

LibreOffice  
Conference  
BUCHAREST



LibreOffice  
The Document Foundation

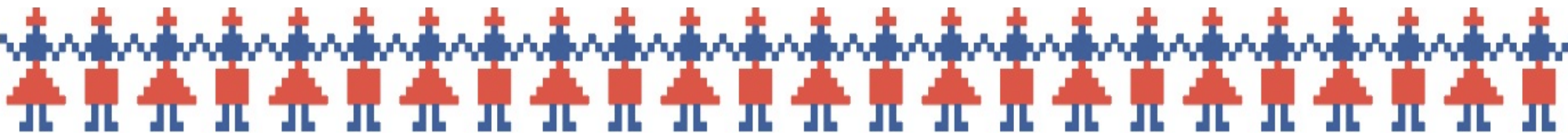
# ODF and it's Toolkit

ODP



Svante Schubert

PDF



# Basics

## What is a Standard?

- A standard is a **blueprint** like a **cooking recipe**!
- A standard creates interoperability (e.g. DINA4)
- A standard improves reusability
- A standard prevents Lock-In-Effect
- A standard is lowering costs (e.g. freetests/validator)

# Basics

## What is a Standard?

- A standard is a **blueprint** like a **cooking recipe**!
- A standard creates interoperability (e.g. DINA4)
- A standard improves reusability
- A standard prevents Lock-In-Effect
- A standard is lowering costs (e.g. freetests/validator)
- A standard is usually published as PDF (digital stone)  
Unfortunately no Digitalisation for Software Standards, yet

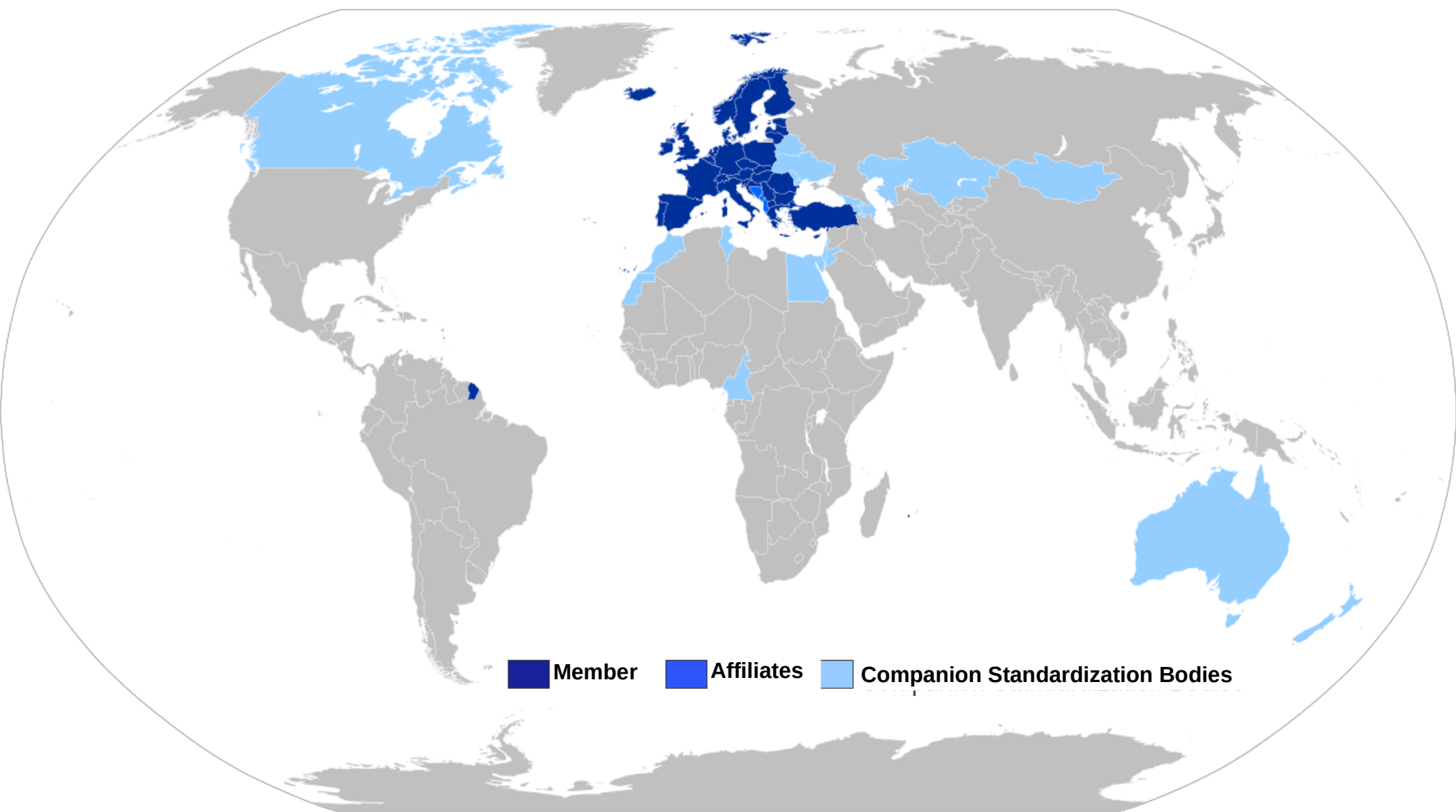
# Digitalisation

## In a Nutshell - Automation

- Everybody should be able at any time to:  
**Throw data over the digital fence & automatically understand received data:**
  - Without human interaction!
  - Without bi-lateral agreement!

