

Document Interop from an Open Source perspective

PHPExcel & OPENXML4J

Maarten Balliauw – <http://blog.maartenballiauw.be>

Julien Chable – <http://blogs.developpeur.org/neodante>

12-Dec-08

Who are we?

- Maarten Balliauw
 - RealDolmen – <http://www.realdolmen.com>
 - <http://blog.maartenballiauw.be>
 - maarten@maartenballiauw.be
- Julien Chable
 - Wygwam – <http://www.wygwam.com>
 - <http://blogs.developpeur.org/neodante>
 - julien@wygwam.com



PHPExcel and OPENXML4J

- PHPExcel



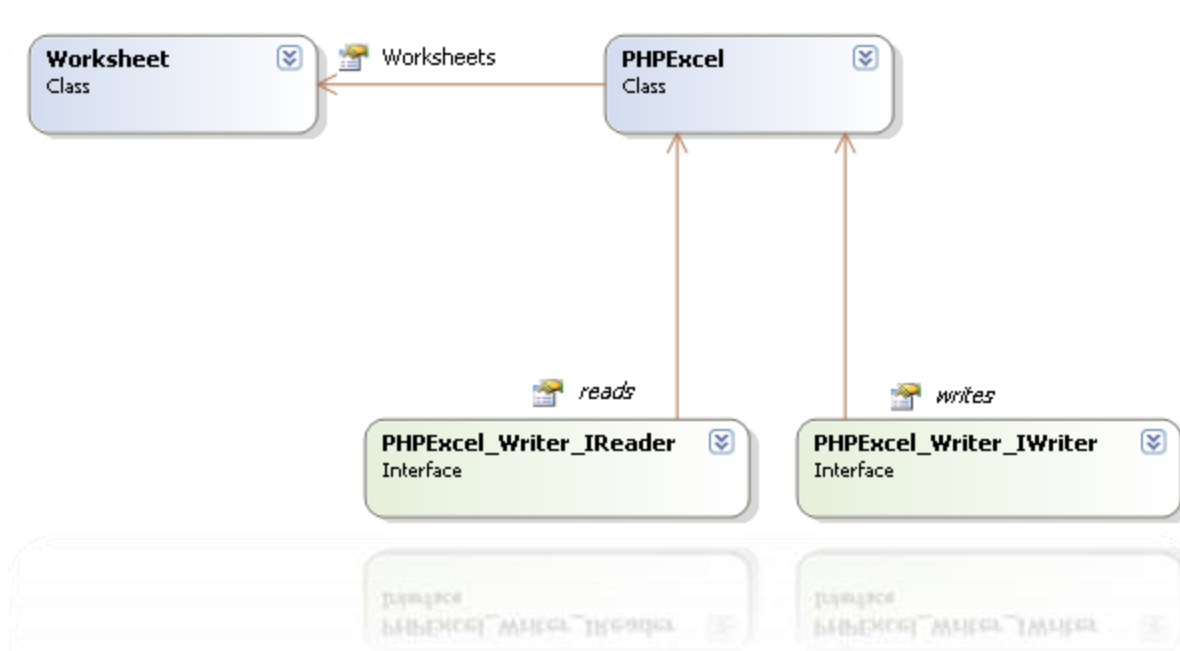
- OPENXML4J



PHPExcel – Usage scenarios

- Generating Excel documents on server
- Reporting from PHP applications in different file formats
- Processing existing Excel documents on server
- Importing data from Excel document
- Re-using business logic stored in Excel document (check my blog!)
- ...

