



The Next Millennium Document Format

XML Prague 15.02.2020

The Next Millennium Document Format

Collection of Challenges & Solutions

- Many possible challenges, but let's focus on three:
 - Full collaboration support by design
 - Agile (Open) Standardization
 - Reactive Interoperable Layout by design

The Next Millennium Document Format

Collection of Challenges & Solutions

- First Challenge - Collaboration:
 - Multiple People working on Single Document Output using different Applications
 - Unsolved by Design by any File/Document Format
- Why a Challenge?
 - Design of many File Formats based on the 80ths
 - Single PC using Modem & Floppy Discs
- Do Email and Cloud Storage help?

The Next Millennium Document Format

Collection of Challenges & Solutions

- Key Question in Collaboration:
What have you changed? (to merge)

The Next Millennium Document Format

Collection of Challenges & Solutions

- Key Question in Collaboration:
What have you changed? (to merge)
- Two semantic equal representations:
OpenDocument Text ↔ **ODT User Changes (JSON)**

Character Attributes

Examples

Bold

ACM POSTER...

Italic

Model is viewed...

Underline

Changes are shown...

```
{ "name": "addParagraph", "start": [4] },  
{ "name": "addText", "start": [4,1], "text": "ACM" },  
{ "name": "addText", "start": [4,12], "text": " POSTER" },
```

