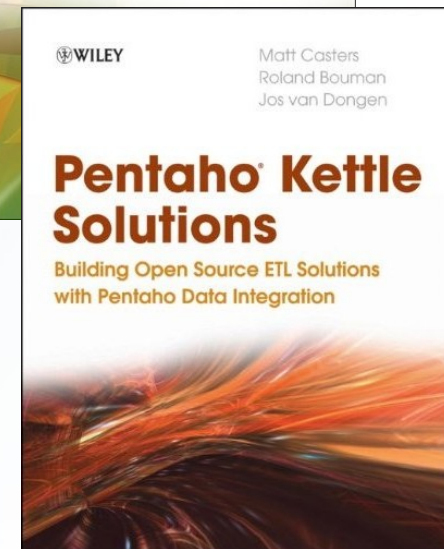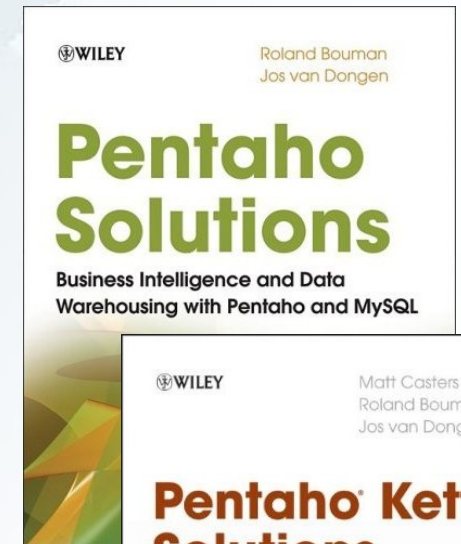# Kettle-Cookbook

**Auto-documentation
for
Kettle jobs and transformations**

Thanks for attending!

- Roland Bouman; Leiden, Netherlands

- Ex MySQL AB, Sun Microsystems

- Web and BI Developer

- Co-author of "Pentaho Solutions"

- ...and "Pentaho Kettle Solutions"

- Blog: http://rpbouman.blogspot.com/

- Twitter: @rolandbouman

Agenda

- Documentation

- Introducing Kettle-Cookbook

- Demonstration

- Roadmap

- Questions and answers

- Links and resources

Agenda

- Documentation

- Introducing Kettle-Cookbook

- Demonstration

- Roadmap

- Questions and answers

- Links and resources

Documentation

Roland Bouman: http://rpbouman.blogspot.com/ Twitter: @rolandbouman
kettle-cookbook: http://code.google.com/p/kettle-cookbook/

Documentation

Documentation

Roland Bouman: http://rpbouman.blogspot.com/ Twitter: @rolandbouman
kettle-cookbook: http://code.google.com/p/kettle-cookbook/

Documentation

Roland Bouman: http://rpbouman.blogspot.com/ Twitter: @rolandbouman
kettle-cookbook: http://code.google.com/p/kettle-cookbook/

