The PAPI Framework

Sean McGrath

Technical Lead

Pan African Parliamentary
Interoperability Framework Initiative

Pre-requisites

- The “Rationale for PAPI” Presentation
- Basic knowledge of XML

Goals of PAPI

- A lingua franca for the semantic interchange of Parliamentary Documents in Africa
- An extensible, implementable baseline for semantic representation of Parliamentary documents
- Extensible metadata model, leveraging Eurovoc
- Extensible resource naming and resource linking (hypertext) models
- Minimal barriers to Parliamentary content creation

Why Semantic Models of Parliamentary Documents?

- Automation, automation, automation
  - Multiple output formats (pdf, html etc.) from single source documents
  - Automated production of derived documents e.g. briefs, amendment lists etc.
  - Automated change control and consolidation
  - Reporting
  - Hypertext management

Overview

- Semantic modelling – content versus presentation and the profit center/cost center model
- PAPI Continuum model & PAPI transforms
- PAPI Hierarchy/Containership model
- PAPI Metadata model
- PAPI Hypertext model & Naming conventions
- PAPI Validation & Business Rules model

The real world – content + presentation

- Semantic models separate contents from presentation
- But
  - Not always possible to totally separate presentation from content
    - Electronic Fidelity & Paper-centricity
    - Historical precedent
    - Tables, graphics etc.
  - Many examples in the PAPI documentation…

The real world – content Vs. presentation

- Profit Center versus Cost Center Markup
  - All tags have business costs. Only some tags have business benefits.
  - Where possible, separate content from presentation but *only where there is business value* in separating content from presentation

Content or presentation?

Short title and commencement----------Section 12

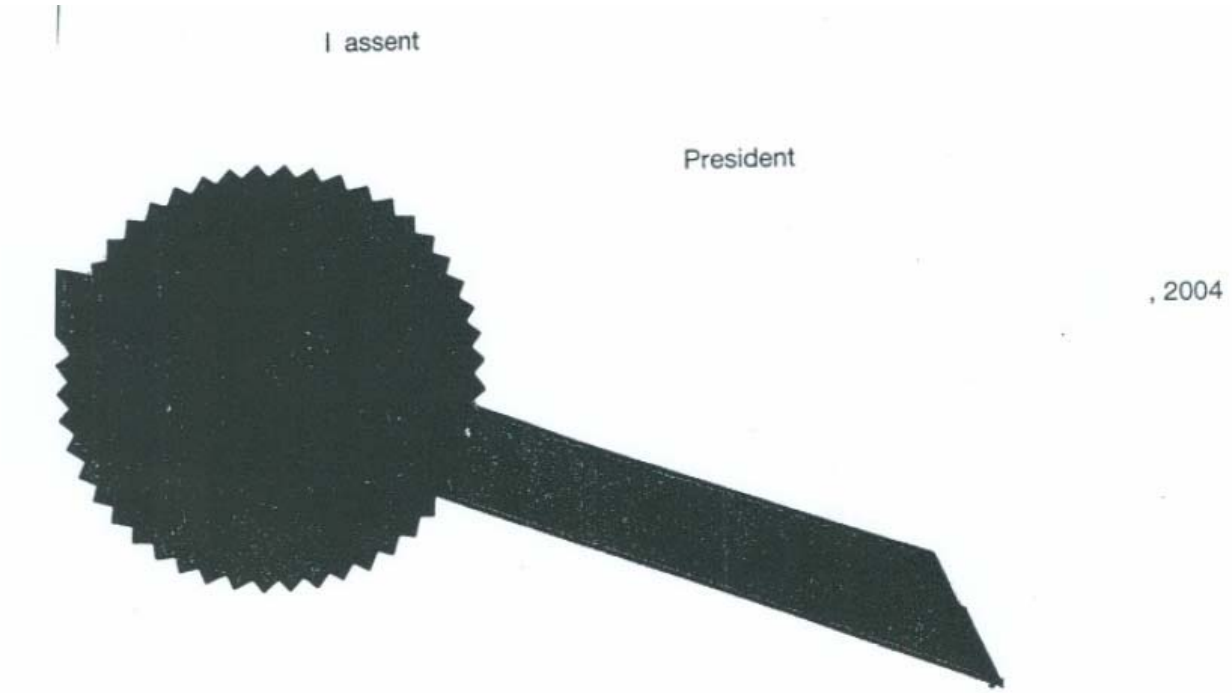
Dot Tab Leader

Side head

Short title.

The Education Act, 2004

Content or presentation?



Content or presentation?

Determination of applications

35. (1) An agency shall determine an application under this Part,

(a) by amending its records in accordance with the application;

or

(b) by refusing to amend its records.

(2) An agency that fails to determine an application within 30 working days after the application is received by the agency shall, for the purposes of this Act, be taken to have determined the application by refusing to amend its records in accordance with the application.

"(a) or (b)" or "(1) or (2)"?

Content or presentation?

87. A person who

(a) seeks or gains access to personal record of any other person under false pretences;

(b) willfully makes a false statement to mislead or gain access to information; or

(c) with intent to deny right of access to information,

(i) destroys, damages or alters a document;

(ii) conceals a document; or

(iii) makes a false entry in a document,

4 column table? White space? Full justification?

Content or presentation?

Bank of Tanzania

About the Bank

- Primary Objective and Function of the Bank
- Bank of Tanzania Act
- Board of Directors
- Board Members
- Calendar of Important Monetary and Economic Policy
- Bank of Tanzania Location
- History of the Bank
- Bank of Tanzania Web Site
- Governors: Past and Present

Financial Markets

Payment System

Bank of Tanzania Act

PRELIMINARY

1. (1) This Act may be cited as the Bank of Tanzania Act, 1995

(2) This Act shall come into operation on such date as the Minister shall, by notice published in the Gazette, appoint.

2. This Act shall extend to Tanzania Zanzibar as well as to Mainland Tanzania.

3. In this Act, unless the context otherwise requires—"the Bank" means the Bank of Tanzania established by this act; bank", except when used in the expressions "central bank" and "foreign bank", means a financial institution authorised to receive money on current account subject to withdrawal by cheque; "banking business" means—

TOC part of the legislation or part of rendering?

Content, presentation & end-use tools

- Presentation-oriented tools predominate
  - WP
  - Web-based editors
- Full structured XML authoring/editing systems not an option for many users:
  - Expense
  - Expertise
  - World view

The classic XML conundrum

- Engineers want pure, hierarchical, semantic models for downstream processing
- End-users want flat, presentation-oriented models for upstream content creation
- Many fine XML designs have failed in the real world because of failures to adequately address the needs of the upstream
- Plus, cannot totally separate content from presentation, so…

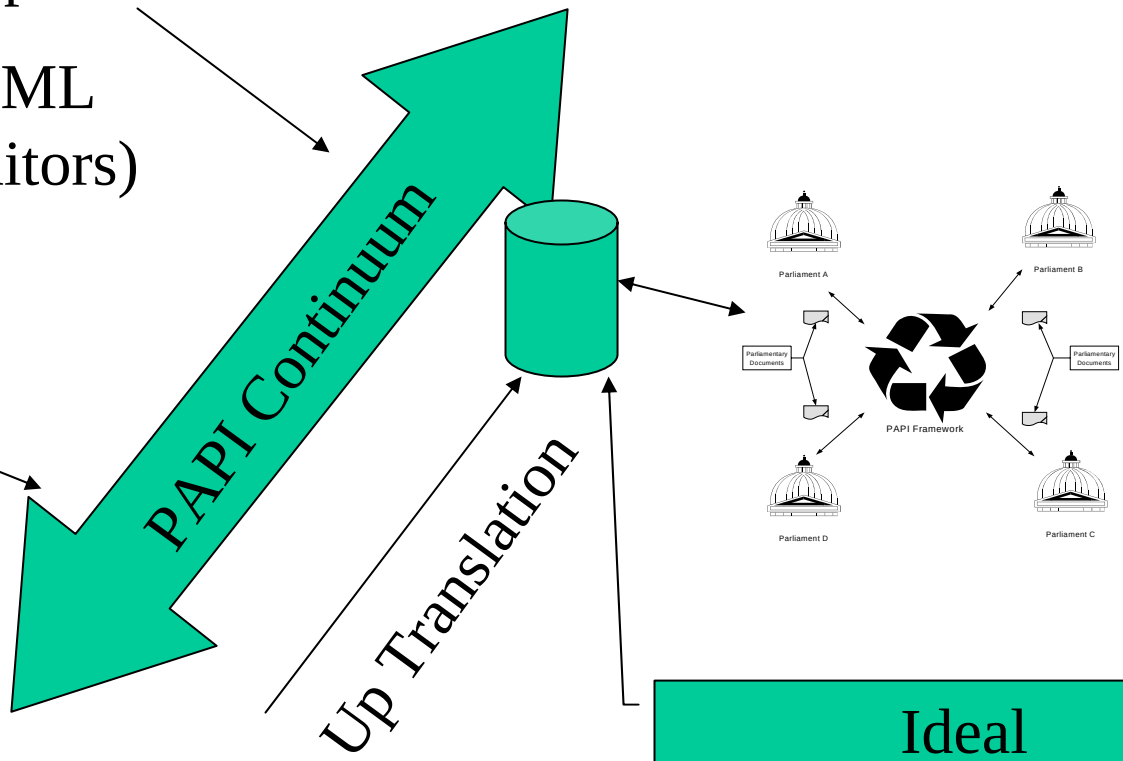
PAPI Continuum

PAPI
Semantic
Implicit

(XML
Editors)