ODT ⇔ Changes

sponsored by PrototypeFund



See <https://github.com/tdf/odftoolkit/tree/1.0.0> SNAPSHOT

Interoperable Collaboration

Exchanging **ODF Changes**



ODF Application

 ODF 4th paragraph new



ODF Application

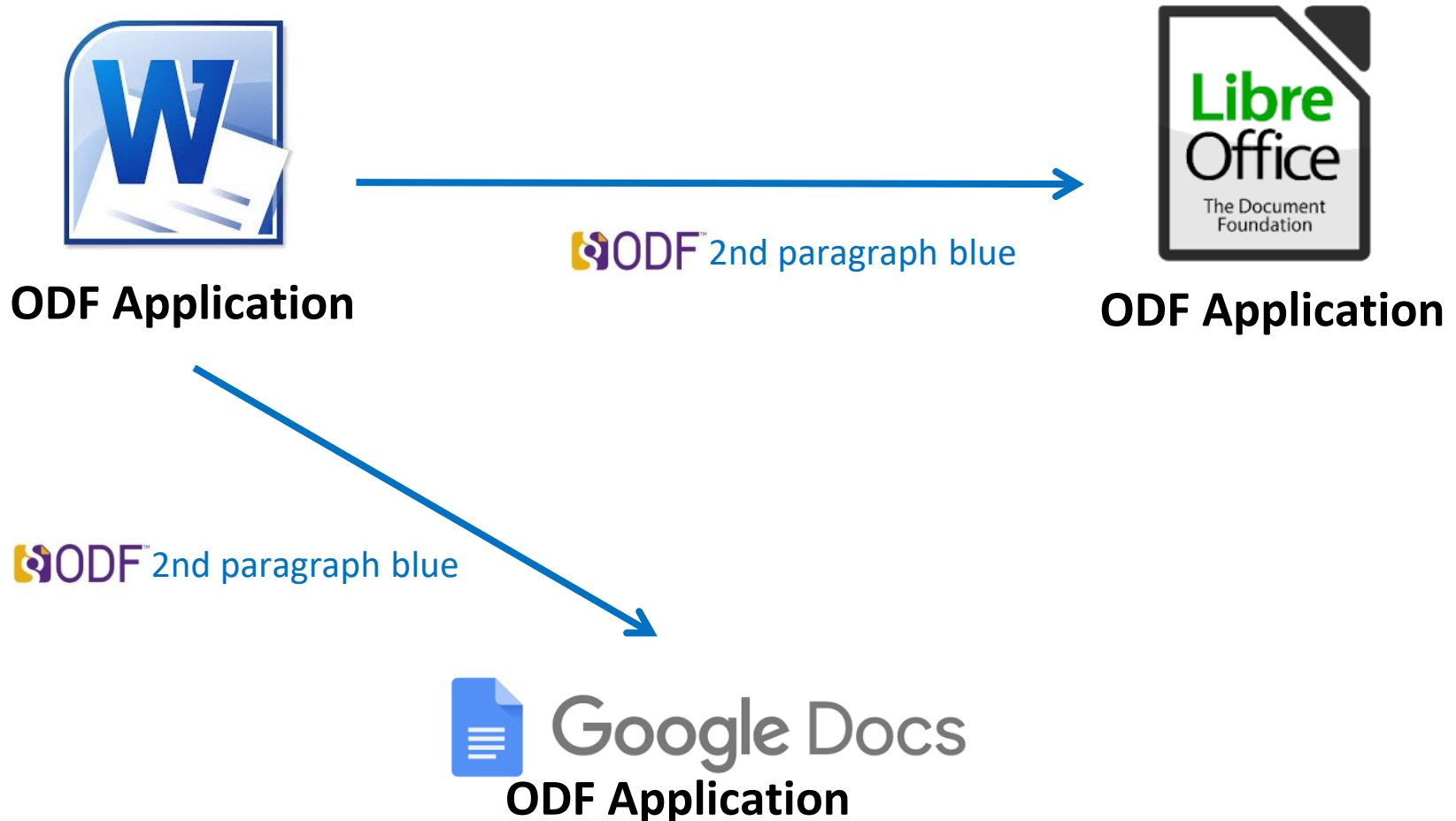
 ODF 4th paragraph new



Google Docs
ODF Application

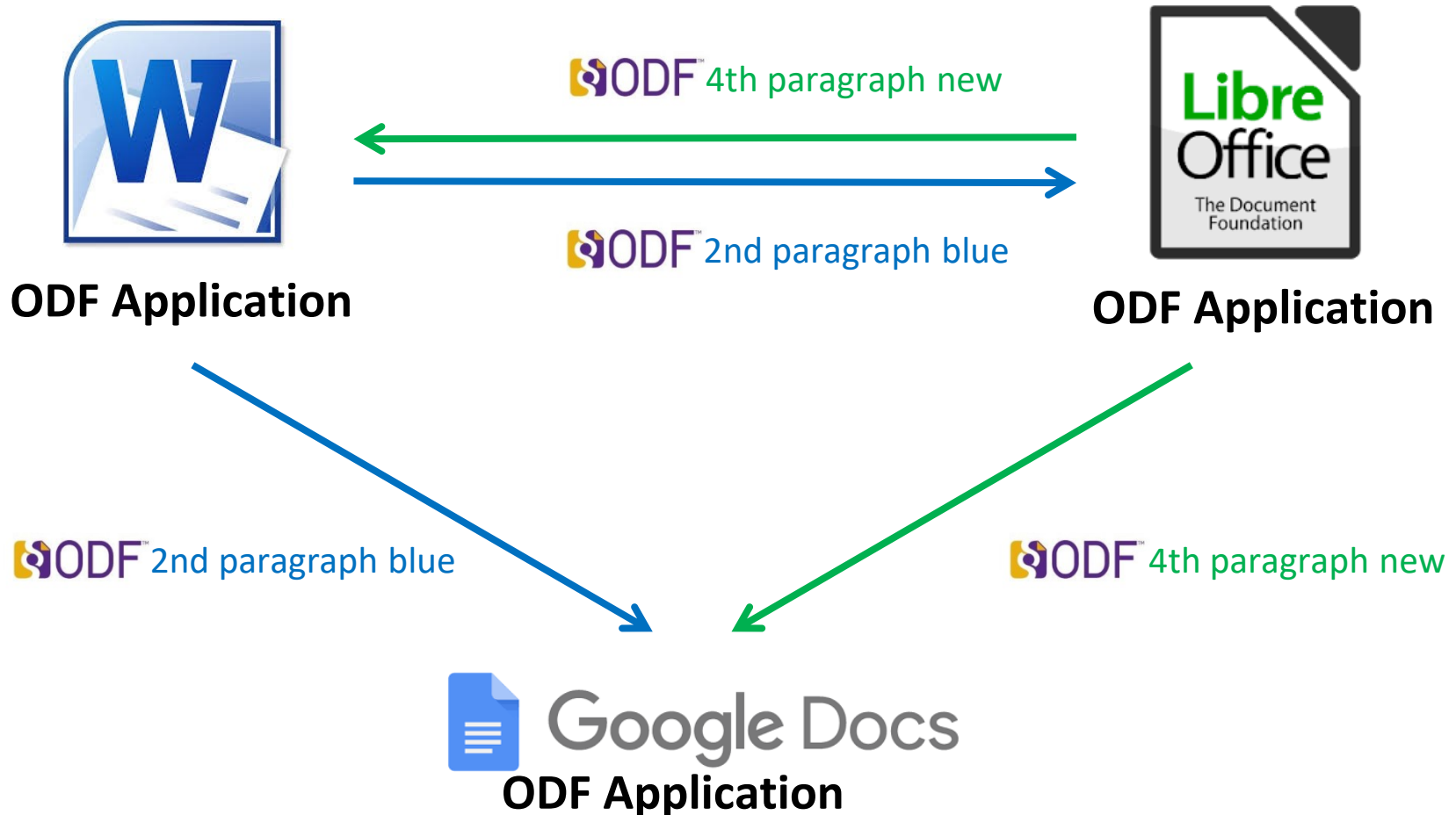
Interoperable Collaboration

Exchanging **ODF Changes**



Interoperable Collaboration

Exchanging **ODF Changes**



Example Document

Exchanging **ODF Changes**

a	b
c	d



Semantic tree:

The underlying **XML tree** is being mapped to larger semantic pieces represented as **Semantic Tree**. Changes refer to those user objects.

a	b
c	d



NOTE:

- Text editor as Emacs does not „see“ the table nor the image!
- The „**W**“ character is for **LibreOffice** at position „3/1“
- The „**W**“ character is for **Emacs** at position „1/1“ (or if table shown as CSV than „2/1“)



Emacs is exciting, but...

Interoperable Collaboration

Exchanging **ODF Changes**



CKEditor 5
ODF Application



ODF Application

 **ODF™** ADD „Hello“ @3/1

 **ODF™** ADD „Hello“ @3/1

ODF Feature Bridge

„**Feature bridge**“ not only temporally adds/deletes changes, but maps them to other „**change dialect**“.
(more detailed view on next 2 slides)

The Next Millennium Document Format

CKEditor 5 – Proof of Concept

- Build my CKEditor 5 example to view CK5 changes:

git clone <https://github.com/svanteschubert/ckeditor5-build-classic>

npm install

npm run build

Open local editor in sample and view changes in console:

./sample/index.html

The Next Millennium Document Format

CKEditor 5 – Results in Chrome Console

- `enventinfo[object Object] args[{"baseVersion":12,"position":{"root":"main","path":[0,6],"stickiness":"toNone"},"nodes":[{"data":"X"}],"shouldReceiveAttributes":true,"__className":"InsertOperation"}]`
- `index.html:48 enventinfo[object Object] args[{"baseVersion":13,"range":{"start":{"root":"main","path":[0,3],"stickiness":"toNext"},"end":{"root":"main","path":[0,7],"stickiness":"toPrevious"}},"key":"italic","oldValue":null,"newValue":true,"__className":"AttributeOperation"}]`



BASIC TECHNIQUES

The 1 x 1 of Changes / Operations

The Next Millennium Document Format

The shell game – keep track of your changes

- You work on the 3rd paragraph
- Someone inserts a new 2nd paragraph
- Your paragraph becomes the 4th paragraph

The Next Millennium Document Format

One Document – Many ways to create it...

„ABC“

Final Document

The Next Millennium Document Format

One Document – Many ways to create it...