Documentation

Roland Bouman: http://rpbouman.blogspot.com/ Twitter: @rolandbouman
kettle-cookbook: http://code.google.com/p/kettle-cookbook/
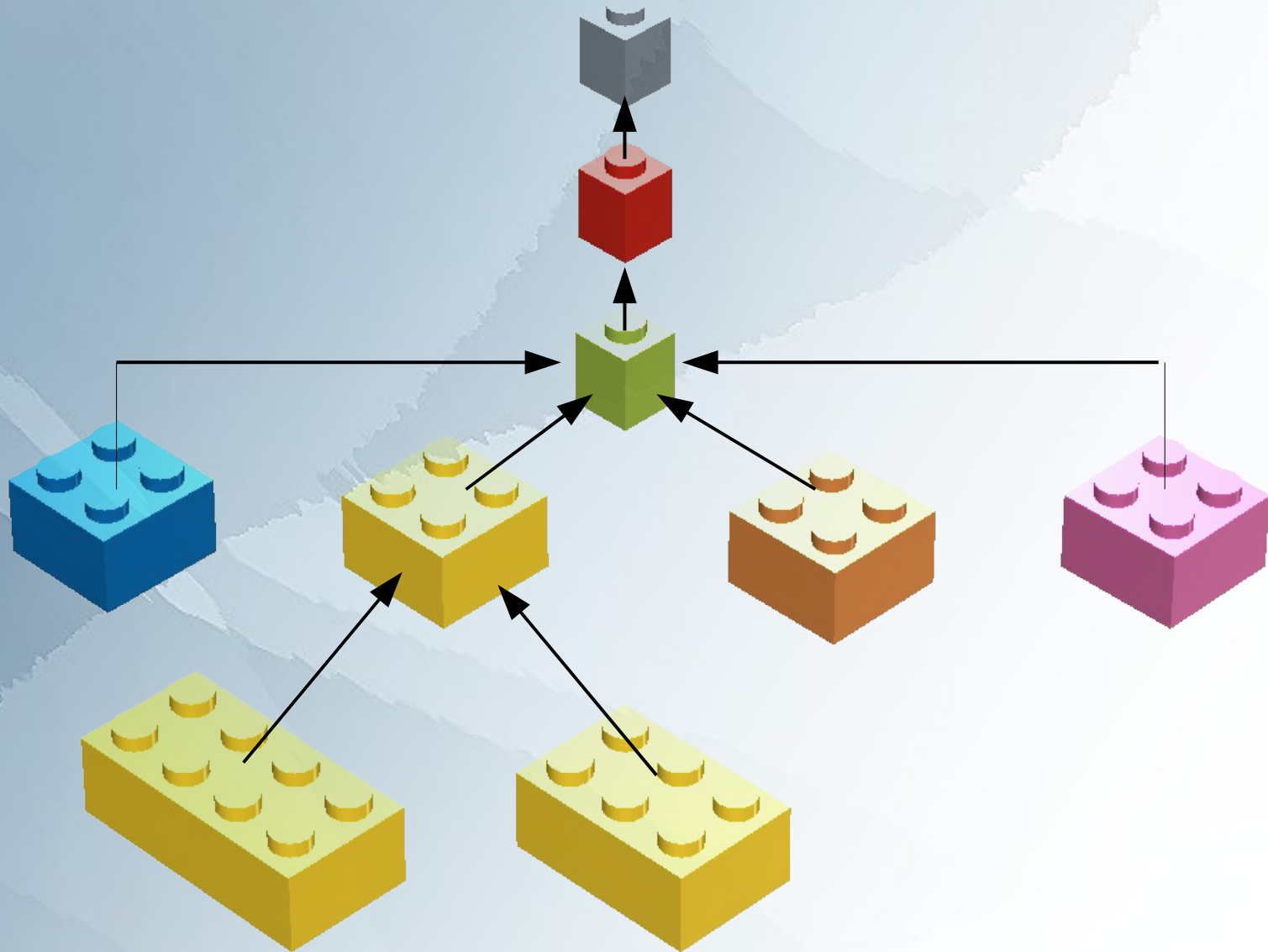
Documentation

Documentation Benefits

- Allows an ETL solution to be verified against design documents

- If done right, can help to train developers

- Can be used to understand data lineage

- Facilitate auditing processes

Documentation?
Whaddya mean, documentation?

WTH isn't there any documentation?