# What is an EU Standard?

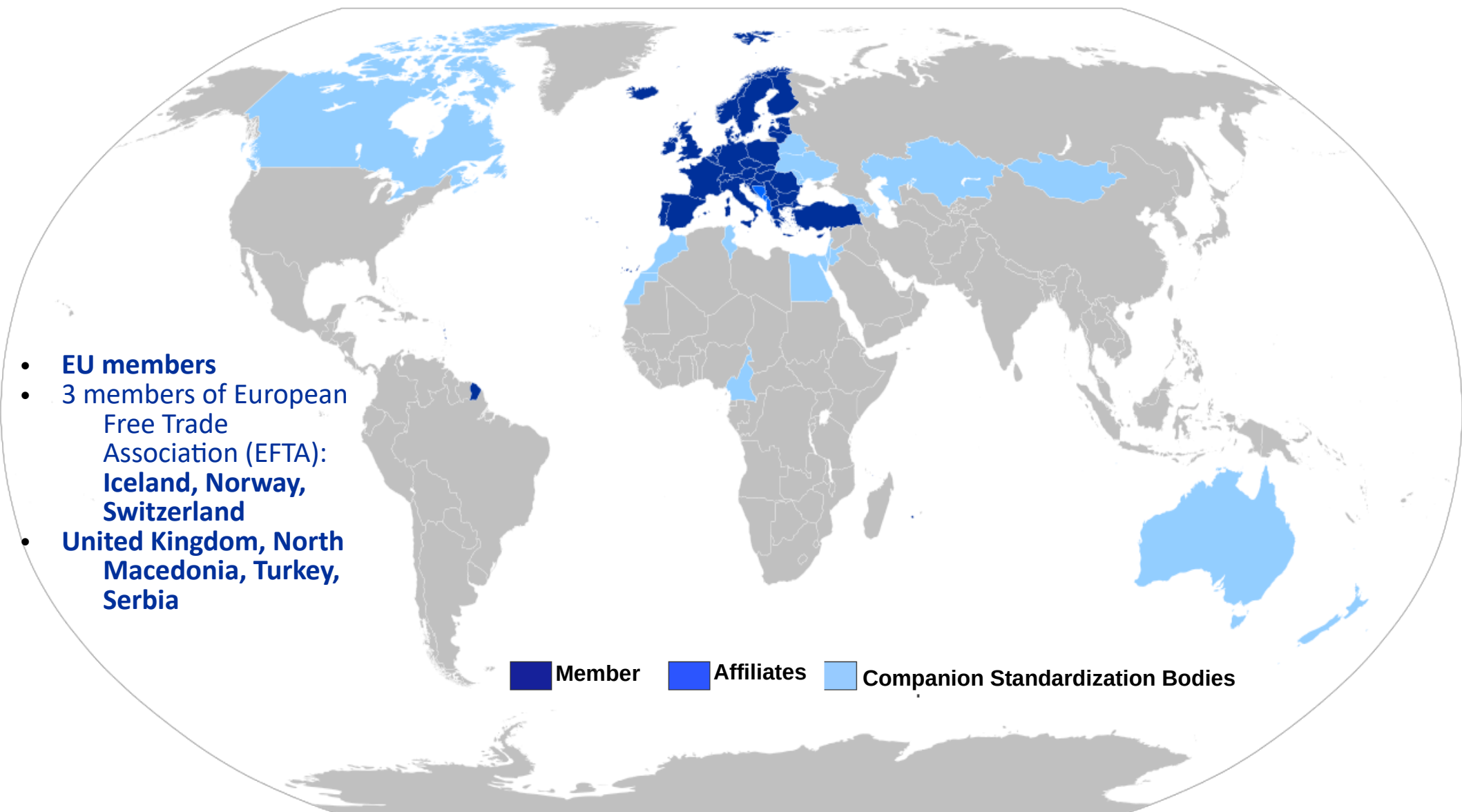
Who is CEN (Comité Européen de Normalisation)?



Member
  Affiliates
  Companion Standardization Bodies

# What is an EU Standard?

Who is CEN (Comité Européen de Normalisation)?

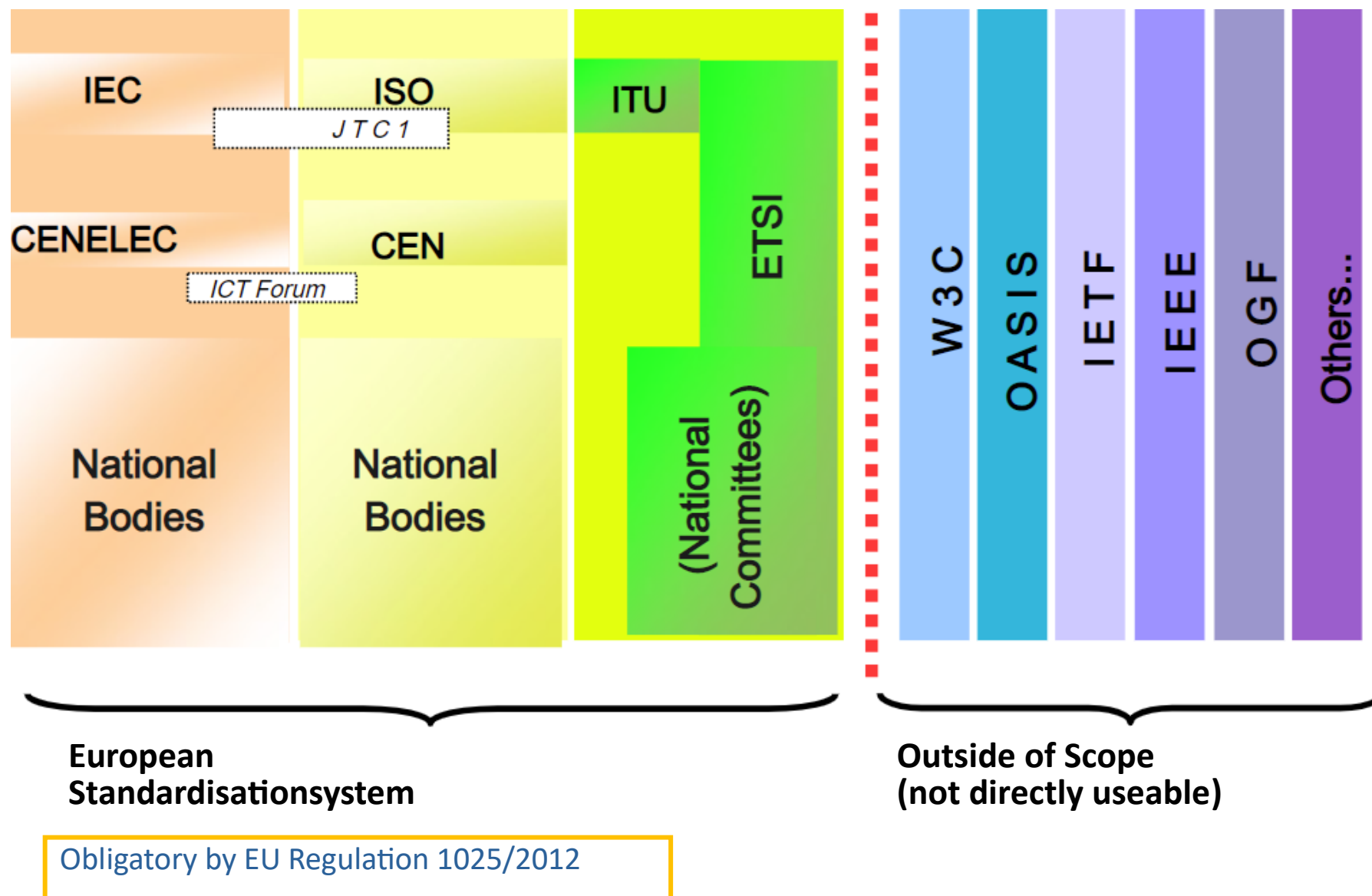


- **EU members**
- 3 members of European Free Trade Association (EFTA): **Iceland, Norway, Switzerland**
- **United Kingdom, North Macedonia, Turkey, Serbia**