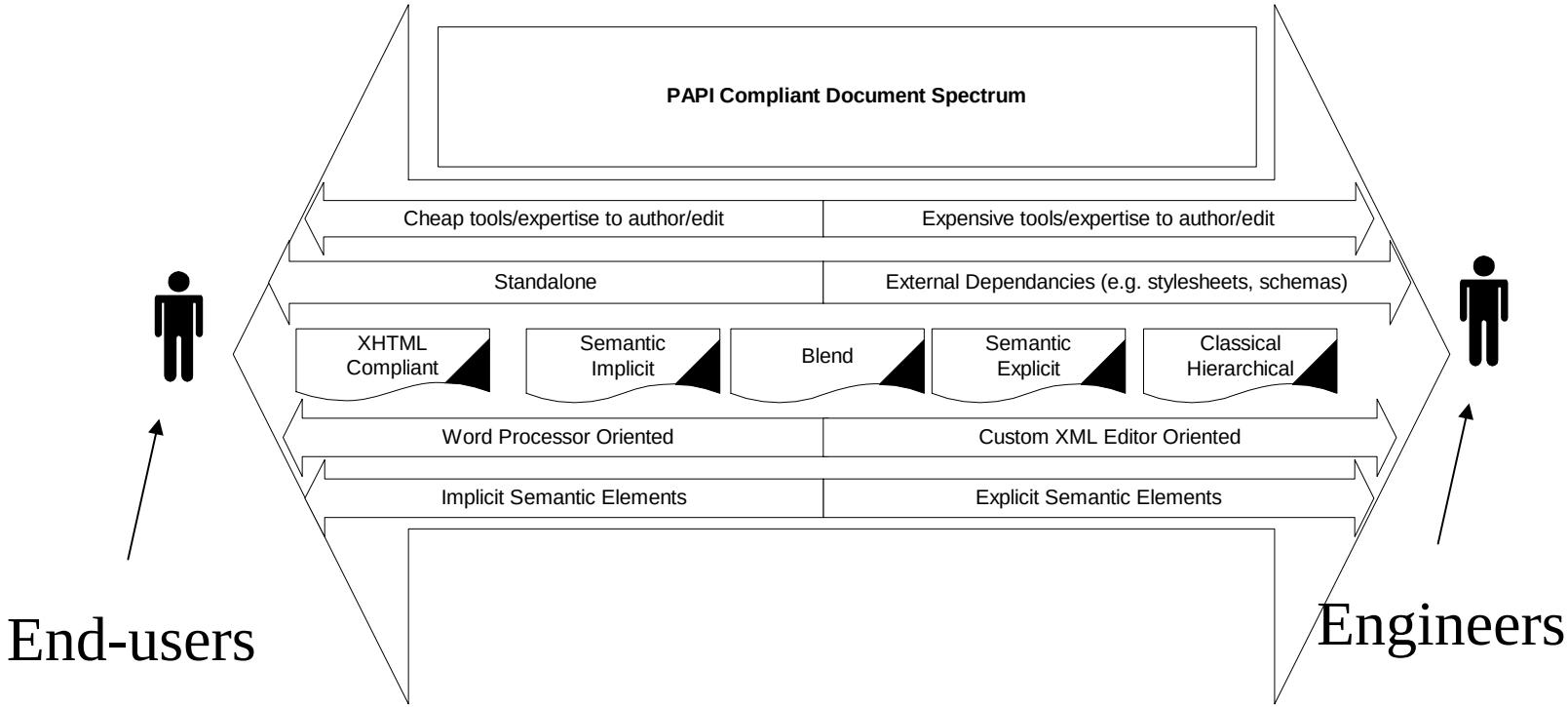
PAPI
XHTML

(XHTML
Editors)



Ideal
Interchange
Notation

The PAPI Continuum Model



Principle differences at ends of the PAPI continuum

- Three major areas:
  - Internal/external styling for presentation
  - Mechanism by which semantic elements are identified
  - Mechanism by which hierarchical structure is represented

The PAPI continuum model

- Identification of semantic elements
  - <Clause>
  - Vs.

  - <Part>
  - Vs. <h1 class="Part">
  - Etc.

The PAPI Continuum Model

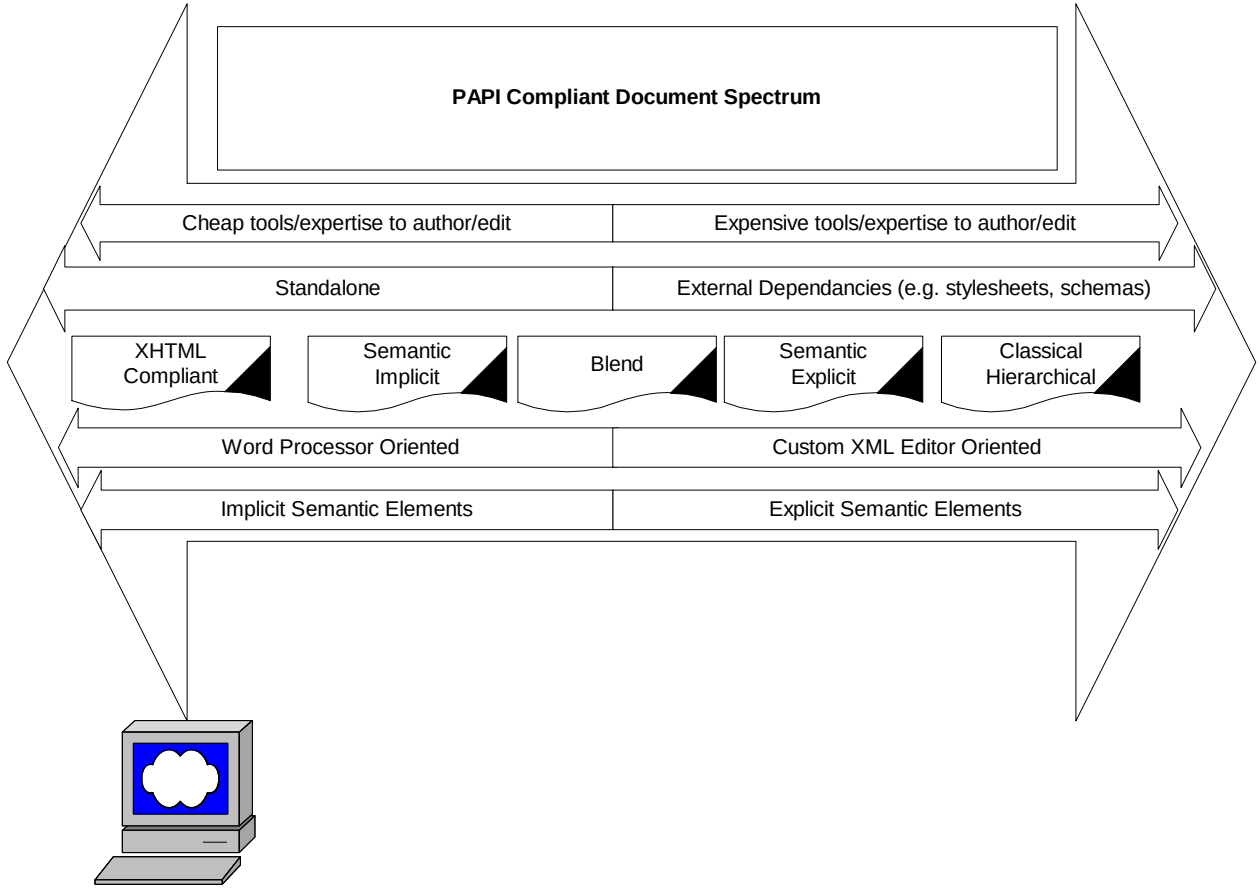
- Hierarchical structure

```
<h1 class="Part">
<title>Part 1</title>
<p>Stuff</p>…</h1>
```

