



ENTHOUGHT
SCIENTIFIC COMPUTING SOLUTIONS

Avoiding the “Excel hell”

using a Reproducible and Flexible Framework for
Research, Analysis, and Production

Didrik Pinte
QPUG – NYC – May 2013

Summary

- Context
- Issues
- Solutions - demos

In the press

Damn Excel! How the 'most important software application of all time' is ruining the world

CNN Money, April 2013

88% of spreadsheets have errors

MarketWatch, April 2013

In the press - Reinhard & Rogoff



PHOTO: ANDREW HARRER/BLOOMBERG/GETTY

In the press - JP Morgan



PHOTO: WAYNE LEVIN/GETTY IMAGES

In the press

- Barclays purchase of Lehman - 2008
- Utah's office of education - 2012
- Fannie Mae - 2003
- etc.



For more: <http://www.eusprig.org/horror-stories.htm>

On the ground

- Spreadsheet/model management projects - Energy sector
- Migrating away from excel for quoting - Bond desk
- Processing workflows, risk management
- UI widgets for large dataset

Growing needs

- Governance
 - Transparency and traceability
 - Replicability
 - Benchmarks
- Big data

Basel III
Solvency II

What are the real issues?

- Referencing
- Numerical issues
- Manual operations (cut & paste, ...)
- Tracing IO's in libraries/API's (e.g. QuantlibXL)
- Version control
- Meetings the needs of researchers, analysts, production reporting, etc.

Any other ??

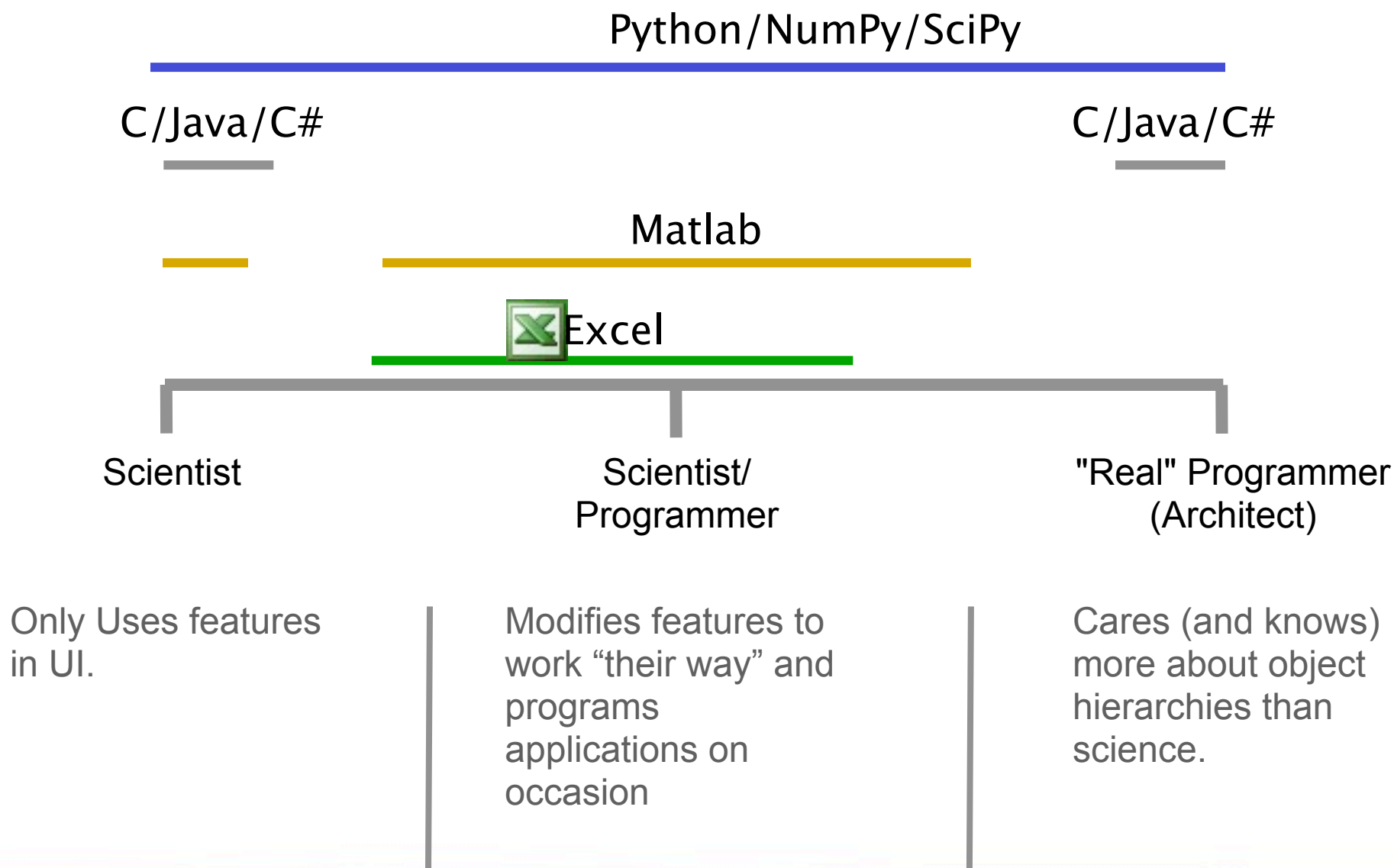
Exhell

Reproducibility



- Testing
- Access to data
- Version control
- Smells:
 - manual steps
 - no testing / dependence between data and processing

The Python ecosystem has a solution

Spectrum of Users



Spectrum of Users

  Python/NumPy/SciPy

C/Java/C#

C/Java/C#

Matlab



Scientist

Scientist/
Programmer

"Real" Programmer
(Architect)

Only Uses features
in UI.

Modifies features to
work "their way" and
programs
applications on
occasion

Cares (and knows)
more about object
hierarchies than
science.

The Python ecosystem as a solution?

- A step-by-step approach:
 - Python as a backend, Excel as the frontend
 - Python within Excel
 - A complete Python solution
- Demo's

Python as a processing engine

- Load -> process -> export
- Tools:
 - xlrd, xlwt for xls
 - openpyxl for xlsx/xlsm
 - reportlab

Python in Excel

- Using Python directly within Excel on validated workflows, processes and function.

pyvot

pywin32

XLLoop



Typical workflows

Research

Analyze

Production

Acquire
data

Process

Share

Publish

Complete Python solutions

- Existing ones?
 - Python based or not
- What do you need?
 - integrated tools
 - scripting environment, notebooks
 - data management done the right way ;-)
 - control on:
 - sharing
 - versions
 - deployment