PHPExcel - Architecture



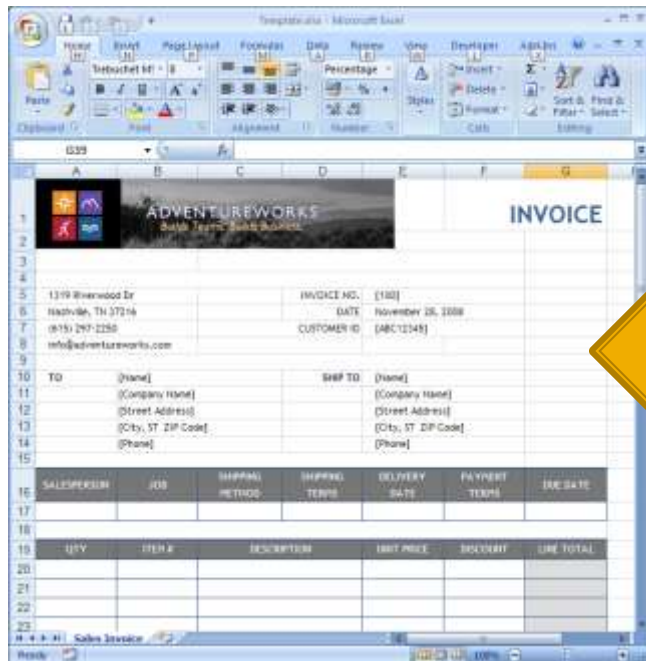
- Core class library featuring spreadsheet engine
- Supported by IReader and IWriter instances

PHPExcel – Readers and writers

- Readers for Excel2007, Excel5, CSV, Serialized format
- Writers for Excel2007, Excel5, CSV, HTML, PDF and Serialized format



Interoperability demo



ADVENTUREWORKS
Build. Terms. Build. Build.

1319 Riverwood Dr
Nashville, TN 37216
(615) 297-2255
info@adventureworks.com

INVOICE NO. [110]
DATE November 28, 2008
CUSTOMER ID [ABC1234]

TO: (Name) (Company Name) (Street Address) (City, ST ZIP Code) (Phone)
SHIP TO: (Name) (Company Name) (Street Address) (City, ST ZIP Code) (Phone)

SALESPERSON	JOB	SHIPPING METHOD	SHIPPING TERMS	DELIVERY DATE	PAYMENT TERMS	DUPLICATE

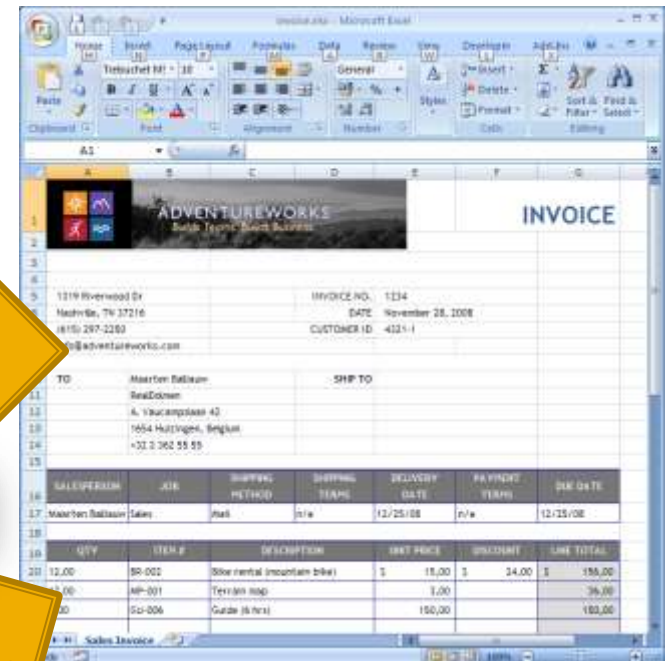
QTY	ITEM #	DESCRIPTION	UNIT PRICE	DISCOUNT	LINE TOTAL

Read/write

PHPExcel

php

Calculation



ADVENTUREWORKS
Build. Terms. Build. Build.

1319 Riverwood Dr
Nashville, TN 37216
(615) 297-2255
info@adventureworks.com

INVOICE NO. 1234
DATE November 28, 2008
CUSTOMER ID 4321-1

TO: Maarten Balleux (Company Name) RealEstate A. Vlaucplasse 42 1654 Hulsinghe, Belgium +32 3 362 55 55
SHIP TO:

SALESPERSON	JOB	SHIPPING METHOD	SHIPPING TERMS	DELIVERY DATE	PAYMENT TERMS	DUPLICATE
Maarten Balleux	Sales	Fast	OTB	12/25/08	n/a	12/25/08

QTY	ITEM #	DESCRIPTION	UNIT PRICE	DISCOUNT	LINE TOTAL
12.00	BR-002	Bike rental (mountain bike)	\$ 15.00	\$ 14.00	\$ 158.00
1.00	MP-001	Terrain map	1.00		36.00
1.00	SC-006	Guide (3 hrs)	180.00		180.00



Before we start...