Versus

```
<h1 class="Part">Part 1</h1>
<p>Stuff</p>
<h1 class ="Part">Part II…
```

PAPI XHTML



Web Editor/WP

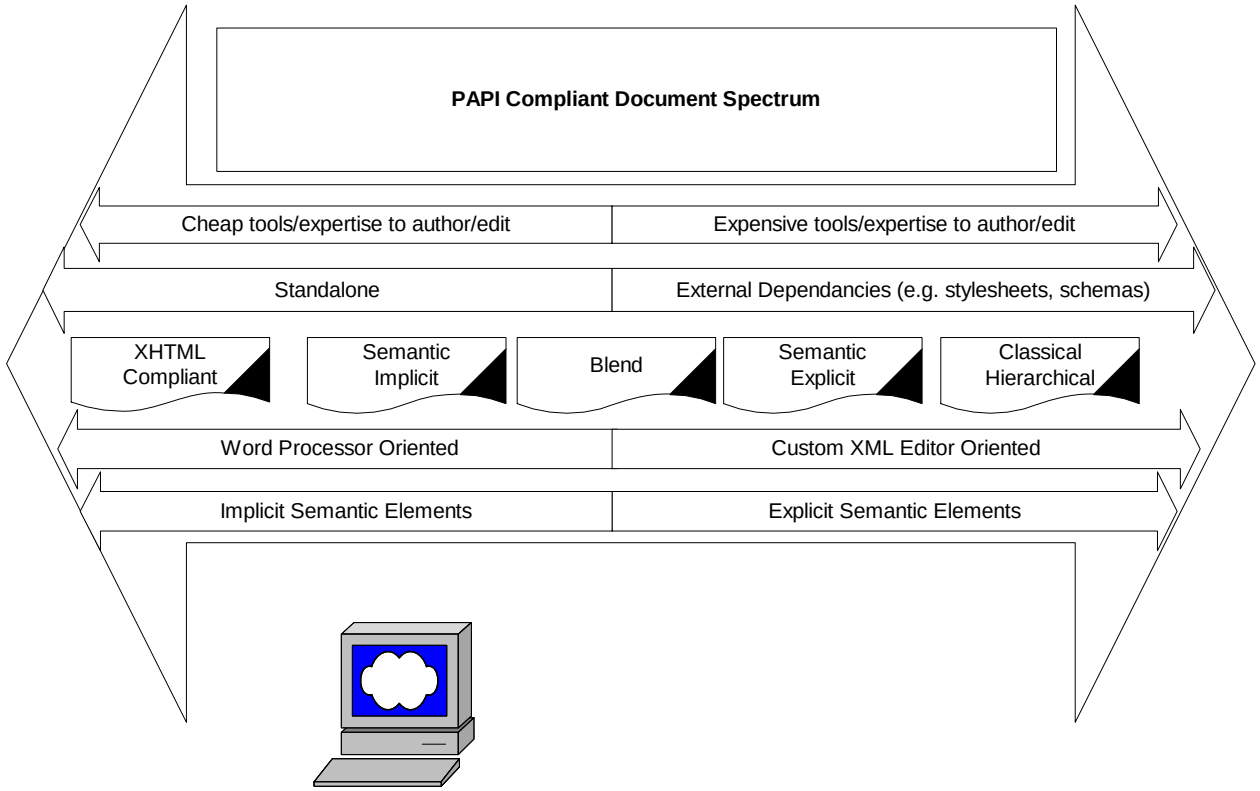
The PAPI Contiuum Model – PAPI XHTML

- PAPI-compliant XML documents can consist completely of XHTML markup. The standard `span` and `div` elements of XHTML can be used to encode PAPI semantics where required. Samples files of this form are in the `samples/xhtml_compliant` subdirectory.

PAPI XHTML

```
<html>
<div class ="ACT:Section">
<p>
<a name="sec1"/>1. (1) there is hereby
  established...</p>
<p>(2) The Commission -</p>
<p style="text-indent:10%">(a) shall be a
  body corporate...</p>
<p style="text-indent:10%">(b) may sue
  and be sued in...</p>
</div>
</html>
```

PAPI Semantic Implicit



WP or XML editor

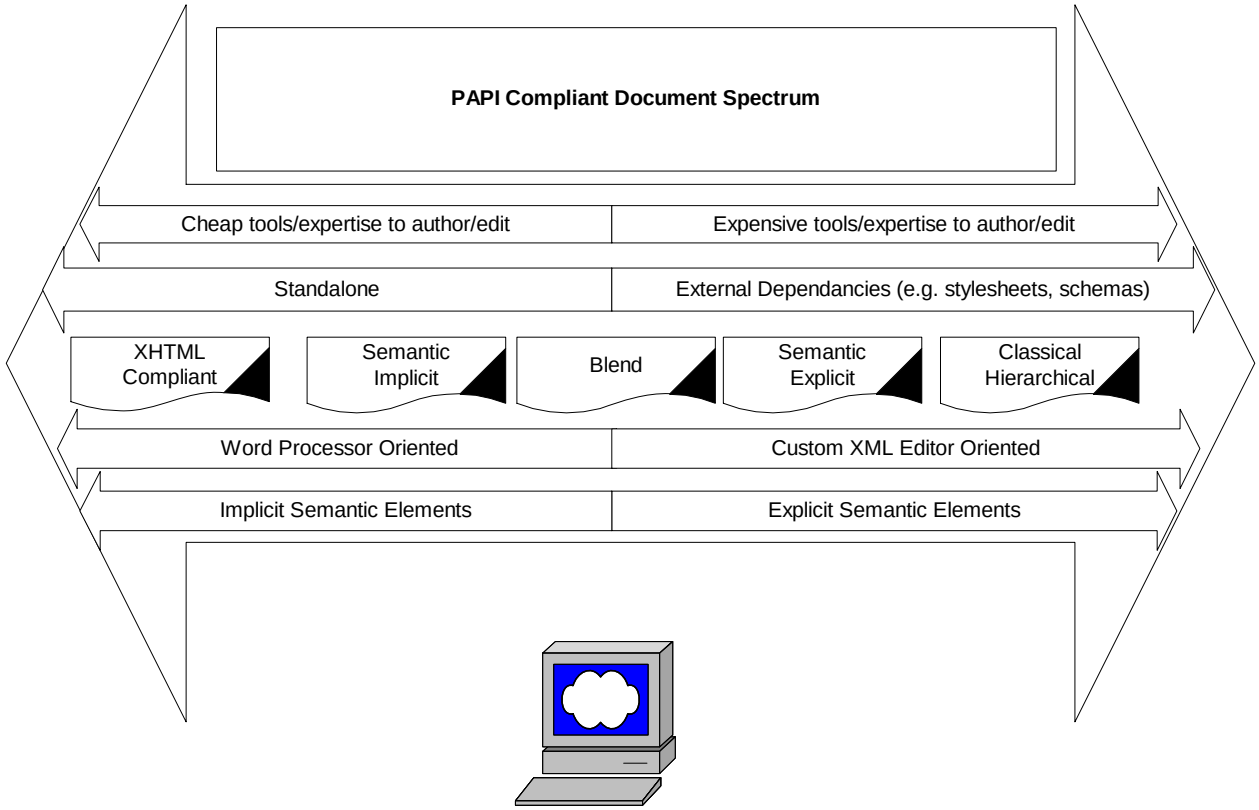
The PAPI Contiuum Model – Semantic Implicit

- PAPI-compliant XML (PAPI document type – not XHTML document type)
- Represent semantic elements using class attributes of elements from the XHTML tagset. Samples files of this form are in the `samples/semantic_implicit` subdirectory.