User 1
changes

ADD „A“ @1

Current
Document

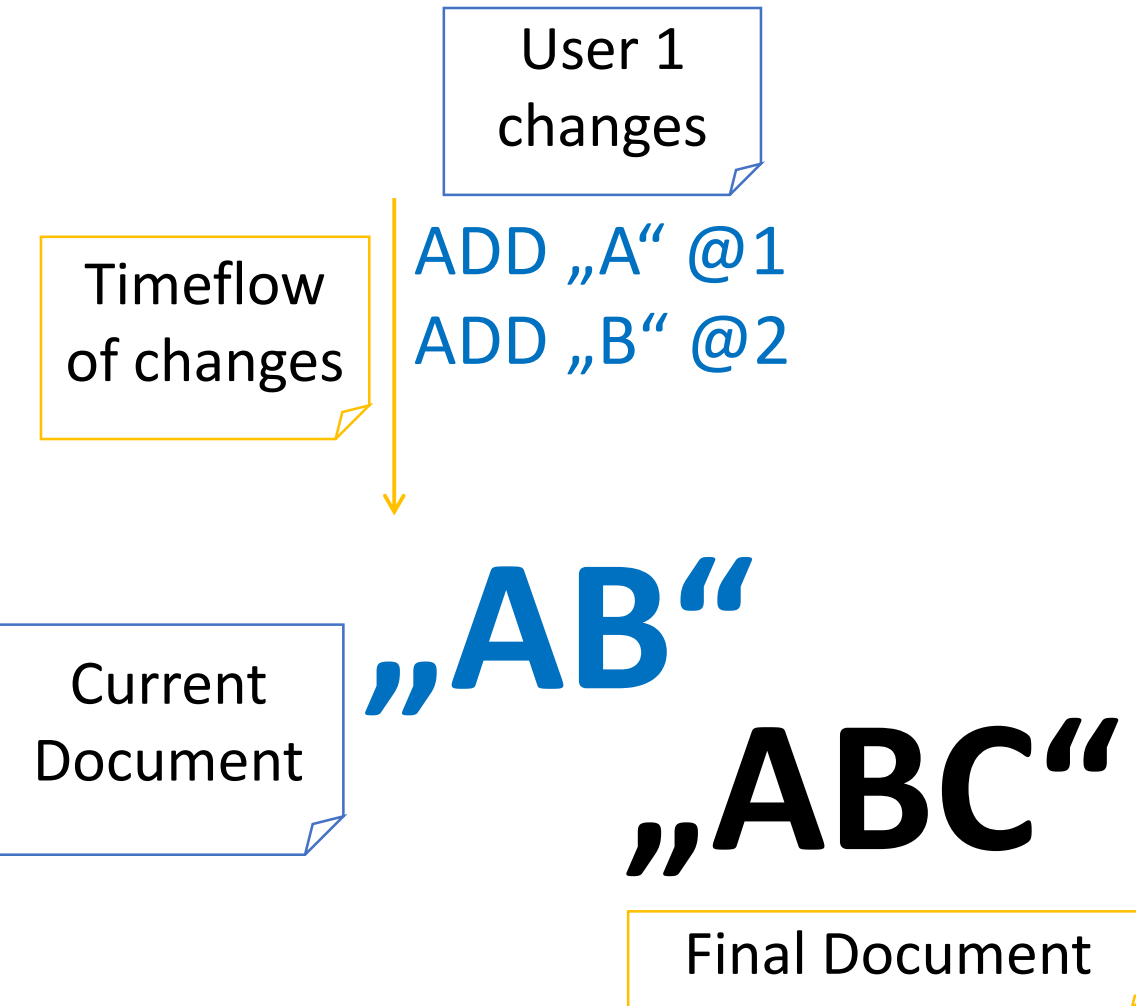
„A“

„ABC“

Final Document

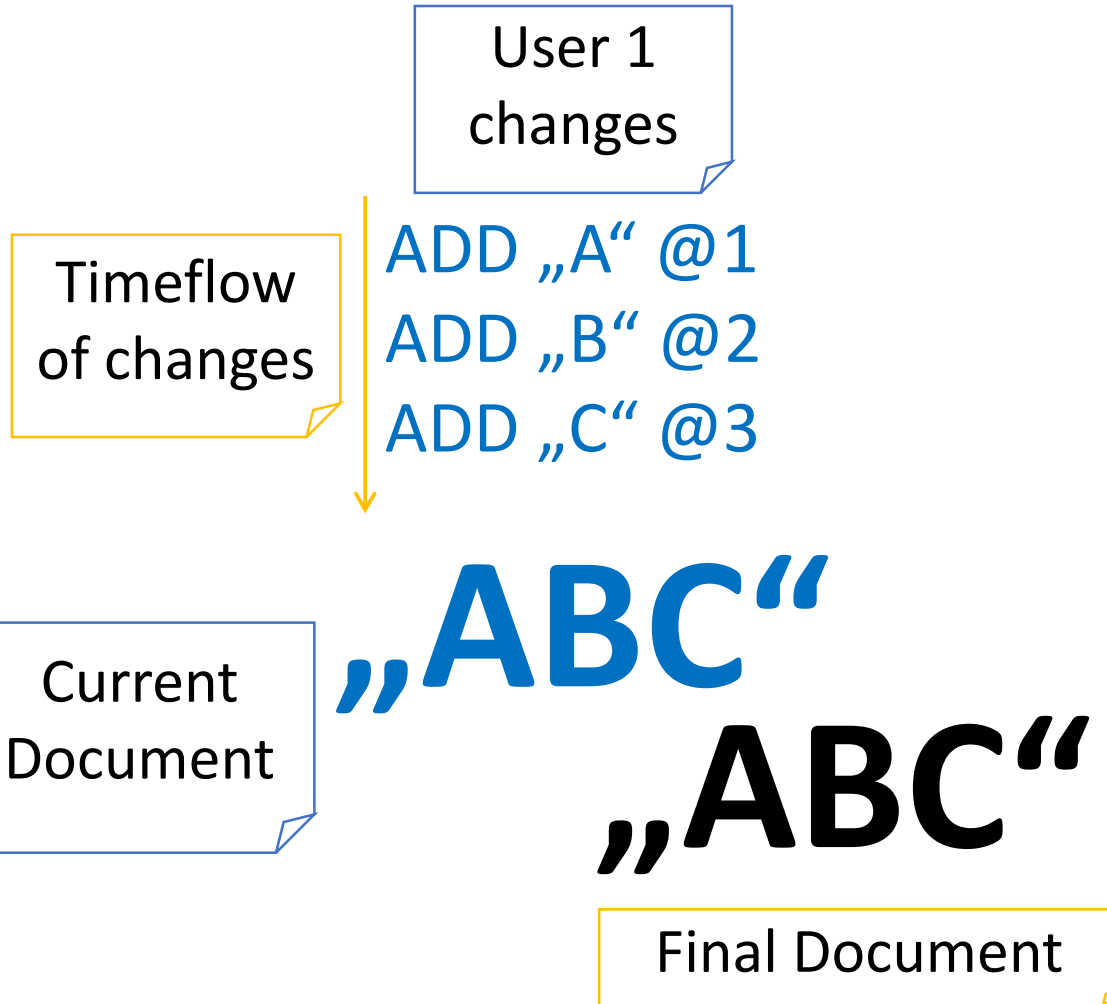
The Next Millennium Document Format

One Document – Many ways to create it...



The Next Millennium Document Format

One Document – Many ways to create it...



The Next Millennium Document Format

One Document – Many ways to create it...

ADD „A“ @1

ADD „B“ @2

ADD „C“ @3

„ABC“

Final Document

The Next Millennium Document Format

One Document – Many ways to create it...

User 2
changes

ADD „A“ @1
ADD „B“ @2
ADD „C“ @3

ADD „C“ @1

„C“
„ABC“

Final Document

The Next Millennium Document Format

One Document – Many ways to create it...

User 2
changes

ADD „A“ @1
ADD „B“ @2
ADD „C“ @3

ADD „C“ @1
ADD „B“ @1

„BC“

„ABC“

Final Document

The Next Millennium Document Format

One Document – Many ways to create it...

ADD „A“ @1

ADD „B“ @2

ADD „C“ @3

ADD „C“ @1

ADD „B“ @1

ADD „A“ @1

„ABC“

„ABC“

Final Document

The Next Millennium Document Format

One Document – Many ways to create it...

ADD „A“ @1		ADD „C“ @1
ADD „B“ @2	↔	ADD „B“ @1
ADD „C“ @3		ADD „A“ @1

QUESTION:

How transforming one into the other? 🤖

„ABC“

The Next Millennium Document Format

One Document – Many ways to create it...