PHP

- `$x = 5;`
- `$x = "hello";`

- `$x = "Hello " . "world";`

- `echo "hello !";`

C#

- `int x = 5;`
- `string x = "hello";`

- `string x = "Hello " + "world";`

- `Console.Write("hello!");`

DEMO

PHPExcel and OPENXML4J

- PHPExcel



- OPENXML4J



Open XML for Java

OpenXML4J

- OpenXML4J launch May 30th, 2007

The collage features several key elements:

- ASP MAGAZINE**: A website header with navigation links like 'Accueil', 'Forums', and 'Offres d'emplois'. A search bar indicates the search engine is not yet active.
- Wygwam lance le projet Open Source OpenXML4J**: A central news article from 'JDN développeur' dated Friday, June 1, 2007. It reports that Wygwam has launched an open source project around the OpenXML format. The article mentions that Wygwam is heavily involved in the format from its origin and that the French company has launched its first version.
- CRAIG KITTERMAN'S INTEROPERABILITY COM**: A banner for a community connecting people, data, and diverse systems.
- JAVA OPEN XML API PROJECT OFFICIALLY LAUNCHED!**: A banner announcing the official launch of the Java API library project, stating it was started by Julien Chabrie (Wygwam France).
- Le Monde Informatique.fr**: A screenshot of a news article from May 30, 2007, titled 'Architecture logicielle' and 'Inscrivez-vous XML'. It discusses the project and its relevance to the .NET and Open Source communities.
- Pro programmez!**: A developer portal banner for architects, designers, and developers.
- GNT**: A banner for a website offering information, hardware, and mobile services.

OpenXML4J

- Multi platform API for Java client/J2SE and serveur/J2EE applications
- Open Source Project under Licence BSD (dual license Apache V2)
- Official website : <http://www.openxml4j.org/>
- Based on ECMA 376 specifications

Why OpenXML4J and why open source ?

- Unified companies and communities efforts into a concrete and common implementation
- Promote Open XML Interoperability in heterogeneous environments

An open standard, an open implementation !

Initial steps overview

- Several steps:
 1. Implementation of *Open Packaging Convention* (part 2)
 2. Implementation of typed parts :
WordprocessingML, SPreadsheetML and PresentationML
 3. Implementation of an object model for each format
- Now, only *step 1* is implemented. We need contributors and ... time !

Overview

Libraries

DOM4J

XMLbeans

...

Open Packaging Convention

Shared schemas

(DrawingML, MathML, CustomXML, Metadata,...)

WordprocessingML

STP

Object model

SpreadsheetML

STP

Object model

PresentationML

STP

Object model

Overview : Open Packaging Convention

Librairies

DOM4J

XMLbeans

...

Open Packaging Convention

Shared schemas
(DrawingML, MathML, CustomXML, Metadata,...)

WordprocessingML

STP

Modèle objets

SpreadsheetML

STP

Modèle objets

PresentationML

STP

Modèle objets

A .NET API clone ?

```
import org.openxml4j.opc;  
...  
Package p = Package.open(filepath,  
PackageAccess.READ);  
    for (PackagePart part : p.getParts())  
        System.out.println(part.getPartName() +  
" -> " + part.getContentType());
```



```
using System.IO.Packaging;  
...  
Package p = Package.Open("",  
System.IO.FileMode.Open);  
    for (PackagePart part in p.GetParts())  
        Console.WriteLine(part.Uri + " -> " +  
part.ContentType);
```



A .NET API clone ?

```
Package p = Package.open("c:\....docx",  
PackageAccess.READ);  
PackageProperties props = p.getPackageProperties();  
System.out.println("Title: " +  
props.getTitleProperty().getValue());  
p.close();
```



```
Package p = Package.Open("c:\....docx",  
FileMode.Open);  
PackageProperties props = p.PackageProperties;  
Console.WriteLine("Title: " +  
props.Title);  
p.Close();
```