PAPI Semantic Implicit

```
<MISC:MISC>
<MISC:Header>
<PAPI:Metadata>…</PAPI:Metadata>
</MISC:Header>
<MISC:Body>
<xhtml:p class="DocTitle">Extraits de la
  constitution</xhtml:p>
<xhtml:p class="MainTitle">TITRE
  III<xhtml:br/>DU POUVOIR
  LEGISLATIF</xhtml:p>
```

PAPI Blend



WP or XML editor

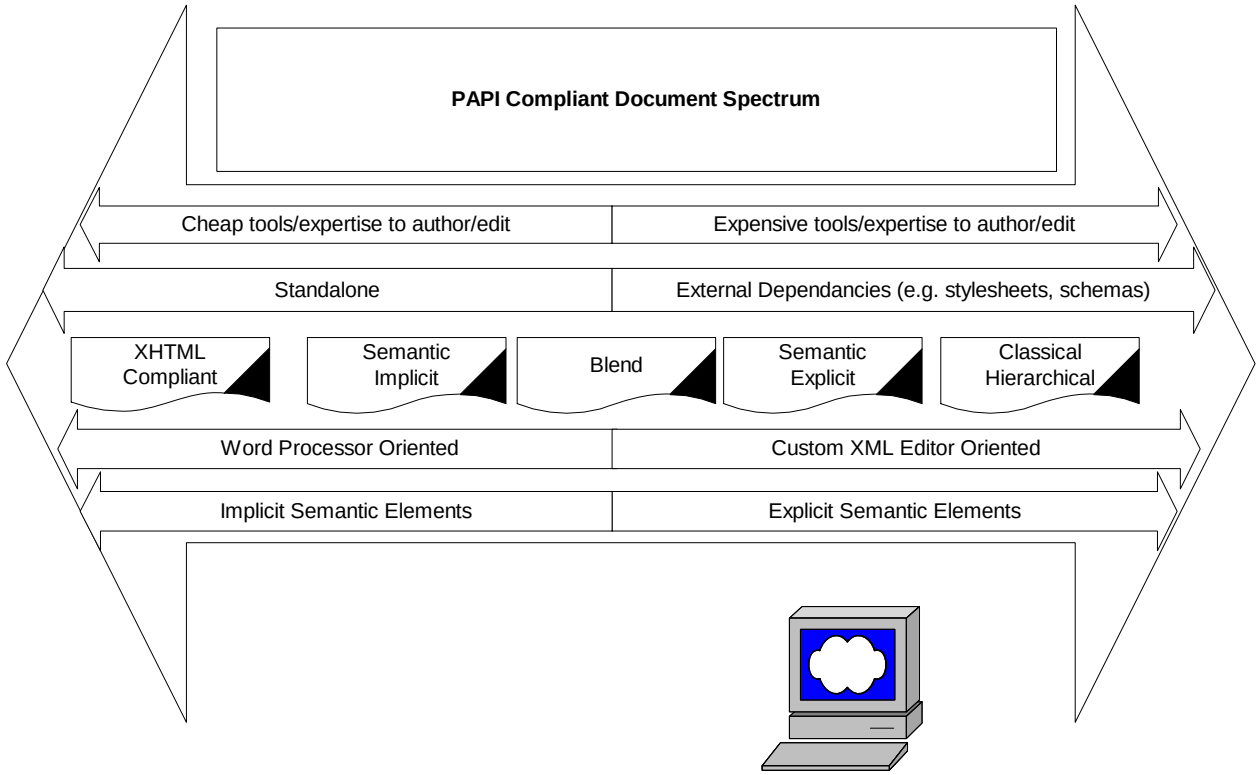
The PAPI Continuum Model - Blend

- PAPI-compliant XML (PAPI document type – not XHTML document type)
- Mix the use of XHTML elements for presentation effects with PAPI-specific XML elements.
- Samples files of this form are in the `samples/blend` subdirectory.

PAPI BLEND

```
<MISC:Body>
<xhtml:p><xhtml:b>Extraits de la
  constitution</xhtml:b></xhtml:p>
<xhtml:p style="text-
  align:center"><xhtml:b>TITRE
  III<xhtml:br/>DU POUVOIR
  LEGISLATIF</xhtml:b></xhtml:p>
```

PAPI Semantic Explicit



XML editor

The PAPI Continuum Model – Semantic Explicit

- PAPI-compliant XML (PAPI document type – not XHTML document type)
- Represent semantic elements using custom XML elements. Sample files of this form are in the `samples/semantic_explicit` subdirectory.

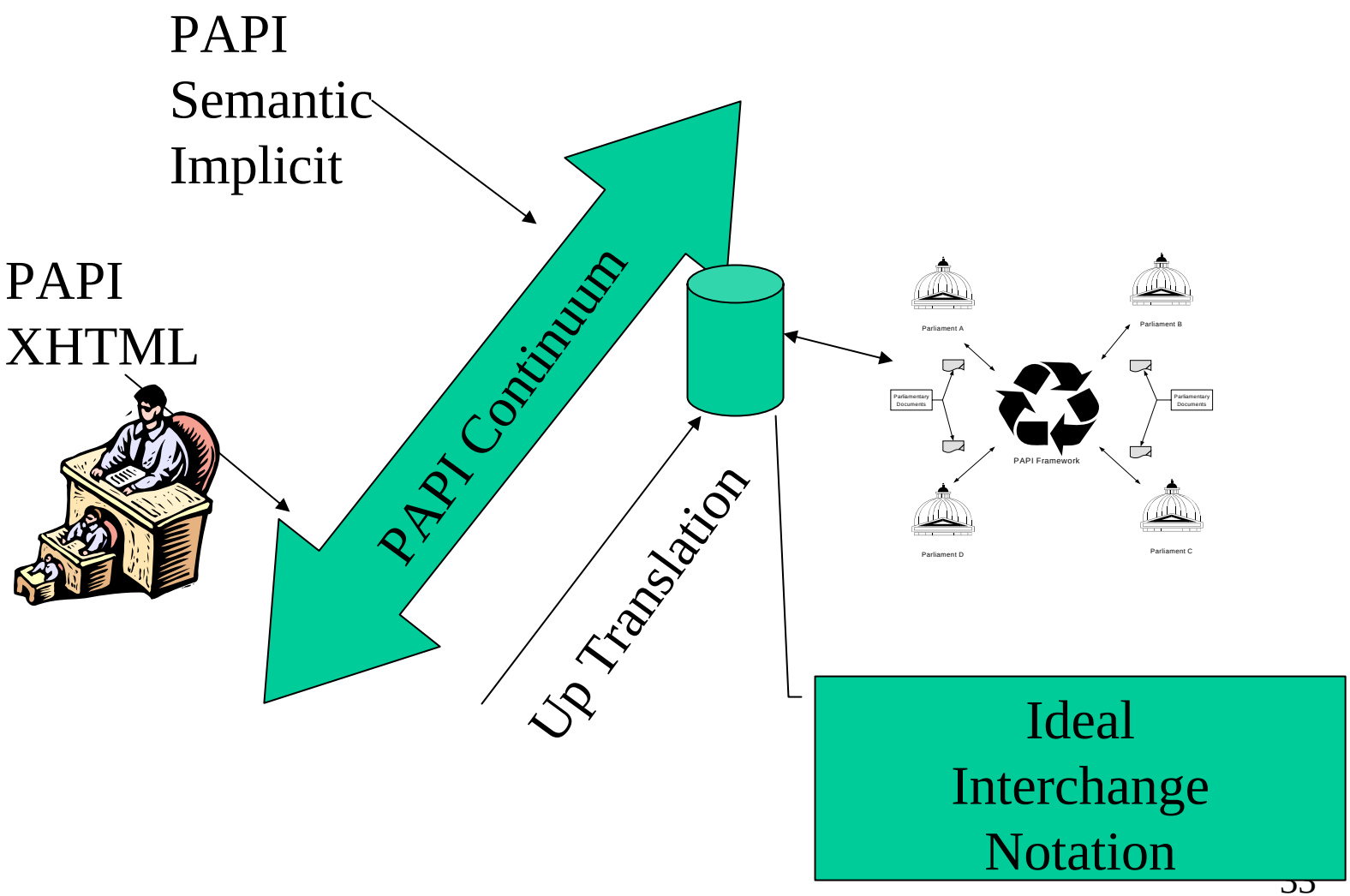
Semantic Explicit

```
<PAPI:Metadata>
<PAPI:Property
  Name="PAPI.NormativeResourceName">
cmr_misc_extraits_de_la_constitution_fr.xml
</PAPI:Property>
</PAPI:Metadata></MISC:Header>
<MISC:Body>
<MISC:DocTitle>
Extraits de la constitution
</MISC:DocTitle>
<MISC:MainTitle>
TITRE III<xhtml:br/>DU POUVOIR LEGISLATIF
</MISC:MainTitle>
```