ADD „A“ @1
ADD „B“ @2
ADD „C“ @3

ADD „C“ @1
ADD „B“ @1
ADD „A“ @1

Move C change
from top to
bottom

Timeflow
of changes

„ABC“

Atomic operation is
switching two
adjacent changes

The Next Millennium Document Format

One Document – Many ways to create it...

ADD „A“ @1
ADD „B“ @2
ADD „C“ @3

ADD „B“ @1
ADD „C“ @2
ADD „A“ @1

Timeflow
of changes

„ABC“

Position of C changes
as B was inserted now
earlier, when the two
changes are being
switched!

The Next Millennium Document Format

One Document – Many ways to create it...

ADD „A“ @1

ADD „B“ @2

ADD „C“ @3

ADD „B“ @1

ADD „A“ @1

ADD „C“ @3

„ABC“

The Next Millennium Document Format

One Document – Many ways to create it...

Move B change
from top to
middle

ADD „A“ @1

ADD „B“ @2

ADD „C“ @3

ADD „B“ @1

ADD „A“ @1

ADD „C“ @3

„ABC“

The Next Millennium Document Format

One Document – Many ways to create it...

Now both lists
of changes are
identical

ADD „A“ @1

ADD „B“ @2

ADD „C“ @3

ADD „A“ @1

ADD „B“ @2

ADD „C“ @3

We might call
the blue list
normalized!

„ABC“

The Next Millennium Document Format

Deletion

How
deleting „B“?

ADD „A“ @1

ADD „B“ @2

ADD „C“ @3

„ABC“

The Next Millennium Document Format

Deletion

Error!!

ADD „A“ @1

ADD „C“ @3

NO GAP
allowed!

NO deletion of a change
from within the stack!

„ABC“

The Next Millennium Document Format

Deletion

ADD „A“ @1

ADD „B“ @2

ADD „C“ @3

B not the last change!
B influences C!

„ABC“

The Next Millennium Document Format

Deletion

ADD „A“ @1

ADD „C“ @2

ADD „B“ @2

B now the last change!
Influences removed by
OT!

„ABC“

OT:

<http://www.codecommit.com/blog/java/understanding-and-applying-operational-transformation>

The Next Millennium Document Format

Deletion

ADD „A“ @1
ADD „C“ @2

„AC“

The Next Millennium Document Format

One Document – Many ways to create it...

ADD „A“ @1
ADD „C“ @2
ADD „B“ @2
DEL „B“ @2



ADD „A“ @1
ADD „C“ @2

Add inverse
operation!
Keep all changes!
(Blockchain Mode)

„AC“

Removes B and
keep changes
normalized.

The Next Millennium Document Format

The Miracle of Merge

- Adapt the GIT concept of Pull & Rebase
- If there are changes on the servers (client pulls)
- New changes have to moved beyond own (OT position adoptions apply)

The Next Millennium Document Format

Collection of Challenges & Solutions

- There is a subcommittee for ODF collaboration:
[OASIS ODF Advanced Documentation Collaboration SC](#)
- I became chair – king without land – as yet no implementors
- After my proposal for future change-tracking was voted over proposals from Microsoft & DeltaXML
- My key ideas:
 - Define what a change is, before try tracking it!
 - Make a change disjoint of the physical document!
- Releasing a [change “reference” implementation](#) soon...

The Next Millennium Document Format

Collection of Challenges & Solutions

- What's next for me:
 - Doing an ODF Toolkit release ;-)
 - For the future: Exchange existing ODFDOM changes implementation with a generated one
 - Generation based on additional XML grammar info
 1. What is a semantic entity: define XML nodes
 2. How can a semantic entity be changed:
Define mapping of change parameter to XML nodes
 - Defining nothing new, just reverse engineering...
 - Generate also “change specification” from it... (not new)

The Next Millennium Document Format

Collection of Challenges & Solutions

9.1.2 <table:table>

The <table:table> element is the root element for a table.

The <table:table> element is usable within the following elements: <chart:chart> [11.1](#), <draw:frame> [10.4.2](#), <draw:text-box> [10.4.3](#), <office:spreadsheet> [3.7](#), <office:text> [3.4](#), <style:footer> [16.11](#), <style:footer-left> [16.13](#), <style:header> [16.10](#), <style:header-left> [16.12](#), <table:covered-table-cell> [9.1.5](#), <table:dde-link> [14.6.4](#), <table:table-cell> [9.1.4](#), <text:deletion> [5.5.4](#), <text:index-body> [8.2.2](#), <text:index-title> [8.2.3](#), <text:note-body> [6.3.4](#) and <text:section> [5.4](#).

The <table:table> element has the following attributes: table:is-sub-table [19.655](#), table:name [19.673.13](#), table:print [19.693](#), table:print-ranges [19.694](#), table:protected [19.696.4](#), table:protection-key [19.697](#), table:protection-key-digest-algorithm [19.698](#), table:style-name [19.726.16](#), table:template-name [19.732](#), table:use-banding-columns-styles [19.736](#), table:use-banding-rows-styles [19.737](#), table:use-first-column-styles [19.738](#), table:use-first-row-styles [19.739](#), table:use-last-column-styles [19.740](#), table:use-last-row-styles [19.741](#) and xml:id [19.914](#).

The <table:table> element has the following child elements: <office:dde-source> [14.6.5](#), <office:forms> [13.2](#), <table:desc> [9.1.14](#), <table:named-expressions> [9.4.11](#), <table:scenario> [9.2.7](#), <table:shapes> [9.2.8](#), <table:table-column> [9.1.6](#), <table:table-column-group> [9.1.10](#), <table:table-columns> [9.1.12](#), <table:table-header-columns> [9.1.11](#), <table:table-header-rows> [9.1.7](#), <table:table-row> [9.1.3](#), <table:table-row-group> [9.1.9](#), <table:table-rows> [9.1.8](#), <table:table-source> [9.2.6](#), <table:title> [9.1.13](#), <text:soft-page-break> [5.6](#).

http://docs.oasis-open.org/office/v1.2/os/OpenDocument-v1.2-os-part1.html#element-table_table