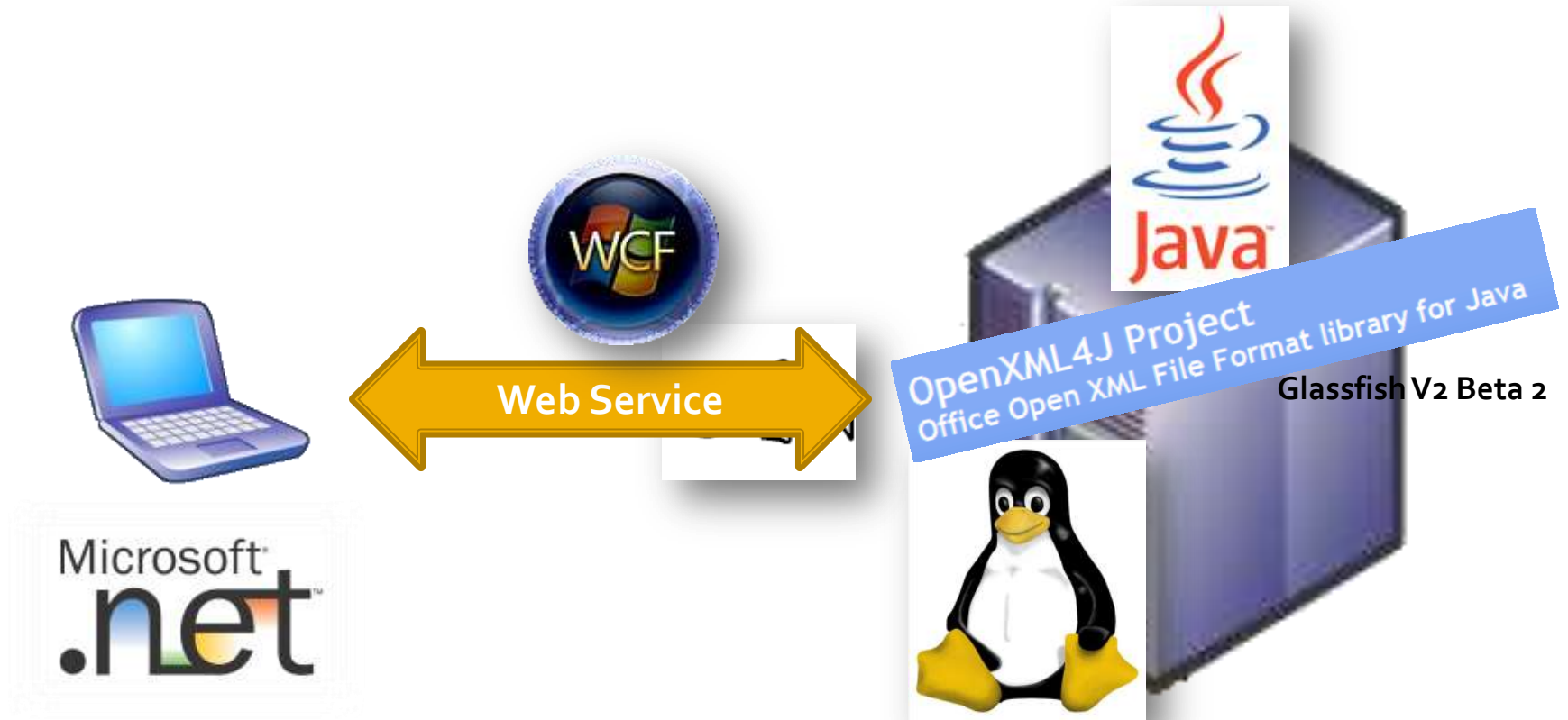


Roadmap

- Beginning in 2006 :
 - Now uses by several open source projects :
 - POI (<http://poi.apache.org>)
 - Doc4J (<http://dev.plutext.org/trac/docx4j>)
 - Several projects in companies (French)
- Today we're joining the Apache foundation and the POI project to offer a single API and point of contact to the community

