Excel 2 XML

Extract data from the most used database in the world and transform it in something useful
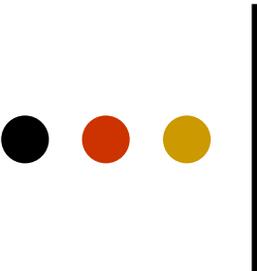
Agenda

- Who?
- Why?
- How?
- What?
- When?
- Resources
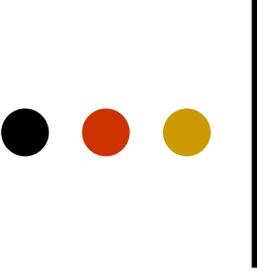
Who am I

○ My Unix box usually answers…
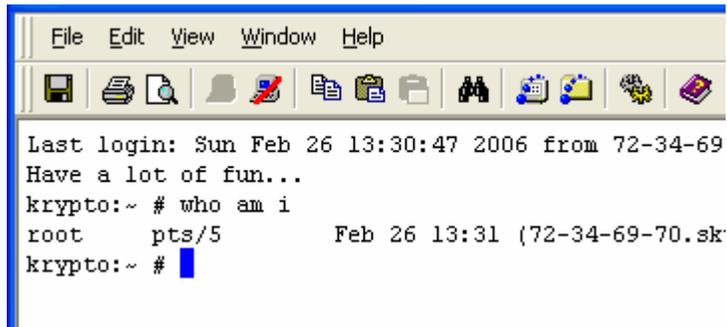
```
File  Edit  View  Window  Help

Last login: Sun Feb 26 13:30:47 2006 from 72-34-69
Have a lot of fun...
krypto:~ # who am i
root      pts/5          Feb 26 13:31 (72-34-69-70.sk
krypto:~ #
```

JGP.net

Ok, no joke now…

- Jean Georges Perrin
- Dev. Tools (xGL, Java EE, PHP)
- Web since 1994
- IIUG BoD since 2002
- Focus on **Software Engineering**
- Proudly lives in

Ville de Haguenau

tourisme-alsace.com
ALSACE

JGP.net
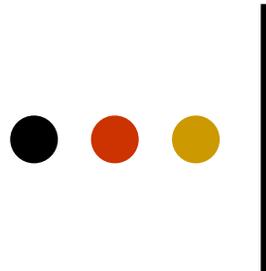
And you?

- Java gurus?
- PHP?
- XML gurus?
- Excel gurus?

JGP.net

Why?

- Need to publish Excel data on the Web
- Need to store Excel data in XML-oriented databases
  - Informix Cheetah, Oracle, DB2 "pureXML"
- Part of an SOA chain

Informix

IBM

ORACLE

JGP.net
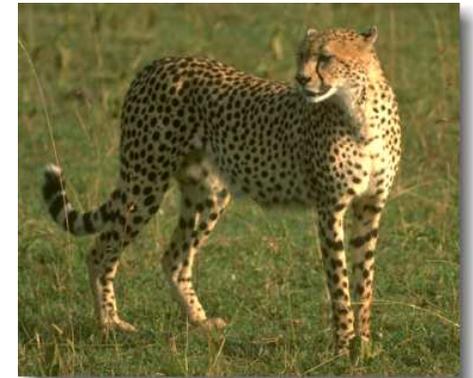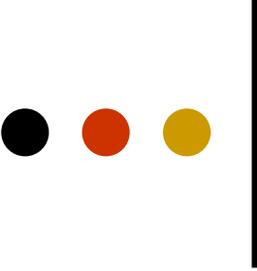
How?

- Go to **File**, **Save As**…

- Why not?
  - Microsoft centric
 - Microsoft XML dependent
  - No (easy) automation