The Next Millennium Document Format

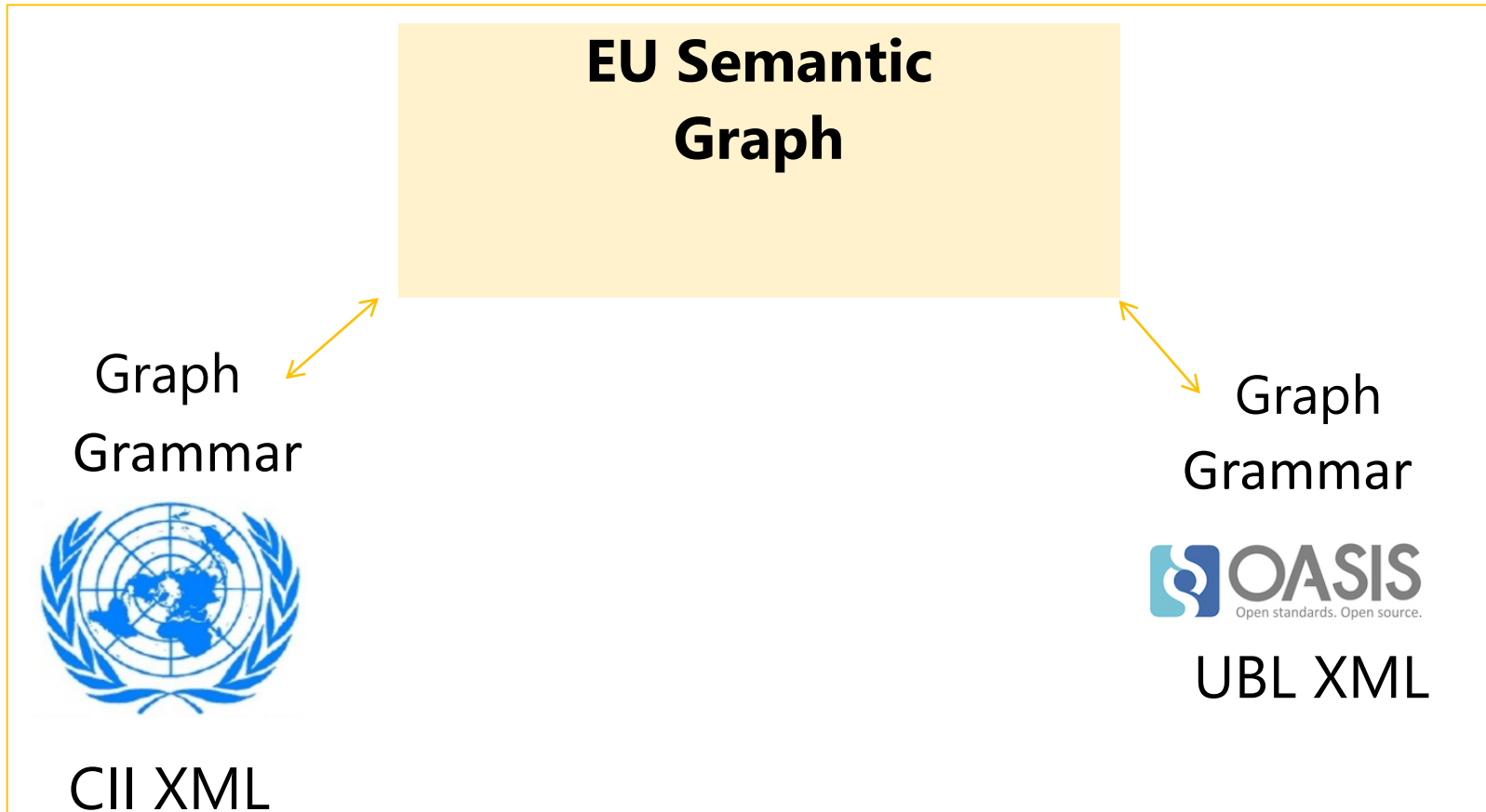
Fragment Identifier on Semantics

- Every MIME-Type may define its own fragment identifier
- Best define build-in identifier on user's semantic:
 - **Office Presentation:** #7 is slide 7
 - **Office Spreadsheets:** #*B5:C7* or #*mysheet.B5:C7*
 - **Office Text:** Every **heading** is by default a fragment identifier

[See https://lists.oasis-open.org/archives/office/200812/msg00036.html](https://lists.oasis-open.org/archives/office/200812/msg00036.html)

Syntax Binding using Graphs

Box represents 3 connected Graphs)



The Next Millennium Document Format

Consider defining Schematron on Semantics

- There are Schematron files for
 - **CII XML:**
 - <https://github.com/ConnectingEurope/eInvoicing-EN16931/tree/master/cii/schematron>
 - **UBL XML:**
 - <https://github.com/ConnectingEurope/eInvoicing-EN16931/tree/master/ubl/schematron>
- Why not avoid duplication of semantic checks?
- Define “Order date before invoice date” once on EU semantic
- Not on syntax level twice, generate it via syntax binding

The Next Millennium Document Format

Syntax Binding

Semantic

Syntax

ID	Level	Card.	BT	Desc.	DT	Path	Type	Card.	Match	Rules
BT-1	1	1..1	Invoice number	A unique identification of the Invoice.	I	/Invoice/cbc:ID	I	1..1		
BT-2	1	1..1	Invoice issue date	The date when the Invoice was issued.	D	/Invoice/cbc:IssueDate	D	1..1		
BT-3	1	1..1	Invoice type code	A code specifying the functional type of the Invoice.	C	/Invoice/cbc:InvoiceTypeCode	C	0..1	CAR-2	
BT-5	1	1..1	Invoice currency code	The currency in which all Invoice amounts are given, except for the Total VAT amount in accounting currency.	C	/Invoice/cbc:DocumentCurrencyCode	C	0..1	CAR-2	
BT-6	1	0..1	VAT accounting currency code	The currency used for VAT accounting and reporting purposes as accepted or required in the country of the Seller.	C	/Invoice/cbc:TaxCurrencyCode	C	0..1	SEM-2	

https://ec.europa.eu/cefdigital/wiki/download/attachments/59180282/CEFeInvoicingWebinar%239UnderstandingUBL_CII_v1.0.pdf?version=1&modificationDate=1520420915552&api=v2

<https://github.com/svanteschubert/en16931-data-extractor>

The Next Millennium Document Format

Agile Specification (or the lack of it)

- Specification being the blueprint of software
- Problematic Time gap:
 - Agile CI/CD software release weekly
 - Specification release every 18 month