DEMO

Interop demo



DEMO

Specifications conformity

- Close to the ECMA standard specifications → Validation tests

ID	Rule	Reference	Package Implementer	Format Designer	Format Producer	Format Consumer	Status	Check location	Associated Test
M1.1	The package implementer shall require a part name.	8.1, 8.1.1	x				OK	PackagePart class and PackagingURIHelper.isValidPartName	OPCCompliance_PartName.testEmptyPartName
M1.2	The package implementer shall require a content type and the format designer shall specify the content type.	8.1	x	x			OK	PackagePart class	NA
M1.3	A part name shall not have empty segments.	0	x				OK	PackagingURIHelper.isValidPartName	OPCCompliance_PartName.testPartNameWithInvalidSegmentsFailure
M1.4	A part name shall start with a forward slash ("/") character.	0	x				OK	PackagingURIHelper.isValidPartName	OPCCompliance_PartName.testPartNameStartsWithAForwardSlashFailure
M1.5	A part name shall not have a forward slash as the last character.	0	x				OK	PackagingURIHelper.isValidPartName	OPCCompliance_PartName.testPartNameEndsWithAForwardSlashFailure
M1.6	A segment shall not hold any characters other than pchar characters.	0	x						OPCCompliance_PartName.testPartNameComparisonFailure
M1.7	A segment shall not contain percent-encoded forward slash ("/"), or backward slash ("\") characters.	0	x						OPCCompliance_PartName.testPartNameComparisonFailure
M1.8	A segment shall not contain percent-encoded unreserved characters.	0	x						OPCCompliance_CoreProperties.testCorePropertiesPart
M1.9	A segment shall not end with a dot (".") character.	0	x				OK	PackagingURIHelper.isValidPartName	OPCCompliance_CoreProperties.testOnlyOneCorePropertiesPart
M1.10	A segment shall include at least one non-dot character	0	x				OK	PackagingURIHelper.isValidPartName	OPCCompliance_CoreProperties.testOnlyOneCorePropertiesPart_AddRelationship
M1.11	A package implementer shall neither create nor recognize a part with a part	8.1.1.1	x						OPCCompliance_CoreProperties.testOnlyOneCorePropertiesPart_AddPart

```

Test for test.org.openxml4j.opc.compliance [Runner: JUnit 3]
├── test.org.openxml4j.opc.compliance.OPCCompliance_PartName
│   ├── testInvalidPartNames
│   ├── testValidPartNames
│   ├── testEmptyPartNameFailure
│   ├── testPartNameWithInvalidSegmentsFailure
│   ├── testPartNameStartsWithAForwardSlashFailure
│   ├── testPartNameEndsWithAForwardSlashFailure
│   ├── testPartNameComparison
│   └── testPartNameComparisonFailure
├── test.org.openxml4j.opc.compliance.OPCCompliance_CoreProperties
│   ├── testCorePropertiesPart
│   ├── testOnlyOneCorePropertiesPart
│   ├── testOnlyOneCorePropertiesPart_AddRelationship
│   ├── testOnlyOneCorePropertiesPart_AddPart
│   ├── testDoNotUseCompatibilityMarkup
│   ├── testDCTermsNamespaceLimitedUse
│   ├── testUnauthorizedXMLLangAttribute
│   ├── testLimitedXSITypeAttribute_NotPresent
│   └── testLimitedXSITypeAttribute_PresentWithUnauthorizedValue

