial intelligence: a modern approach. Pearson 3rd Edition: Chapters 22-23
- [3] Görz, G. 2003. Handbuch der künstlichen Intelligenz. Oldenbourg 4th Edition