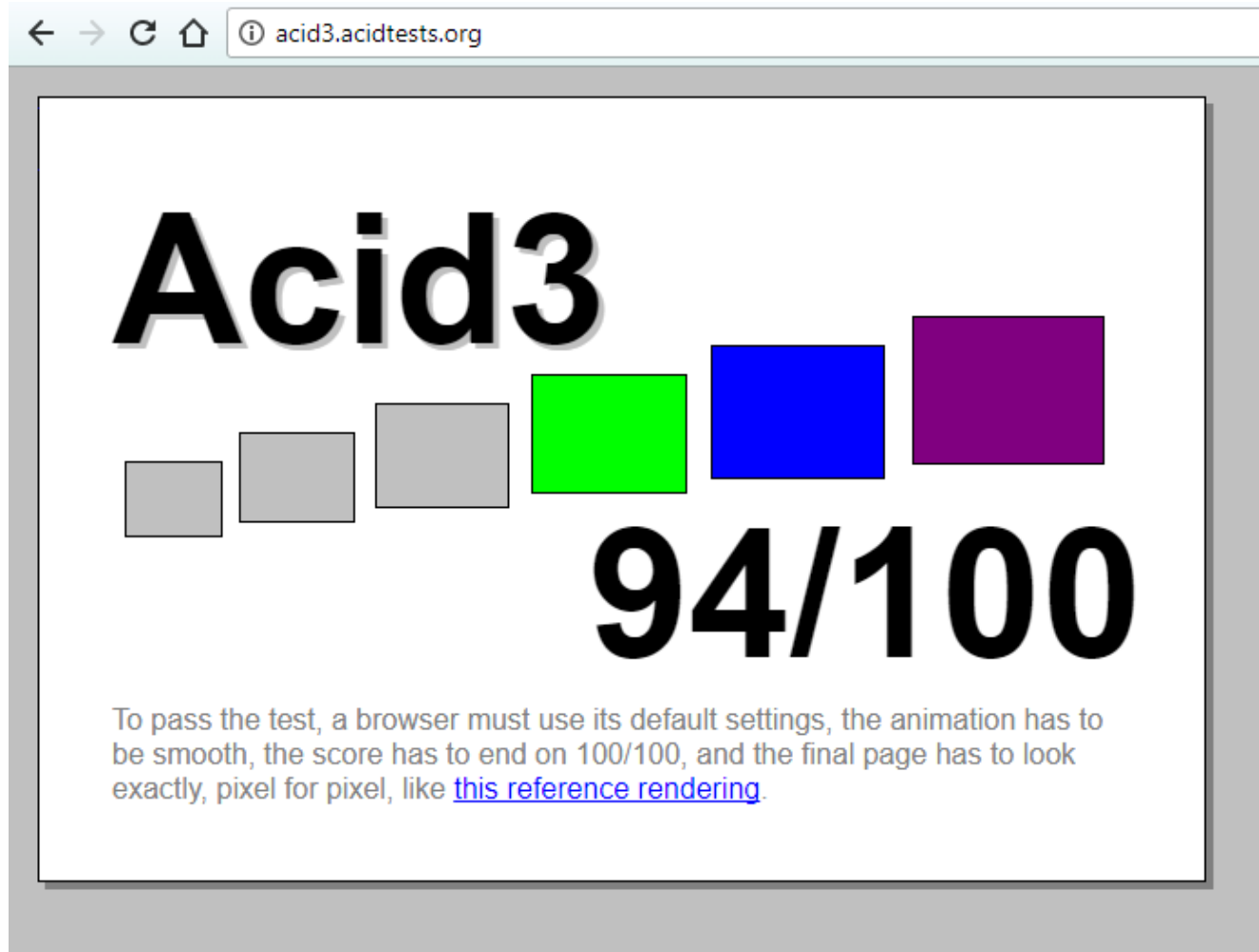
The Next Millennium Document Format

Agile Specification

- Specification being the blueprint of software
- Problematic Time gap:
 - Agile CI/CD software release weekly
 - Specification release every 18 month
- Obvious solution:
 - Let software extract the required data from specification
 - Generate as much as possible from specification
 - Tests at Spec level
 - Transitions from Spec Version to Version at Spec level

The Next Millennium Document Format

Conformance Test belong to Specification




The Next Millennium Document Format

Conformance Test belong to Specification




[Home](#) [News](#) [Oct 08, 2017 - 2 new features added: Streams & :focus-ring CSS pseudo-class](#) [Compare browsers](#) [About](#)

Can I use @page ? [Settings](#)
1 result found

CSS Paged Media (@page)  - WD Global 61.43% + 11.61% = 73.03%


CSS at-rule (@page) to define page-specific rules when printing web pages, such as margin per page and page dimensions.

[Current aligned](#) [Usage relative](#) [Date relative](#) [Show all](#)

IE	Edge *	Firefox	Chrome	Safari	Opera	iOS Safari *	Opera Mini *	Android Browser *	Chrome for Android
			49			10.2			
		 55	60	10.1		10.3		4.4	
 11	 15	 56	61	11	47	11	all	56	61
	 16	 57	62	TP	48				
		 58	63		49				
		 59	64						

[Notes](#) [Known issues \(1\)](#) [Resources \(6\)](#) [Feedback](#)

Currently no browsers appear to support the marks & bleed properties from the latest version of the specification.