Member
  Affiliates
  Companion Standardization Bodies

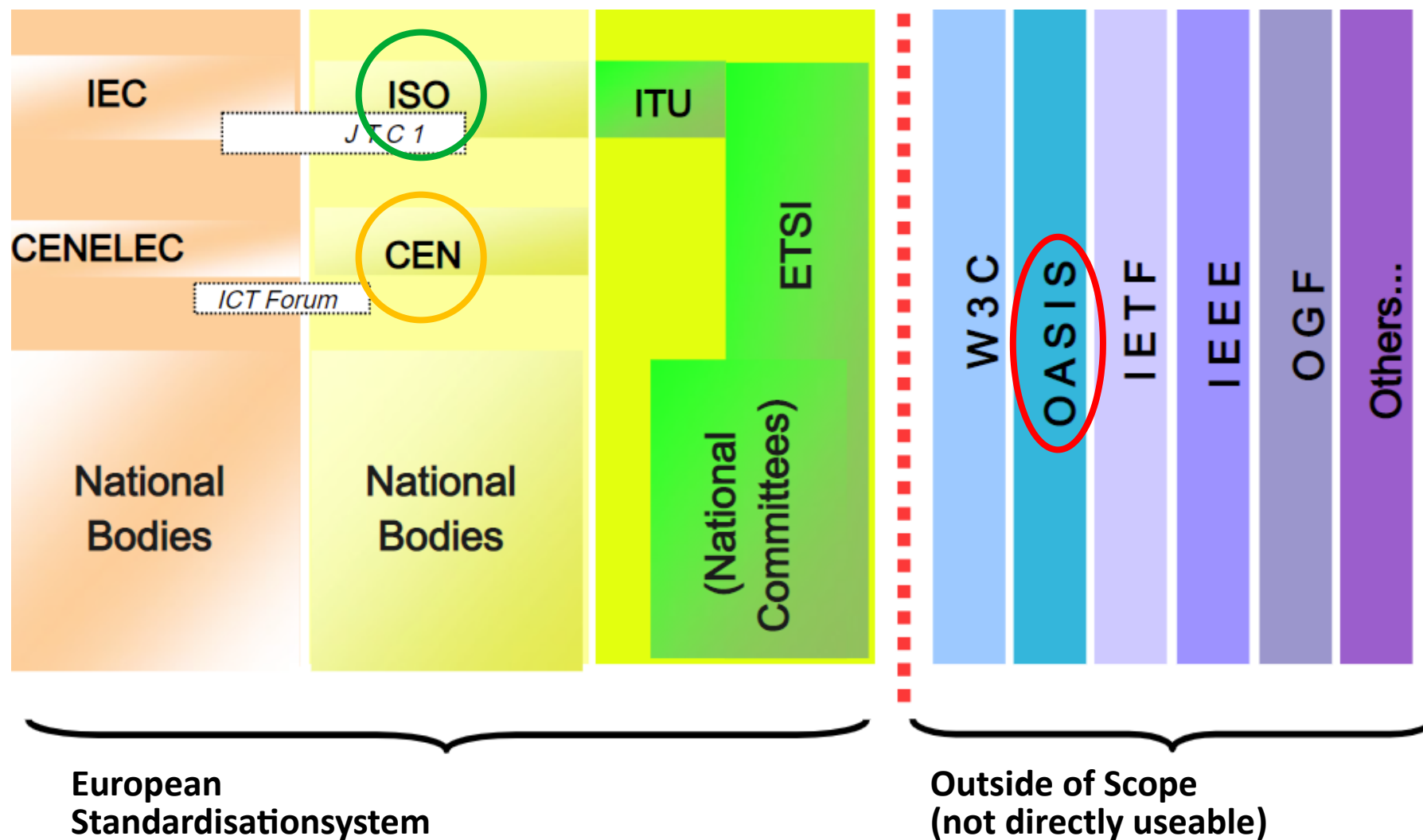
# What is EU ICT Standardisation?

Information and Communication Technologies (ICT)



# What is EU ICT Standardisation?

Information and Communication Technologies (ICT)