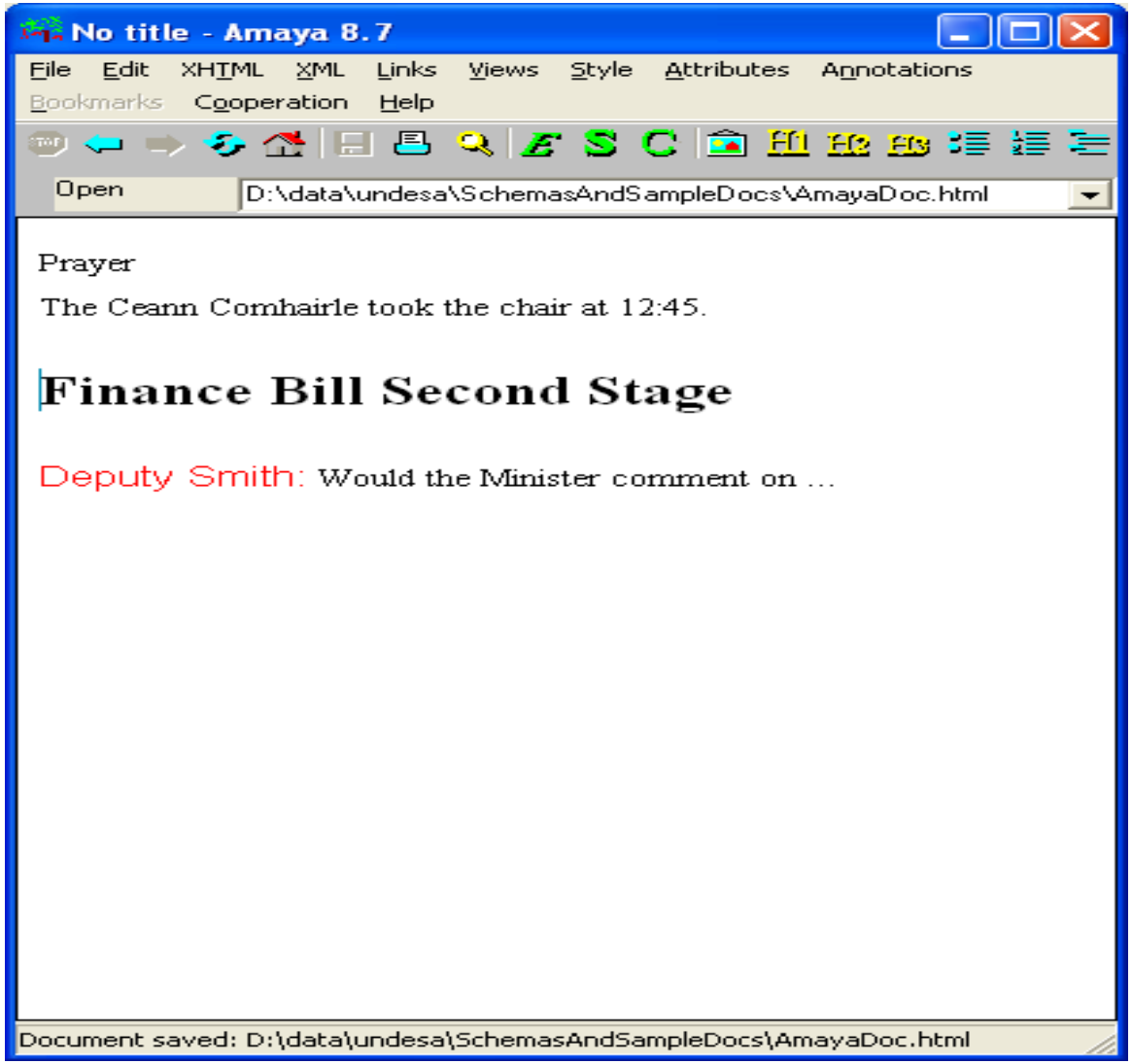
PAPI Transformations and equivalences

- Semantic implicit and semantic explicit are equivalent:
  `<p class="DecreeDetail"> == <DecreeDetail>`
- PAPI exploits this equivalence to simplify content origination – authoring/editing

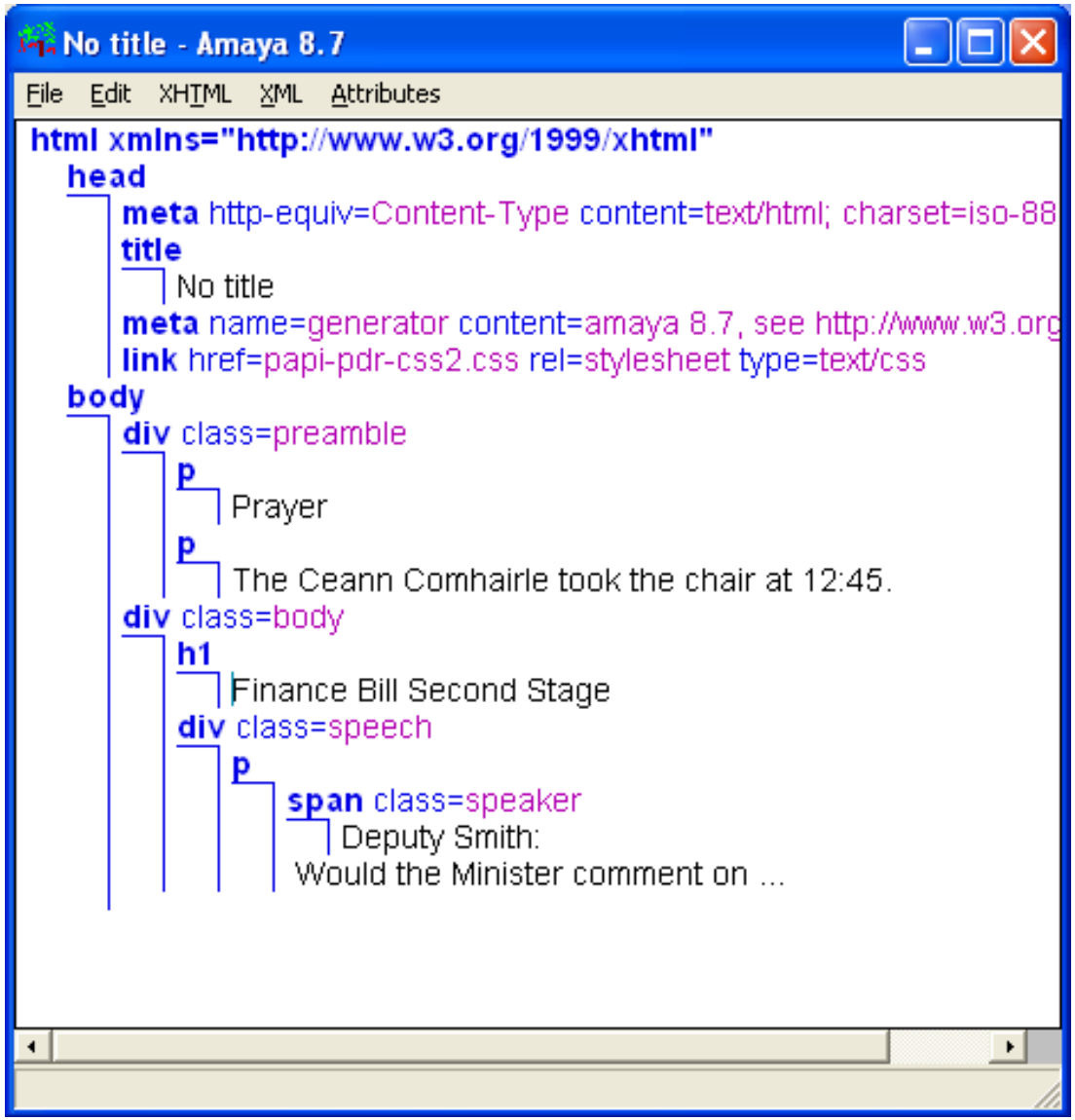
PAPI Transformations



PAPI XHTML


No title - Amaya 8.7
File Edit XHTML XML Links Views Style Attributes Annotations
Bookmarks Cooperation Help
Open
D:\data\undesa\SchemasAndSampleDocs\AmayaDoc.html
Prayer
The Ceann Comhairle took the chair at 12:45.
Finance Bill Second Stage
Deputy Smith: Would the Minister comment on ...
Document saved: D:\data\undesa\SchemasAndSampleDocs\AmayaDoc.html