 Does not support the size property

<https://caniuse.com/>

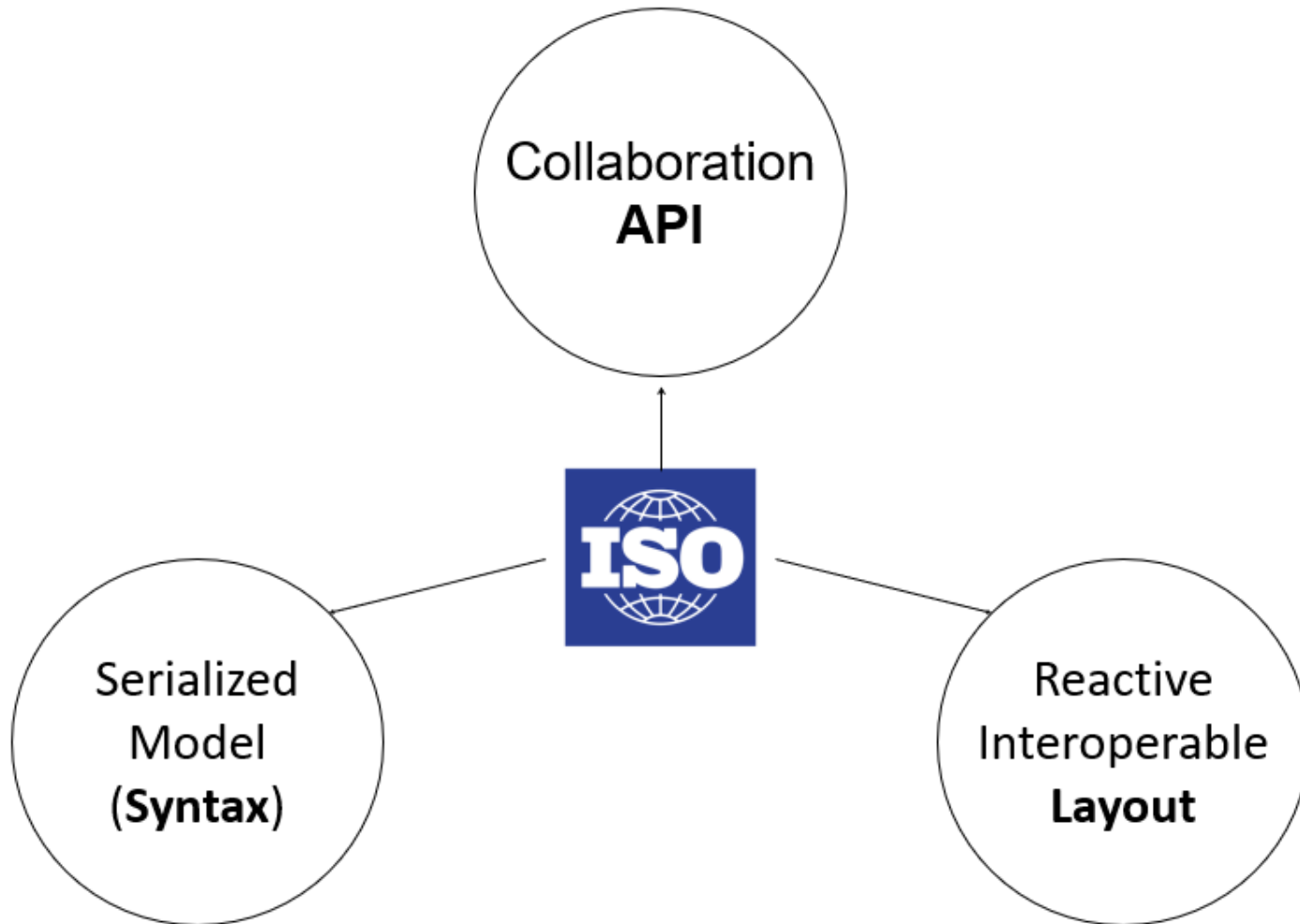
The Next Millennium Document Format

Deciding on a Document Application should be as easy as buying a toaster

Illustration							
model	Krups ControlLine KH442	Tefal TT 5500	Severin AT 2509	Severin AT 2514	WMF Stelio Toaster	Severin AT 2287	Grundig TA 6330
comparison result	 TUV-tested test & comparison procedure Vergleich.org Review 1.2 very good 09/2017	Vergleich.org Rating 1.3 very good 09/2017	Vergleich.org Review 1.5 good 09/2017	Vergleich.org Review 1.6 good 09/2017	Vergleich.org Review 1.7 good 09/2017	Vergleich.org Rating 1.8 Good 09/2017	Vergleich.org Excellent 1.9 good 09/2017
Customer rating at Amazon	★★★★☆ 4 Reviews	★★★★★ 364 reviews	★★★★☆ 2 business days	★★★★★ 963 ratings	★★★★☆ 158 reviews	★★★★☆ 511 reviews	★★★★☆ 44 Reviews
Sheets per pass	2	2	4	2	2	2	2
browning levels	6	8th	6	6	7	7	6
Power (watts)	720 watts	1.200 watts	1.400 watts	850 watts	900 watts	700 watts	850 watts
Dimensions (LxWxH)	33,2 x 24,4 x 20,2 cm	40 x 23,4 x 22,8 cm	12,6 x 37,1 x 18,2 cm	27,1 x 15,5 x 18,3 cm	32,5 x 20 x 27,5 cm	32 x 18 x 18,5 cm	34 x 21,5 x 24,5 cm
mass	0,8 kg	3,1 kg	1,0 kg	1,5 kg	1,9 kg	1,2 kg	2,0 kg
Heat insulation	✓	✓	✓	✓	✓	✓	✓
rolls rust	✓	✓	✓	✓	✓	✓	✓
Bread disc centering Bread disc	✓	✓	✓	✓	✓	✓	✗
Toast lifting function	✓	✓	✓	✓	✓	✓	✓
defrost	✓	✓	✓	✓	✓	✗	✓
manual stuffing function	✓	✓	✓	✓	✓	✓	✓
	⊕ Testieger Foundation Warentest 04/2016	⊕ including egg cooker and egg pans	⊕ Housing is not hot ⊕ also suitable for bread	⊕ very quiet ⊕ integrated roll holder	⊕ tans evenly ⊕ illuminated key labels	⊕ integrated roll holder ⊕ good workmanship	⊕ stable crumb tray ⊕ good workmanship