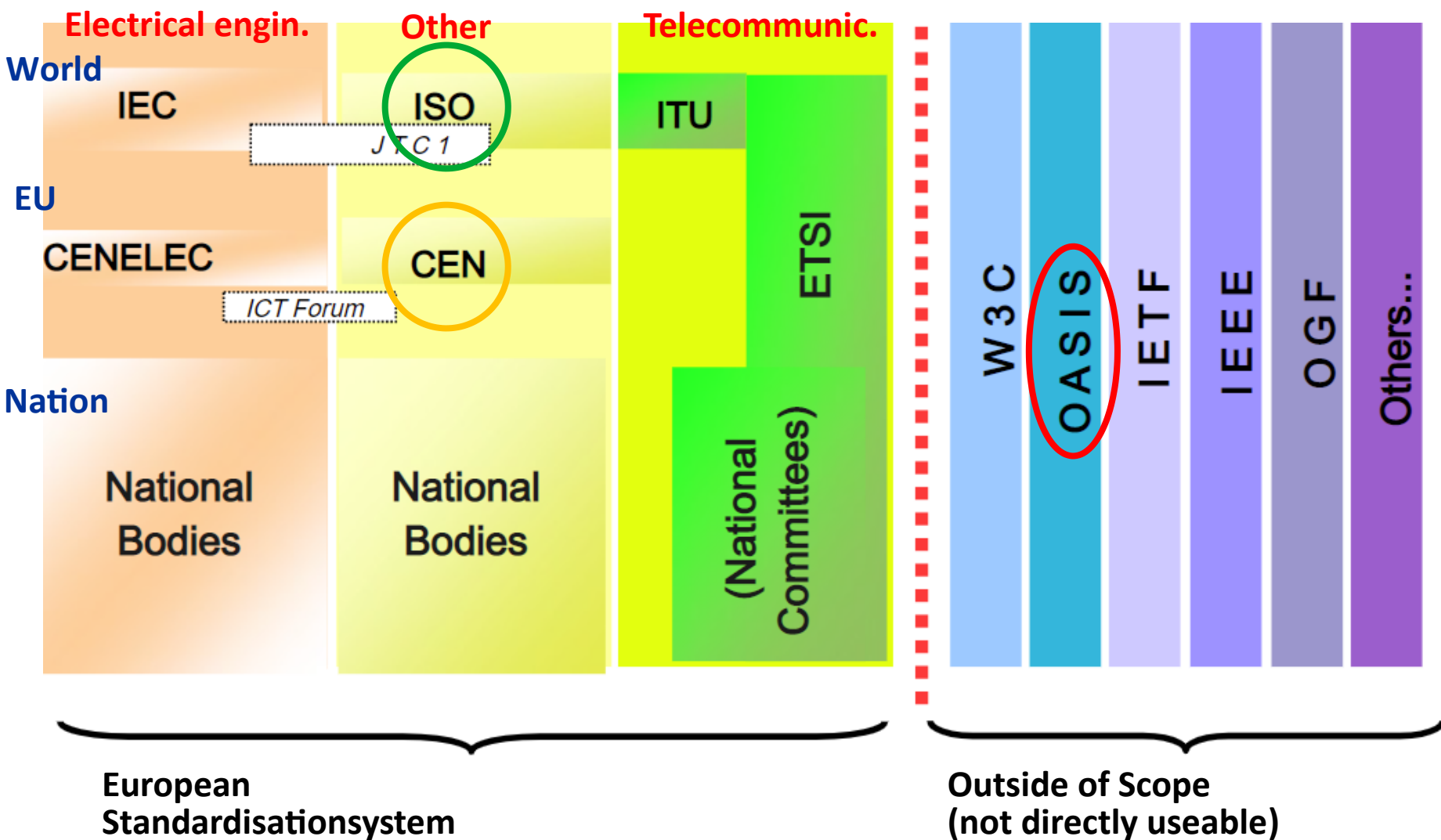


Obligatory by EU Regulation 1025/2012



# What is EU ICT Standardisation?

Information and Communication Technologies (ICT)



Obligatory by EU Regulation 1025/2012

# OASIS and ISO ODF Standard

## Difference

- ISO standard is the “queen” of standards
- Governments are obligated to use ISO standards
- Great Britain Government demands the use of ODF
- ISO standard is usually not free and not open
- OASIS ODF standard is given to ISO via „fast path“
- ODF ISO standard only differs on editorial level
- ODF Standard is only maintained on OASIS level

# OASIS and ISO ODF Standard

## Status

- ODF 1.3 was published (27. April 2021)
- ISO standard ODF 1.3 is “in the queue”
- OASIS TC has a feature freeze for ODF 1.4

# OASIS OpenDocument Technical Committee (TC)

## Voting Members

- Patrick Durusau [**OASIS TC chair & editor**]
- Svante Schubert [**OASIS TC chair & editor**]
- Francis Cave [**OASIS TC secretary & editor**]
- Regina Henschel (*TDF*)
- Michael Stahl (*allotropia*)
- Alfred Hellstern (*Microsoft*)
- Prof. Andreas Guelzow (worked on Calligra - left mid 2023)

# ODF Features

## OASIS ODF 1.4 issues

- Feature freeze for ODF 1.4 at the OASIS TC
- **OASIS issues tracker** (for ODF Syntax and/or Semantic)
  - 77 issues for ODF 1.4 in OASIS issue tracker
- **GitHub issue tracker** (for editorial issues of specification)
  - 5 editorial issues on ODF 1.4 specification
  - GitHub for Software Tooling required for ODF spec delivery

# ODF Features

## ODF Status of LibreOffice

- **TDF Budget 2023:**
  - Missing ODF Features (general collection)
  - Missing ODF Features: Attribute `svg:d` of `<draw:path>` some of the possible commands are missing
  - Missing ODF Features: `Draw:shadow-offset-x/y` only partially implemented
  - Missing ODF Features: The attribute `draw:text-rotate-angle` is interpreted, but there exists no user interface to change it.
  - Missing ODF Feature: `<draw:regular-polygon>` missing completely
  - **Improve release process of ODF versions (Automation)**
- **TDF Budget 2024: Transition to ODF 1.4**

# ODF SPECIFICATION

## GENERATION MISSING



### 13.8 <form:property>

The <form:property> element defines the name, type and value of a property.

The <form:property> element is usable within the following element: <form:properties> [13.7](#).

The <form:property> element has the following attributes: form:property-name [19.300](#), office:boolean-value [19.371](#), office:currency [19.373](#), office:date-value [19.374](#), office:string-value [19.383](#), office:time-value [19.386](#), office:value [19.388](#), office:value-type [19.389](#).

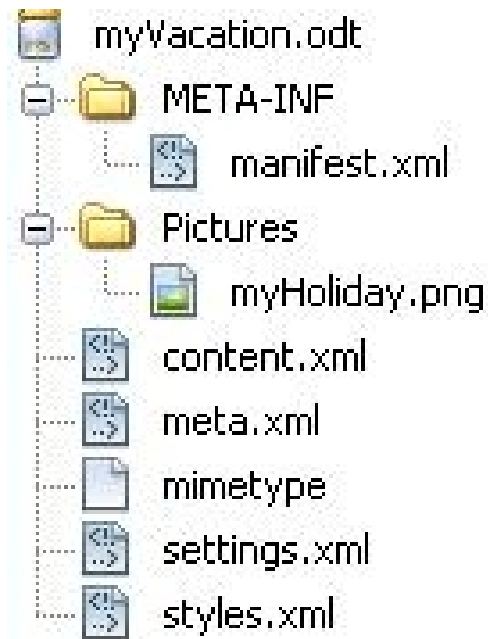
The <form:property> element has no child elements.