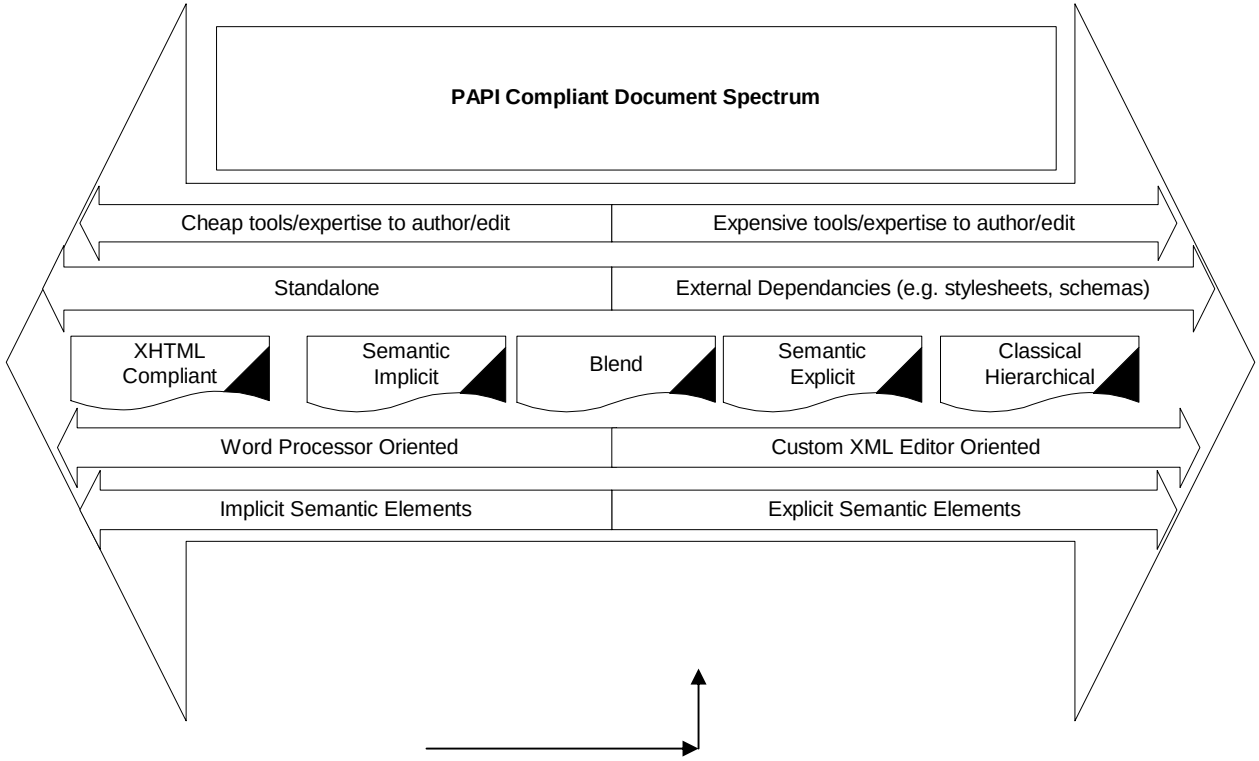
PAPI XHTML


No title - Amaya 8.7
File Edit XHTML XML Attributes
html xmlns="http://www.w3.org/1999/xhtml"
head
meta http-equiv=Content-Type content=text/html; charset=iso-88
title
No title
meta name=generator content=amaya 8.7, see http://www.w3.org
link href=papi-pdr-css2.css rel=stylesheet type=text/css
body
div class=preamble
p
Prayer
p
The Ceann Comhairle took the chair at 12:45.
div class=body
h1
Finance Bill Second Stage
div class=speech
p
span class=speaker
Deputy Smith:
Would the Minister comment on ...

PAPI Transformations

- Up translations
  - Some semi-automated. Some fully automated.
- Down translations
  - Fully automated e.g. publishing systems
- Cross translations
  - Semantic implicit <-> semantic explicit
  - Flat headings <-> hierarchical headings

PAPI Transformations



Normative version

End User Benefit

Engineering Benefit

PAPI Persistence

- For long term storage & archiving, favour semantic implicit/blend PAPI
- Some implementors may prefer to use semantic explicit
- Some my prefer to use classical hierarchical
  - Fine as long as two-way transformations are fully automated
- Sample transforms included with the PAPI materials

The real world – common containership elements

- Hierarchical Structure of Documents
  - Part/Chapter or Chapter/Part?
- In Namibia, the words **Chapter** and **Part** are used in an opposite nesting sense to the way they are used in South Africa
- **Chapter** within **Part** in South Africa. **Part** within **Chapter** in Namibia.
- Chapter and Part are not used in semantically equivalent ways across Parliaments.

The real world – common containership elements

In Sierra Leone, a self-contained numbered unit of legislation is known as a **Section**. In Madagascar, a **Section** is a container element for **Clauses**.

-> **Sections** in Sierra Leone are not semantically equivalent to **Sections** in Madagasgar.

In Algeria, **Books** are used as a container element. Sometimes units of numbered legislation are **sections**, other times they are **articles**. Sometimes they are c**lauses**. Sometimes **section** is used as a container for **clauses**. Other times, **sections** *are* **clauses**.

In Nigeria, **sections** are used like **sections** in South Africa but they do not have numbers…

The real world – common containership elements

- We need an extensible, pragmatic way of capturing part/chapter/clause/section containership information, carefully differentiating between what has pan-African semantics from what has local semantics.

Nested versus flat headings levels

- Flat headings – Word processors, HTML/XHTML
- Hierarchical headings – “classical” SGML/XML model

Nested versus flat heading levels

- Equivalence of “flat” and “hierarchical” models.
- Flat Model:

```
<ACT:ACT>
    <h1 class=”Book”>This is a book</h1>
    <h2 class=”Part”>This is a part</h2>
    <h3 class=”Chapter”>This is a chapter</h3>
    <h4 class=”Clause”>This is a clause</h4>
</ACT:ACT>
```

Nested versus flat heading levels

- Equivalent to hierarchical model - 1:

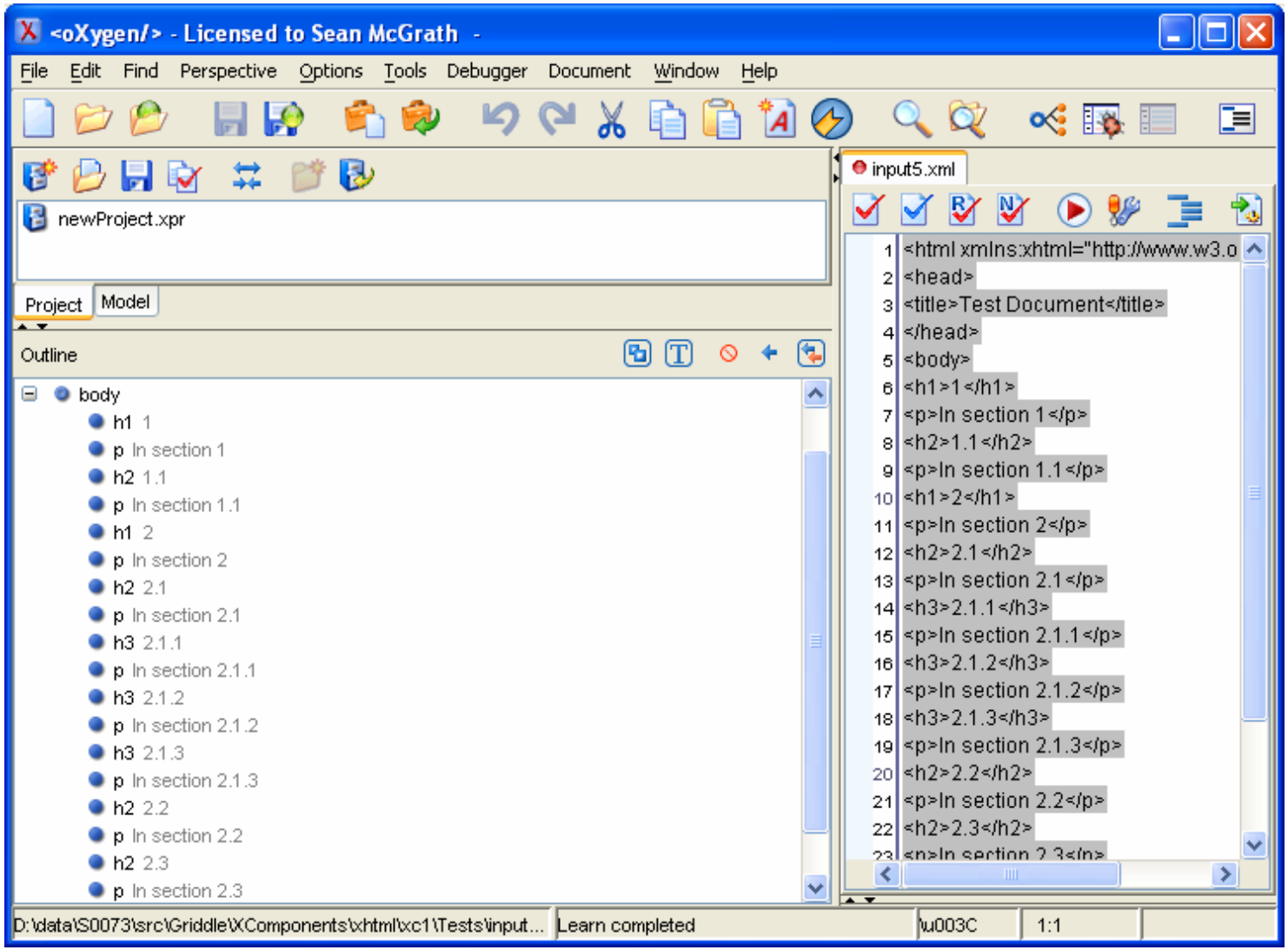
```
<ACT:ACT>
 <h1 class=”Book”><title>This is a book</title>
  <h2 class=”Part”><title>This is a part</title>
   <h3 class=”Chapter”><title>This is a
                                chapter</title>
    <h4 class=”Clause”><title>This is a
                                 clause</title>
    </h4>
   </h3>
  </h2>
</h1>
</ACT:ACT>
```

Nested verus flat heading levels

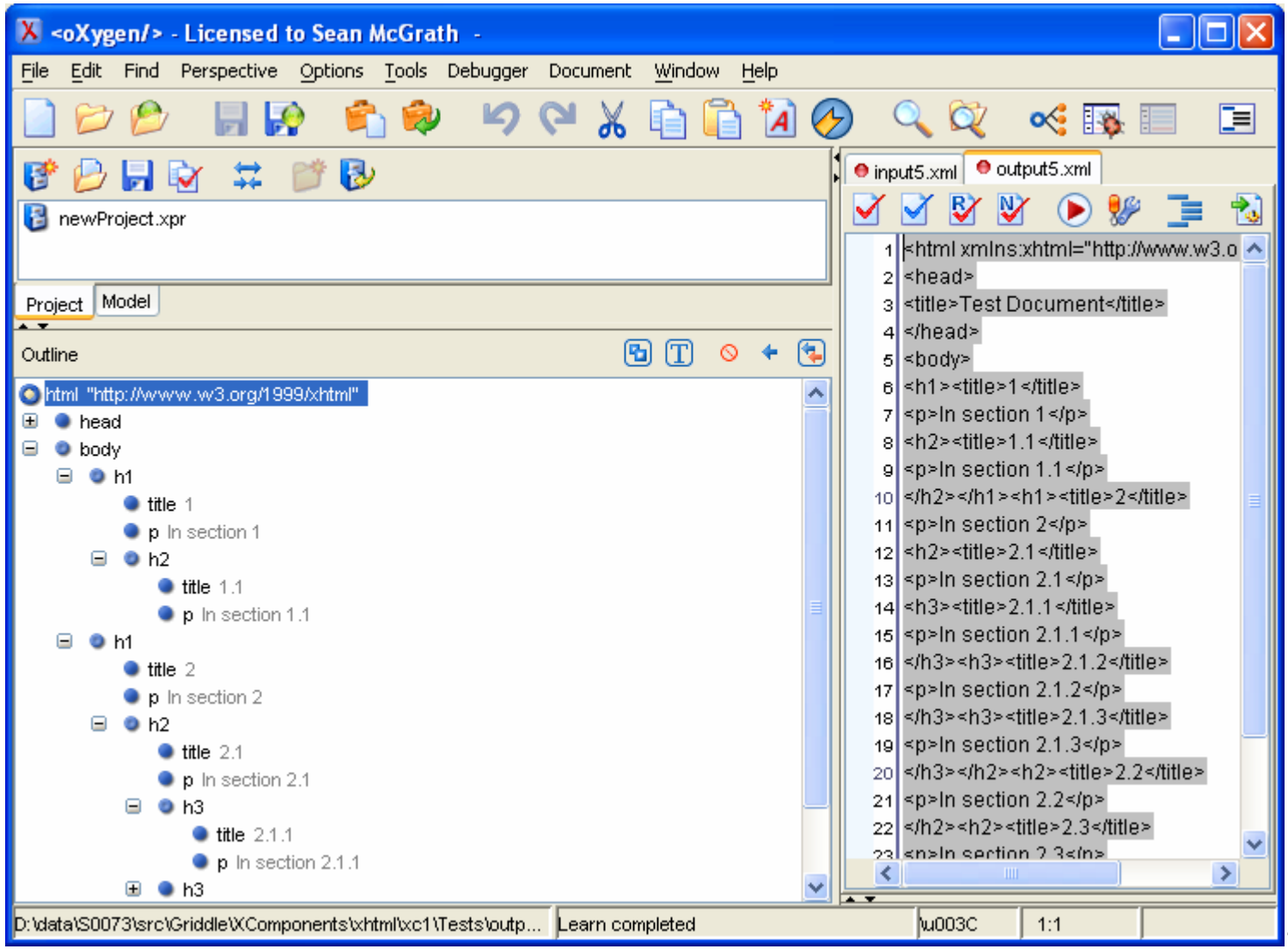
- Equivalent to hierarchical model – 2:

```
<ACT:ACT>
 <ACT:Book><title>This is a book</title>
  <ACT:Part><title>This is a part</title>
   <ACT:Chapter><title>This is a chapter</title>
    <ACT:Clause><title>This is a clause</title>
    </ACT:Clause>
   </ACT:Chapter>
  </ACT:Part>
 </ACT:Book>
</ACT:ACT>
```

PAPI Transformations



PAPI Transformations



The PAPI Metadata Model

- Property/Value Pairs
- Reversed domain names for uniqueness
- Some reserved naming patterns

Naming Metadata Terms

Tanzania National Assembly

URI : http://www.bunge.go.tz/bunge/bunge.asp

Reversed Domain Name : “tz.go.bunge.”

Metadata Example:” tz.go.bunge.ReferringMinistry”

Rwanda National Parliament

URI : http://www.rwandaparliament.gov.rw/

Reversed Doman Name : “rw.gov.rwandaparliament”

Metadata Example :
“rw.gov.rwandaparliament.CommitteeChairperson”

Reserved Prefixes

- PDR = Parliamentary Debate Record Metadata
- PAPI = Pan African Interoperability Framework Metadata
- ACT = Enacted Legislation Metadata
- BILL = Pre-enacted Legislation Metadata

Examples

- PAPI.NormativeResourceName
- PAPI.EurovocDescriptor
- PAPI.DC.*

PAPI Metadata Markup

```
<ACT:Header>
 <PAPI:Metadata>
  <PAPI:Property
   Name="PAPI.NormativeResourceName">
   dze_le_code_de_famille_fr.xml
  </PAPI:Property>
  <PAPI:Property
   Name="org.dz-apn.date">
   9 JUIN 1984
  </PAPI:Property>
 </PAPI:Metadata>
</ACT:Header>
```

Eurovoc Africa

```
<PAPI:Property
Name="PAPI.EurovocDescriptor">
Community programme;
Internet ;
information technology ;
new technology ;
data protection
</PAPI:Property>
```

Hypertext in PAPI

- URIs
  - A normative resource name
    - E.g. gmn_pdr_19062003_en.xml
  - An optional resolution scheme
    - E.g. http://
  - An optional fragment identifier
    - E.g. #motion1
- **“http://gm_pdr_19062003_en.xml#motion1”**

Hypertext in PAPI

- Mapping Table Hypertext Resolution

```
E.g. transform this:
<a href="gmn_pdr_19062003_en.xml#motion1">
Second meeting (Part 1), 2003 legislative
  year. Motion 1.
</a>
```

  – to this:

```
<a
  href="http://wwww.statehouse.gm/gmn_pdr_1906
  2003_en.xml#motion1">
Second meeting (Part 1), 2003 legislative
  year. Motion 1.
</a>
```

Hypertext in PAPI

- “gmn” -> http://wwww.statehouse.gm/
- “sle” -> “http://www.sierra-leone.org/”
- Etc.
- See PAPI documentation for full list.