The Next Millennium Document Format

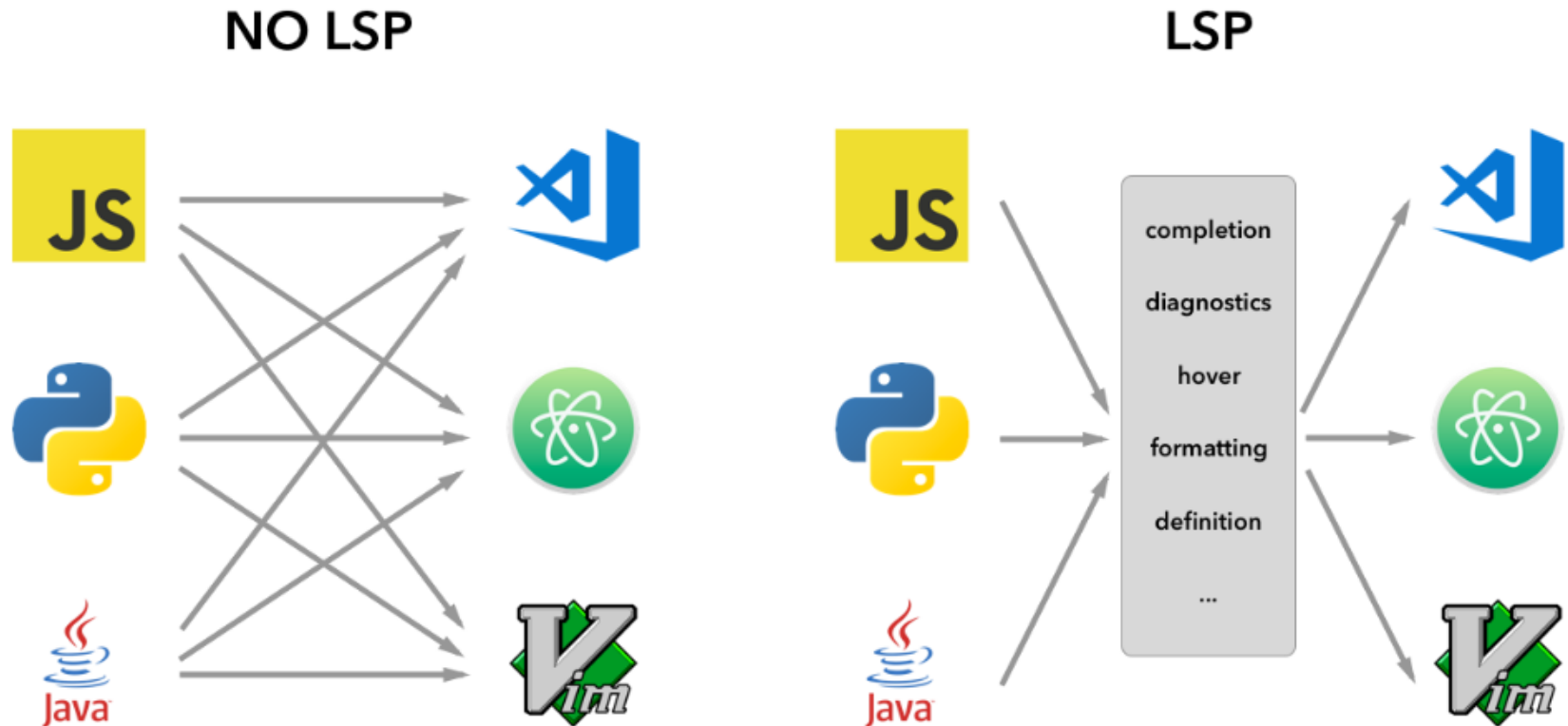
Collection of Challenges & Solutions



The Next Millennium Document Format

Modularize and Reuse Functionality

Language Server Protocol (LSP) from Visual Studio Code



See https://fosdem.org/2020/schedule/track/free_tools_and_editors/

The Next Millennium Document Format

Gap: ODF User Feature \leftrightarrow ODF XML Syntax

- Page orientation on Paragraph (or Table)
 - **User View:**
Paragraph \rightarrow Orientation
 - **Syntax View:**
Paragraph \rightarrow Style \rightarrow
Master Page \rightarrow Page Layout \rightarrow
Page Layout Properties \rightarrow Orientation

The Next Millennium Document Format

Gap: ODF User Feature ↔ ODF XML Syntax

```
<?xml version="1.0" encoding="UTF-8"?>
<office:document-content xmlns:office="urn:oasis:names:tc:opendocument:xmlns:office:1.0" xmlns:style="urn:oasis:names:tc:opendocument:xmlns:style:1.0" xmlns:text="urn:oasis:names:tc:opendocument:xmlns:text:1.0" xmlns:draw="urn:oasis:names:tc:opendocument:xmlns:draw:1.0" xmlns:form="urn:oasis:names:tc:opendocument:xmlns:form:1.0" xmlns:table="urn:oasis:names:tc:opendocument:xmlns:table:1.0" xmlns:tablexforms="urn:oasis:names:tc:opendocument:xmlns:tablexforms:1.0" xmlns:xforms="urn:oasis:names:tc:formscript:xmlns:xforms:1.0" xmlns:xlink="http://www.w3.org/1999/xlink">
  <office:scripts/>
  <office:font-face-decls>
    <style:font-face style:name="Lucida Sans1" svg:font-family="&apos;Lucida Sans&apos;" style:font-family-generic="swiss"/>
    <style:font-face style:name="Liberation Serif" svg:font-family="&apos;Liberation Serif&apos;" style:font-family-generic="roman" style:font-pitch="variable"/>
    <style:font-face style:name="Liberation Sans" svg:font-family="&apos;Liberation Sans&apos;" style:font-family-generic="swiss" style:font-pitch="variable"/>
    <style:font-face style:name="Lucida Sans" svg:font-family="&apos;Lucida Sans&apos;" style:font-family-generic="system" style:font-pitch="variable"/>
    <style:font-face style:name="Microsoft YaHei" svg:font-family="&apos;Microsoft YaHei&apos;" style:font-family-generic="system" style:font-pitch="variable"/>
    <style:font-face style:name="SimSun" svg:font-family="SimSun" style:font-family-generic="system" style:font-pitch="variable"/>
  </office:font-face-decls>
  <office:automatic-styles>
    <style:style style:name="P1" style:family="paragraph" style:parent-style-name="Standard" style:master-page-name="Standard">
      <style:paragraph-properties style:page-number="auto"/>
    </style:style>
    <style:style style:name="P2" style:family="paragraph" style:parent-style-name="Standard" style:master-page-name="First_20_Page">
      <style:paragraph-properties style:page-number="auto"/>
    </style:style>
    <style:style style:name="P3" style:family="paragraph" style:parent-style-name="Header">
      <style:paragraph-properties fo:text-align="center" style:justify-single-word="false"/>
    </style:style>
  </office:automatic-styles>
  <office:body>
    <office:text text:use-soft-page-breaks="true">
      <text:sequence-decls>
        <text:sequence-decl text:display-outline-level="0" text:name="Illustration"/>
        <text:sequence-decl text:display-outline-level="0" text:name="Table"/>
        <text:sequence-decl text:display-outline-level="0" text:name="Text"/>
        <text:sequence-decl text:display-outline-level="0" text:name="Drawing"/>
      </text:sequence-decls>
      <text:p text:style-name="P1">Hello</text:p>
      <text:p text:style-name="Standard">(some standard style text</text:p>
      <text:p text:style-name="P1">World!</text:p>
    </office:text>
  </office:body>
</office:document-content>
```

The Next Millennium Document Format

Gap: ODF User Feature ↔ ODF XML Syntax

```
128 <style:header-footer-properties fo:min-height="0cm" fo:margin-top="0.499cm"/>
129 </style:footer-style>
130 </style:page-layout>
131 <style:page-layout style:name="Mpm1">
132 <style:page-layout-properties fo:page-width="29.7cm" fo:page-height="21.001cm" style:num-format="1" style:print-orientation="landscape" fo:margi
133 <style:footnote-sep style:width="0.018cm" style:distance-before-sep="0.101cm" style:distance-after-sep="0.101cm" style:line-style="solid" st
134 </style:page-layout-properties>
135 <style:header-style>
136 <style:header-footer-properties fo:min-height="0cm" fo:margin-bottom="0.499cm"/>
137 </style:header-style>
138 <style:footer-style>
139 <style:header-footer-properties fo:min-height="0cm" fo:margin-top="0.499cm"/>
140 </style:footer-style>
141 </style:page-layout>
142 </office:automatic-styles>
143 <office:master-styles>
144 <style:master-page style:name="First_20_Page" style:display-name="First Page" style:page-layout-name="Mpm3" style:next-style-name="Standard">
145 <style:header>
146 <text:p text:style-name="Header">First Page Header</text:p>
147 </style:header>
148 <style:footer>
149 <text:p text:style-name="Footer">First Page Footer</text:p>
150 </style:footer>
151 </style:master-page>
152 <style:master-page style:name="Standard" style:page-layout-name="Mpm1">
153 <style:header>
154 <text:p text:style-name="Mpm1">Default Style Header (landscape)</text:p>
155 </style:header>
156 <style:footer>
157 <text:p text:style-name="Footer">Default Style Footer (landscape)</text:p>
158 </style:footer>
159 </style:master-page>
160 <style:master-page style:name="Landscape" style:page-layout-name="Mpm2">
161 <style:header>
162 <text:p text:style-name="Header"/>
163 </style:header>
164 </style:master-page>
```

The Next Millennium Document Format

To remember...

- Cross-Format & Cross-Applcation offline collaboration
- Declare Semantics & their Changes (API)
- **Must** have “Machine-Readable” Specs in our CI/CD
- Modularize & decentralize Editor functionality (e.g. LSP)
- Syntax volatile, Semantic abstraction stable