[https://docs.oasis-open.org/office/OpenDocument/v1.3/os/part3-schema/OpenDocument-v1.3-os-part3-schema.html#element-form\\_property](https://docs.oasis-open.org/office/OpenDocument/v1.3/os/part3-schema/OpenDocument-v1.3-os-part3-schema.html#element-form_property)

# Basics

## What is a ODF?

- **ODF document** is a ZIP (of mostly XML)



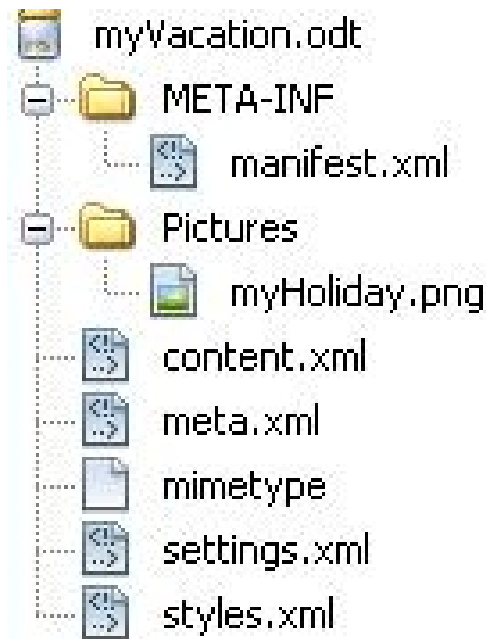
- **ODF Package**  
(OASIS spec. part 2)
- **ODF XML**  
(OASIS spec. Part 3)



# Basics

## What is a ODF?

- **ODF document** is a ZIP (of mostly XML)



- **ODF Package**  
(OASIS spec. part 2)
- **ODF XML**  
(OASIS spec. Part 3)

- **OASIS ODF Primer** missing:  
Semantic and Syntax in a Nutshell

# Basics

What is Semantik & Syntax?

- **Semantic** (*Wine*) described by **Syntax**/Language (*wine/vino*)



# Office Document Semantics

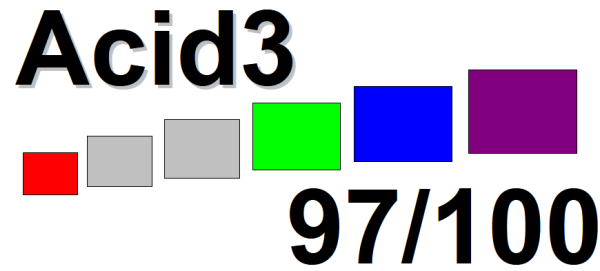
## Features and Subfeatures

- Required to split complexity: How to eat an elephant?
- Semantics is what the common user “knows”
- Semantics are known accross file formats (e.g. table)
- Semantics of file format are the features
- Subfeature  
If a feature (e.g. table background color) does not exist without its parent (table) it is a subfeature
- There can be many syntaxes for a semantic (XML, JSON..)

# Document Standards

## API on Semantic & Syntax

- ODF XML only the final state of a document (load / save)
- ODF defines no state changes (e.g. column insertion in table)
- No state changes == no API == no regression test at standard  
Like [HTML/CSS ACID regression tests](#)

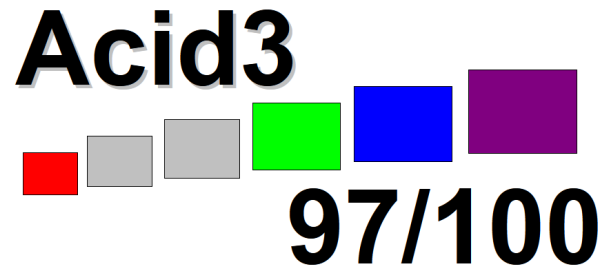


To pass the test, a browser must use its default settings, the animation has to be smooth, the score has to end on 100/100, and the final page has to look exactly, pixel for pixel, like [this reference rendering](#).

# Document Standards

## API on Semantic & Syntax

- ODF XML only the final state of a document (load / save)
- ODF defines no state changes (e.g. column insertion in table)
- No state changes == no API == no regression test at standard  
Like [HTML/CSS ACID regression tests](#)



To pass the test, a browser must use its default settings, the animation has to be smooth, the score has to end on 100/100, and the final page has to look exactly, pixel for pixel, like [this reference rendering](#).

- Remember Change Tracking becomes easier after defining Changes!
- Are such interoperable API / Tests / Macros desired?

# Document Standards

## Semantic Features and their Changes

- 1) Find semantic features (feature tree of ODF)
- 2) What can be added / changed / deleted via LO GUI?  
User feature (and their state changes)
- 3) Analysis of ODF grammar – What elements stick together?
- 4) What is the vocabulary on features used by ODF specification?
- 5) What is the vocabulary on features used by LO documentation and / or testing?



# ODF Validator

This service checks conformance of ODF documents based on their OpenDocument Format specification. It does not cover all conformance criteria, yet (see [implementation details](#)).

## [ODF Version:](#)

auto-detect

## [Logging:](#)

verbose

Choose ODF documents for validation:

list.odt

This service is provided to you by [The Document Foundation](#)



**LibreOffice**  
The Document Foundation

This service does not cover all conformance criteria of the OpenDocument Format specification. It is not applicable for formal validation proof. Problems reported by this service only indicate that a document may not conform to the specification. It must not be concluded from errors that are reported that the document does not conform to the specification without further investigation of the error report, and it must not be concluded from the absence of error reports that the OpenDocument Format document conforms to the OpenDocument Format specification.