Hypertext in PAPI

- URN based hypertext resolution

E.g. transform this:

Second meeting (Part 1), 2003 legislative year. Motion 1.

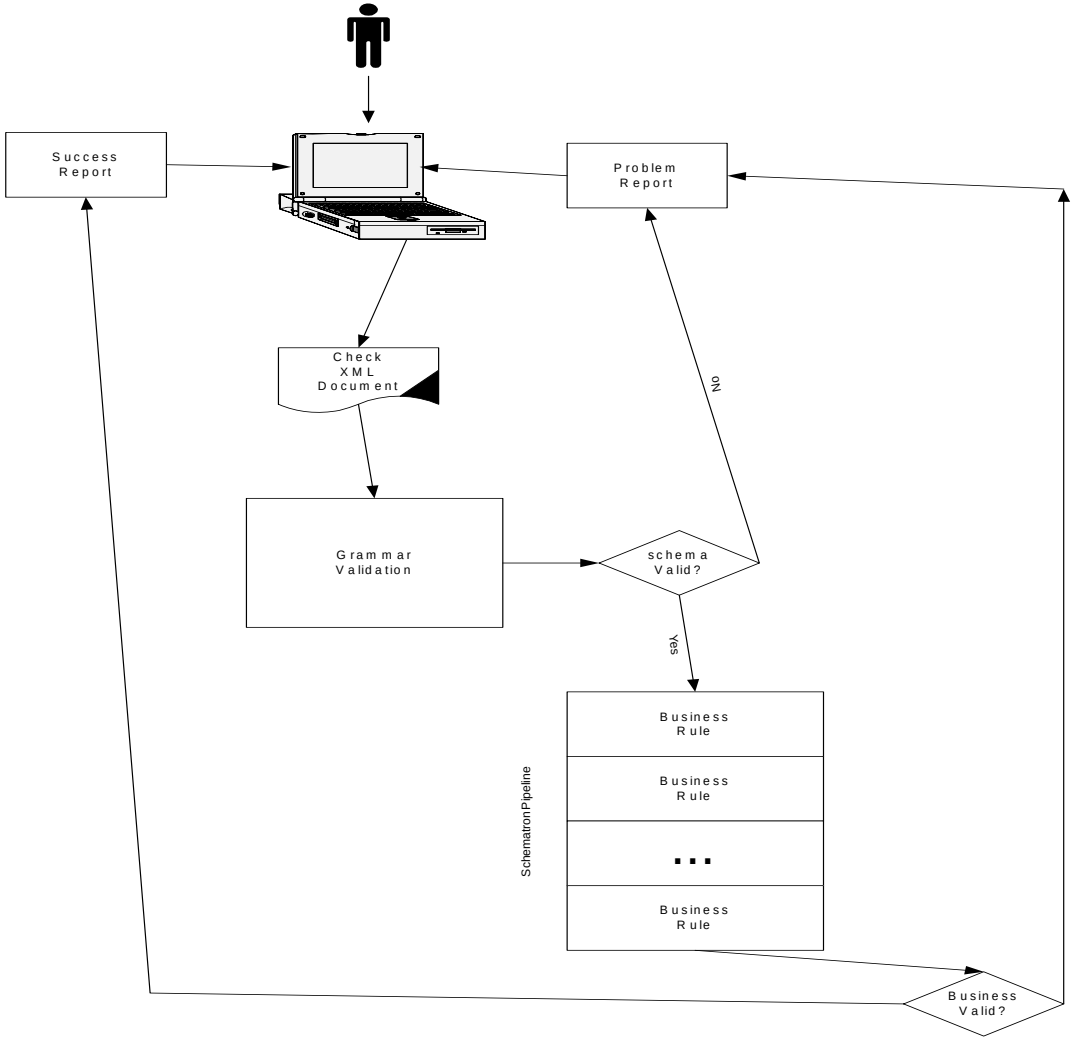
To this:

Second meeting (Part 1), 2003 legislative year. Motion 1.

Hypertext in PAPI

- Anchor naming conventions:
  - Algeria:
    - “art” <N> where N is an integer. e.g.
    - <xhtml:a name="art2"/>
  - Sierra Leone:
    - “secN” where N is the number of the section.
    - <xhtml:a name="sec2"/>

PAPI Document Validation & Business Rules



Schematron

- ISO/IEC 19757 part 3
- XSLT based
- Assertions -> XSLT stylesheet -> validation of business rules

PAPI Business Rule Example

Rule : “All PAPI documents must specify a normative resource name”

```
<sch:rule context="//PAPI:Metadata">
   <sch:assert test=“
count(
PAPI:Property[@Name='PAPI.NormativeResourceName']
) = 1
">
     Expected to find exactly 1 PAPI metadata item named
   PAPI.NormativeResourceName.
   </sch:assert>
 </sch:rule>
```

PAPI Business Rule Example

Rule : A speaker element must have content. Its parent must be a paragraph and must be the first child of that paragraph.

```
<sch:rule context="//PDR:Speaker">
  <sch:assert test="count(text())>0">
    Speaker name cannot be blank.
  </sch:assert>
  <sch:assert test="parent::xhtml:p">
        Parent of PDR:Speaker must be xhtml:p
  </sch:assert>
  <sch:assert test="count(../preceding-sibling::*)=0">
        Parent xhtml:p must be first child of PDR:Speech
  </sch:assert> </sch:rule>
```

Naming conventions

Resource_Name := Country_Code “_” Document_Detail “_” [Authorities “_”] Language_Codes [Incrementing_Number] [“.” Extension]

Authorities := String

Country_Code := “dza” | “cmr” ... (An ISO 3166 country code)

Document_Detail := Debate_Detail |Act_Detail |Bill_Detail |Report_Detail |Constitution_Detail |Proclaimation_Detail |Miscellaneous_Detail

|Debate_Detail := “pdr” “_” Date
Act_Detail := “act” “_” ((Act_Number “_” Act_Year) | String)
Bill_Detail := “bill” “_” ((Bill_Number “_” Bill_Year “_” Bill_Stage) | String)
Miscellaneous_Detail := “misc” “_” String
Act_Number := <Number>
Act_Year := YYYY
Bill_Number := <Number>
Bill_Year := YYYY
Bill_Stage := String

Date := DDMMYYYY (Date in Gregorian Calendar notation)

Language_Codes := Language_Code (“_” Language_Code)*
Language_Code := “en” | “fr” .. (An ISO 639 Language Code)

Extension := “.xml” | “.html” | “.jpg” | ...

IncrementingNumber := “1” “2” etc. (Used for graphic images only)

String := Any sequence of Unicode characters complying to the PAPI
Unicode Profile, excluding white space and “_”

Naming examples

- **dza_pdr_21122004_en.xml**
  - Algerian parliamentary debate record, 21st December 2004. English version. XML notation.
- **nam_bill_19_2003_1_en.xml**
  - Namibia Bill number 19 of 2003, first stage, English version. XML notation.
- **zaf_bill_12_2005_2_MinistryOfAgriculture_en.xml**
  - South Africa. Bill 12 of 2005 (second stage), issued by the Ministry Of Agriculture. English version. XML.

PAPI Summary

- Continuum model designed to simplify the process of getting content into PAPI compliant form
- An evolutionary, rather than “big bang” approach to semantic interoperability
- Extensible metadata framework incorporating Eurovoc
- Extensible business-rule based docunent validation
- Naming conventions for information assets, metadata and hypertext

Status and Roadmap

- Framework has been baselined
- Sample documents, schemas, business rules etc. available
- Full documentation available
- Governance structure under discussion
- Comments and contributions welcome