JGP.net

How?
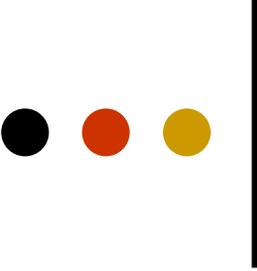
- Parse Excel files with POI

- Build XML document with Xerces

- Using Java 5.0, running on **any platform**

JGP.net

Excel Anatomy

○ **Workbook**

    ● Contains **sheets**

        ● Contains **rows**

           ● Contains **cells**
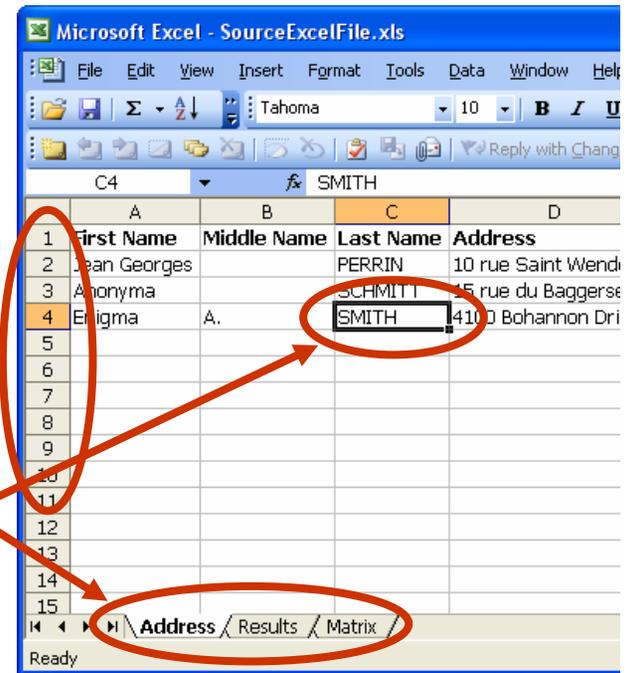
○ Similar to something you know?

JGP.net
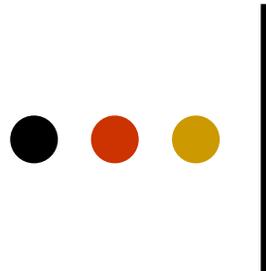
Database Anatomy

○ **Database** ◄┄┄┄┄┄► ○ **Workbook**

● Contains **tables** ◄┄┄► ● **sheets**

● Contains **rows** ◄┄┄┄► ● **rows**

• Contains **fields** ◄┄┄┄┄► • **cells**

○ Excel is the most popular database and nobody knows it!

Understanding the "table"

- 3 type of tables in Excel
  - Vertical
  - Horizontal
  - Raw

XML
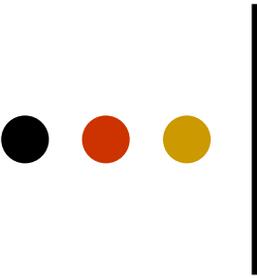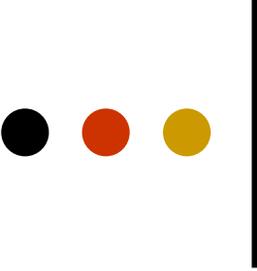
- a
  - contains one or more <table/>
    - contains one or more
      - contains one or more

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<database>
  <table name="Address" orientation="vertical">
    <row id="1">
      <field dataType="text" name="First Name">Jean Georges</field>
      <field dataType="unknown" name="Middle Name"></field>
      <field dataType="text" name="Last Name">PERRIN</field>
      <field dataType="text" name="City">Haguenau</field>
      <field dataType="unknown" name="State"></field>
      <field dataType="text" name="Country">France</field>
      <field dataType="date" name="Birth Date">Tue Oct 05 00:00:00 CET 1971</field>
      <field dataType="text" name="e-mail">jgp@jgp.net</field>
    </row>
    <row id="2">
      <field dataType="text">Anonyma</field>
…
```

What?

Have a look at what it looks like…

JGP.net

When?

- A bit of code to write
  - Vertical tables
  - Raw data
  - Consistent field type for the column
- A few challenges to solve
  - Formulas
  - Storage of meta data
- Better analysis and factoring
  - Better understanding of Excel
  - Better code design
- Test
- Documentation
- Any volunteers?

JGP.net

Resources

- JGP.net
  - Source code & project
  - http://www.jgp.net
- APACHE POI

  - Part of the Jakarta Apache project
  - http://jakarta.apache.org/poi
- Xerces
  - Apache project
  - http://xerces.apache.org
- Eclipse, JDeveloper, NetBeans…

Thanks

Feel free:

jgp@jgp.net

jgp@iiug.org

JGP.net