# ODF Toolkit

## Use Cases (1/2)

- Online Validator (or via commandline)  
<https://odfvalidator.org/>
- Running XSLT directly on ODF document (no unzipping XML)



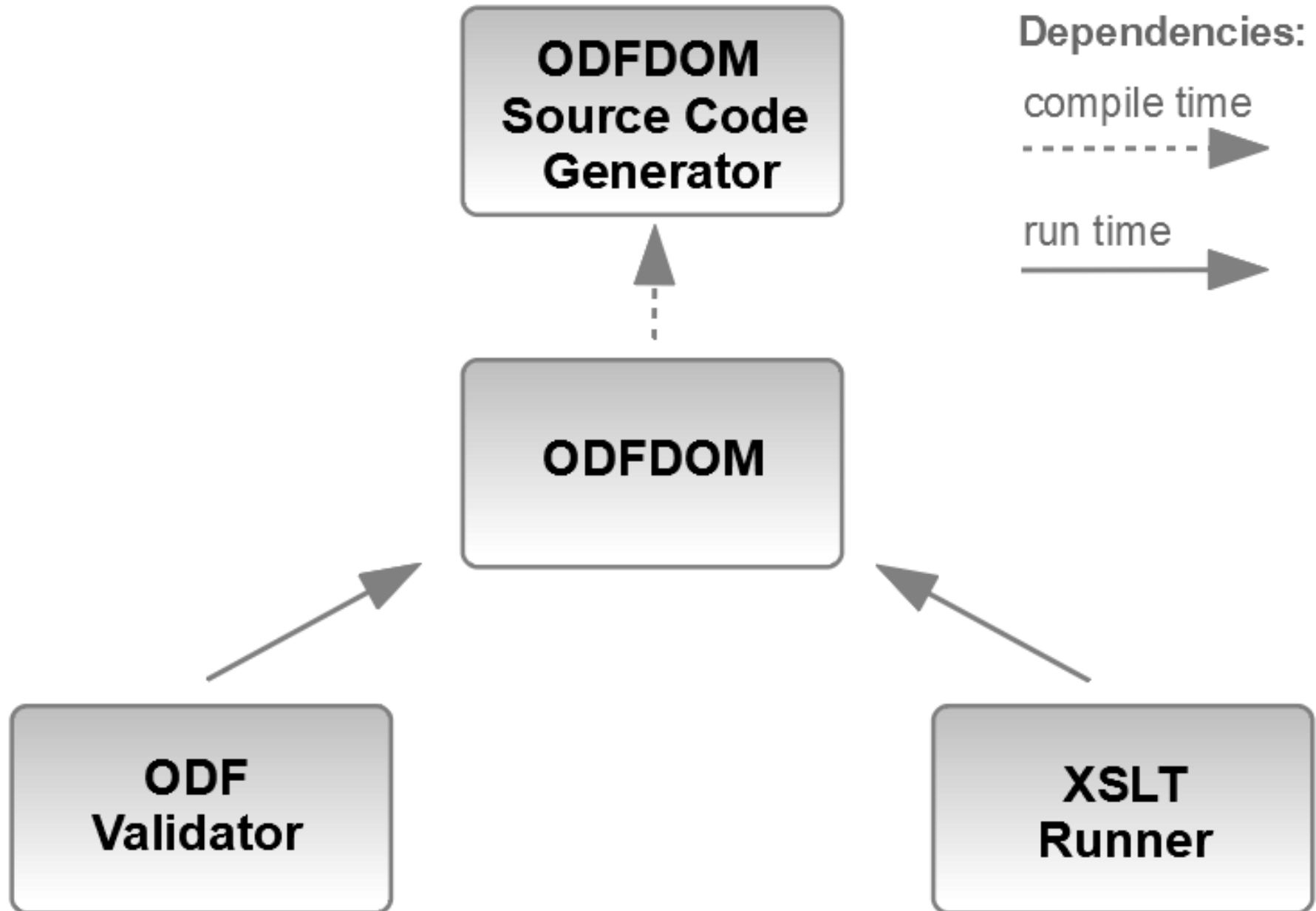
# ODF Toolkit

## Use Cases (2/2)

- Editing an ODF document (e.g. Cloud)
  - by API without Layout
  - for Data Insertion (e.g. by Database)
  - for Data Extraction (e.g. Translation)
- Collaboration on Text Documents (ODT)
  - backend for Web Offices  
(e.g. OX Documents)  
***(started with v0.10.0 – Nov '21)***

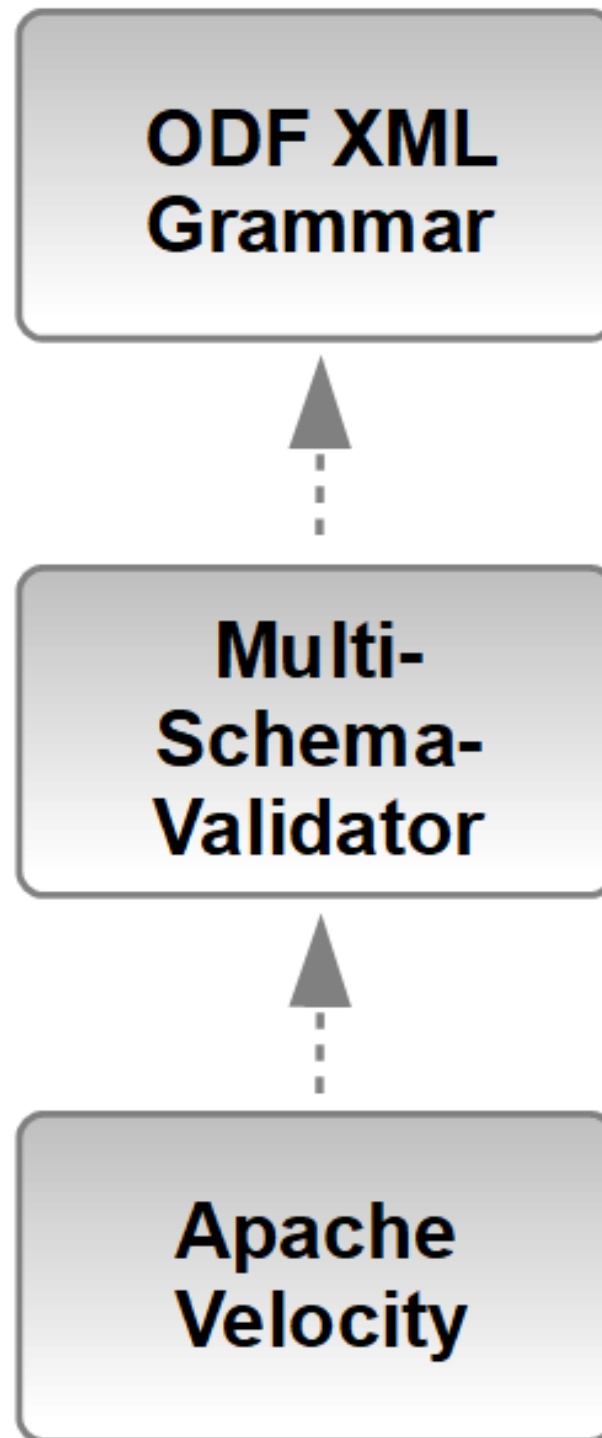
# ODF Toolkit

## Architecture



# ODFDOM Source Code Generator

Architecture



**OASIS XML grammar:**  
~18000 text lines  
~600 XML elements  
~1200 XML attributes

XML Validator -  
**reads** many  
XML grammars

**template engine**  
generating sources  
by text templates  
having a context  
with Java access

# ODF GRAMMAR - TEXT

## HARD TO ANSWER



Can a  
paragraph **<text:p>**  
be nested  
in a valid document?