```

POI

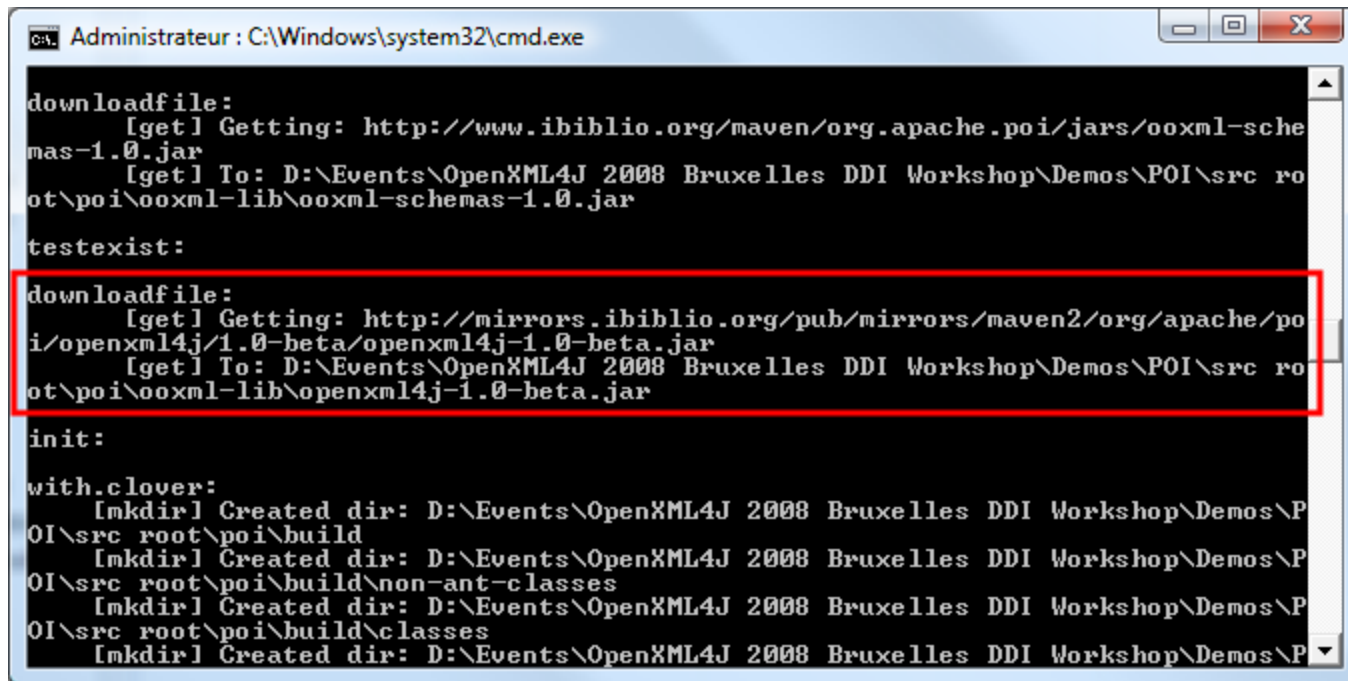
- A set of APIs to read and write Excel files using Java
- Still very active since April 2001: Top Level Project within Apache !
- Also support several MS file formats : Word, PowerPoint, Visio, Publisher
- Version 3.5 will support XLSX and PPTX

POI APIs

- Several APIs for several file formats:
 - **HSSF and XSSF**: Read and write Excel documents (97 – 2003 and 2007)
 - **HWPF**: Read (and partly write) Word 97 documents. Early stages of development.
 - **HSLF**: Read and write PowerPoint 97-2003 documents.
 - **HPSF**: Read and write OLE 2 property (title, author, etc)
 - **HDGF**: Read at very low level (and simple text extraction) Visio 97-2003 documents
 - **HPBF**: Read at very very low level (and simple text extraction) Publisher 98-2007 documents

POI – OpenXML4J

- POI compilation with Ant :



```
Administrateur : C:\Windows\system32\cmd.exe

downloadfile:
  [get] Getting: http://www.ibiblio.org/maven/org.apache.poi/jars/ooxml-schemas-1.0.jar
  [get] To: D:\Events\OpenXML4J 2008 Bruxelles DDI Workshop\Demos\POI\src root\poi\ooxml-lib\ooxml-schemas-1.0.jar

testexist:

downloadfile:
  [get] Getting: http://mirrors.ibiblio.org/pub/mirrors/maven2/org/apache/poi/openxml4j/1.0-beta/openxml4j-1.0-beta.jar
  [get] To: D:\Events\OpenXML4J 2008 Bruxelles DDI Workshop\Demos\POI\src root\poi\ooxml-lib\openxml4j-1.0-beta.jar

init:

with.clover:
  [mkdir] Created dir: D:\Events\OpenXML4J 2008 Bruxelles DDI Workshop\Demos\POI\src root\poi\build
  [mkdir] Created dir: D:\Events\OpenXML4J 2008 Bruxelles DDI Workshop\Demos\POI\src root\poi\build\non-ant-classes
  [mkdir] Created dir: D:\Events\OpenXML4J 2008 Bruxelles DDI Workshop\Demos\POI\src root\poi\build\classes
  [mkdir] Created dir: D:\Events\OpenXML4J 2008 Bruxelles DDI Workshop\Demos\POI\src root\poi\build\classes
```

DEMO

Questions?



Thank you!

- Maarten Balliauw
 - <http://blog.maartenballiauw.be>
 - maarten@maartenballiauw.be
- Julien Chable
 - <http://blogs.developpeur.org/neodante>
 - julien@wygwam.com