### ODF 1.2 XML:

- 598 **XML Elements**
  - 1301 **XML Attributes**
- > 18k lines

# ODF GRAMMAR - TEXT

## HARD TO READ



```
15688 <rng:define name="table-table">
15689   <rng:element name="table:table">
15690     <rng:ref name="table-table-attlist"/>
15691     <rng:optional>
15692       <rng:ref name="table-title"/>
15693     </rng:optional>
15694     <rng:optional>
15695       <rng:ref name="table-desc"/>
15696     </rng:optional>
15697     <rng:optional>
15698       <rng:ref name="table-table-source"/>
15699     </rng:optional>
15700     <rng:optional>
15701       <rng:ref name="office-dde-source"/>
15702     </rng:optional>
15703     <rng:optional>
15704       <rng:ref name="table-scenario"/>
15705     </rng:optional>
15706     <rng:optional>
15707       <rng:ref name="office-forms"/>
15708     </rng:optional>
15709     <rng:optional>
15710       <rng:ref name="table-shapes"/>
15711     </rng:optional>
15712     <rng:ref name="table-columns-and-groups"/>
15713     <rng:ref name="table-rows-and-groups"/>
15714     <rng:optional>
15715       <rng:ref name="table-named-expressions"/>
15716     </rng:optional>
15717   </rng:element>
```

## ODF 1.3 XML:

- 606 XML Elements
  - 1317 XML Attributes
- > 18k lines

# ODF GRAMMAR - TEXT

ANALYZE GRAMMAR  
GENERATE SOURCES



## Generation of ODFDOM sources

- Java class for each **XML Elements**
- Java class for each **XML Attributes**

## ODF 1.3 XML:

- 606 **XML Elements**
  - 1317 **XML Attributes**
- > 18k lines

# SOURCES GENERATED FOR ODF XML

ONE NAME

DIFFERENT CONTENT

DIFFERENT PARENT



Beyond parent <manifest:manifest>

```
189 <rng:element name="manifest:manifest">
201 <rng:attribute name="manifest:version">
202   <rng:value>1.3</rng:value>
203 </rng:attribute>
```

Two possibilities for  
@manifest:version

- "1.3"
- String

Beyond parent <manifest:file-entry>

```
119 <rng:element name="manifest:file-entry">
149   <rng:optional>
150     <rng:attribute name="manifest:version">
151       <rng:ref name="string"/>
152     </rng:attribute>
153   </rng:optional>
```

# SOURCES GENERATED FOR ODF XML

ONE NAME  
DIFFERENT CONTENT  
ONE PARENT



Same parent <manifest:key-derivation>

```
163 <rng:choice>
164   <rng:attribute name="manifest:key-derivation-name">
165     <rng:value>PGP</rng:value>
166   </rng:attribute>
167   <rng:interleave>
168     <rng:attribute name="manifest:key-derivation-name">
169       <rng:choice>
170         <rng:value>PBKDF2</rng:value>
171         <rng:ref name="anyURI"/>
172       </rng:choice>
173     </rng:attribute>
174     <rng:attribute name="manifest:salt">
175       <rng:ref name="base64Binary"/>
176     </rng:attribute>
177     <rng:attribute name="manifest:iteration-count">
178       <rng:ref name="nonNegativeInteger"/>
179     </rng:attribute>
180   </rng:interleave>
181   <rng:optional>
182     <rng:attribute name="manifest:key-size">
183       <rng:ref name="nonNegativeInteger"/>
184     </rng:attribute>
185   </rng:optional>
186 </rng:choice>
```

@manifest:key-derivation-name

- Value „PGP“ →  
no other attributes
- Value „PBKDF2“ or anyURI →  
3 other attributes



# ODFDOM

## Architecture

3.

ODF User API

**ODF Semantic Layer**

2.

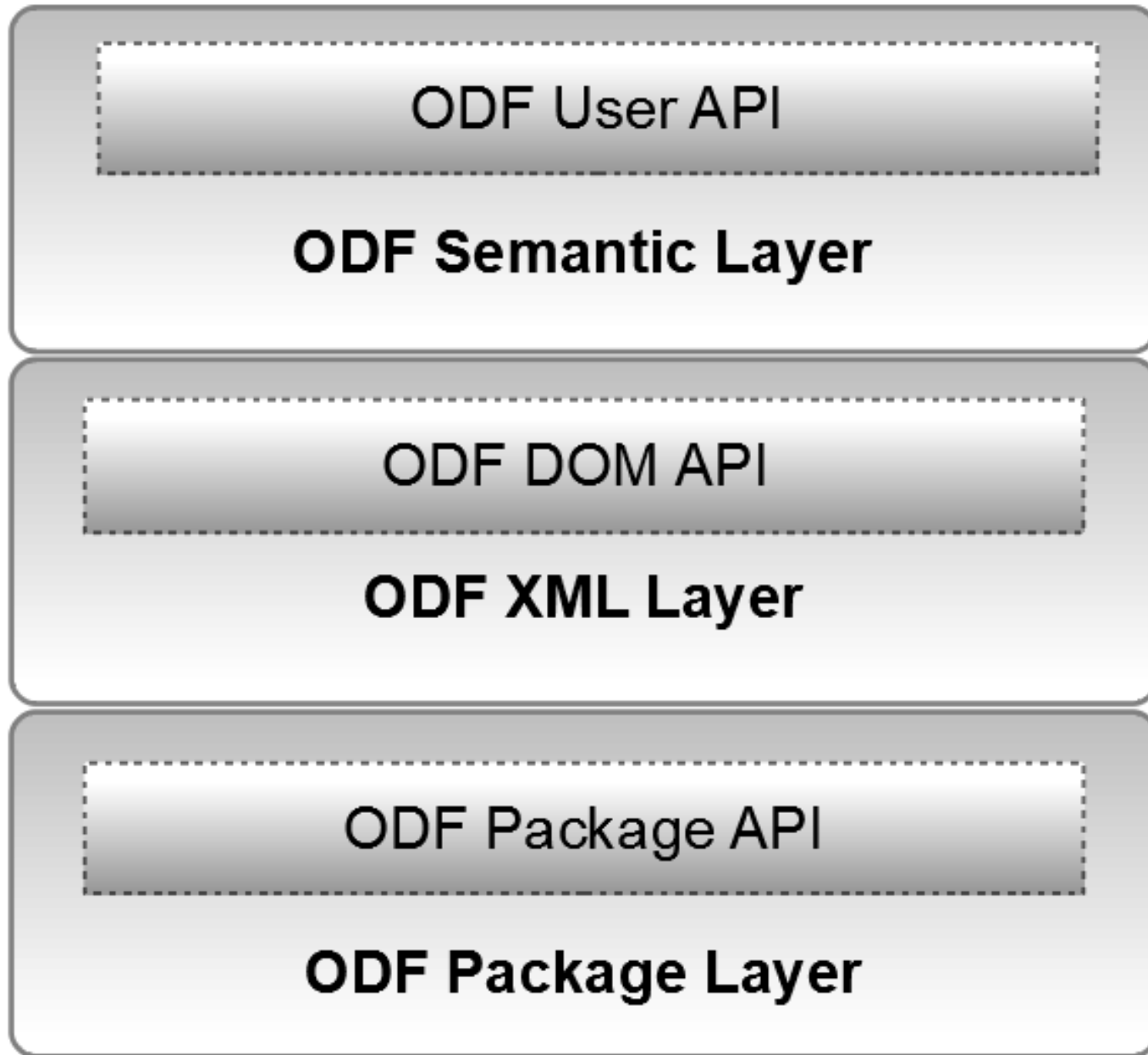
ODF DOM API

**ODF XML Layer**

1.

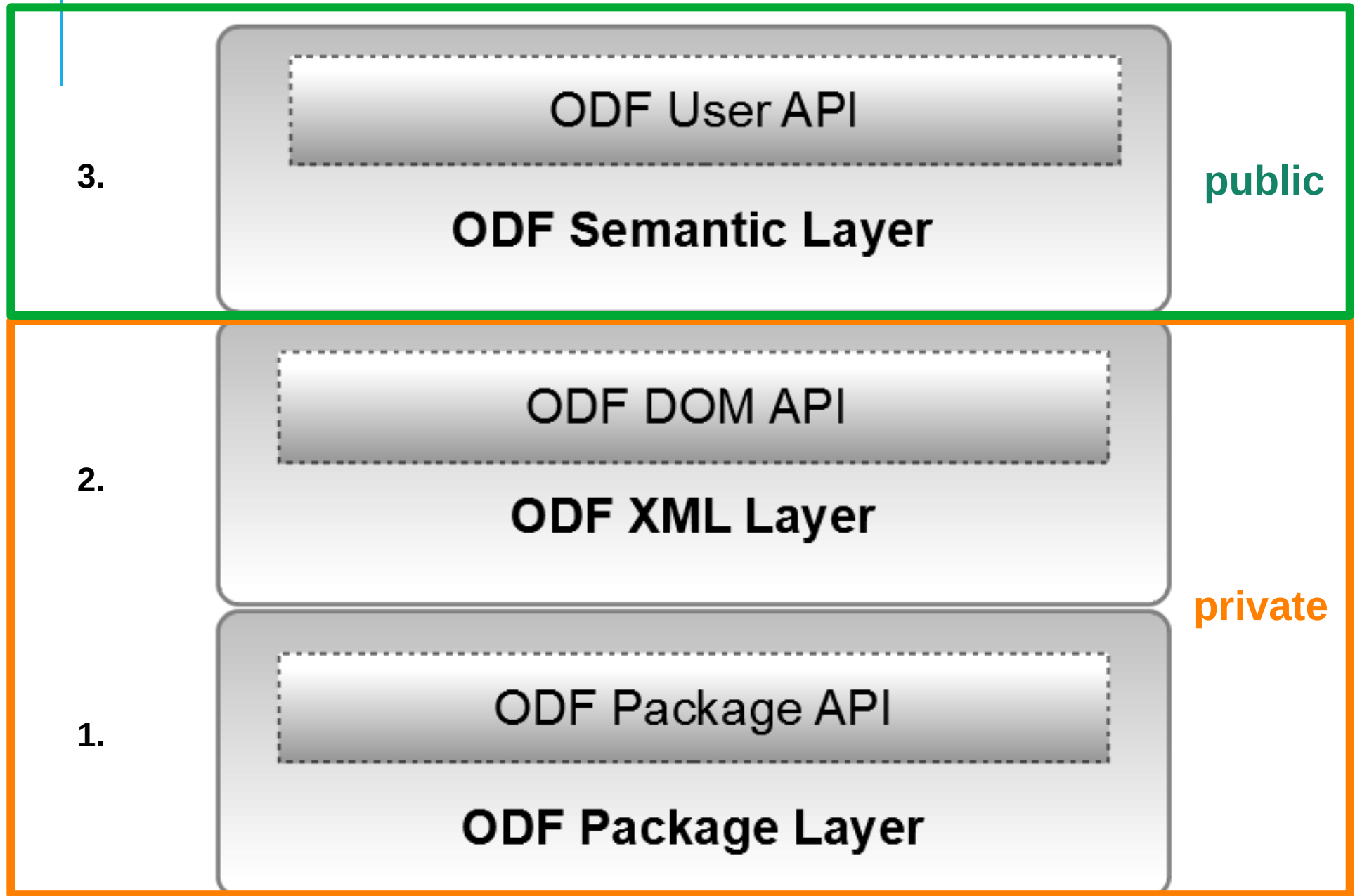
ODF Package API

**ODF Package Layer**



# ODFDOM

Architecture (in spe)



# ODF Toolkit

## Resources

- **Website:**

<https://odftoolkit.org/>

<https://tdf.github.io/odftoolkit/docs/> (latest)

- **Sources:**

<https://github.com/tdf/odftoolkit>

- **Online Validator (hosted by TDF)**

<https://odfvalidator.org/>

- **ODF Specification**

<http://docs.oasis-open.org/office/OpenDocument/v1.3/os/>

- **ODF Specification Tooling**

<https://github.com/oasis-tcs/odf-tc/>