- Benefits are not immediate

- Not popular w/ developers

- Documentation Myths
    - My software is self-explanatory
    - Documentation is always outdated
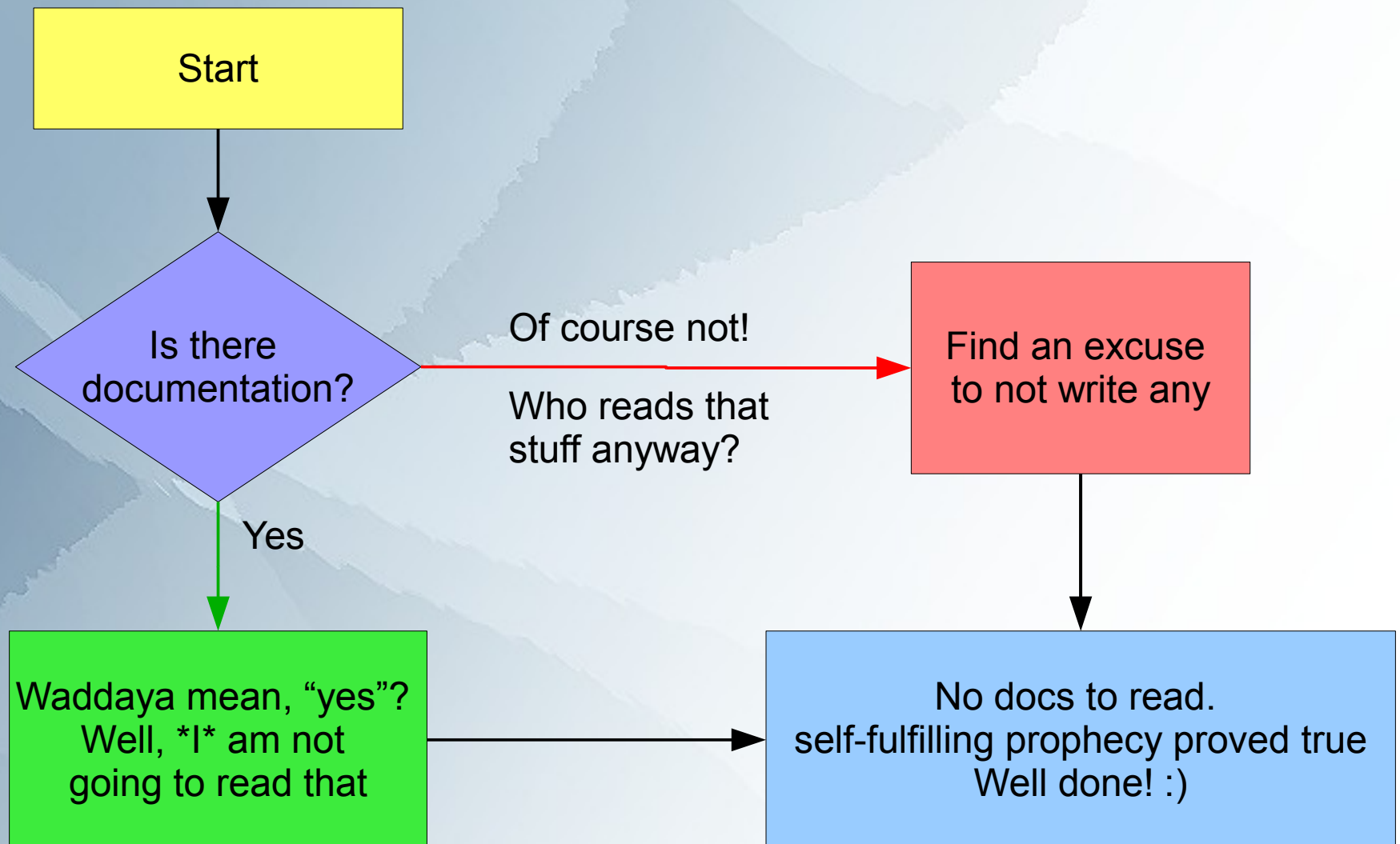   - Who reads documentation anyway?

Documentation Myths:
My Software is self-explanatory

- I already explained, it's self-explanatory.

- Software is only self-describing in the sense that it may be clear *what* it does.

- By itself, software cannot explain *why* it was built this way.

Documentation Myths: Docs are always outdated

- Yeah, documentation is always outdated. Let's blame documentation

- Documenting should be part of the development process

- You can test documentation like you can test software

Documentation Myths:
Who reads docs anyway?

Start

Is there documentation?

Of course not!

Who reads that stuff anyway?

Find an excuse to not write any

Yes

Waddaya mean, "yes"? Well, *I* am not going to read that

No docs to read. self-fulfilling prophecy proved true Well done! :)

Agenda

- Documentation
- **Introducing Kettle-Cookbook**
- Demonstration
- Roadmap
- Questions and answers
- Links and resources

Kettle-cookbook: What is it?

- A documentation generator for Kettle ETL solutions

- Built in Kettle

- Inspired by Benjamin Kallman's Kettle documentation generator (Mainz, 2008)

- Open Source (LGPL)

- Available on google code

Kettle-cookbook: How to use

- While creating/designing, enter descriptions:
  - Job and Transformation Settings
    - Description
    - Extended Description
  - Job entry, Transformation Step:
    - Description
- Run kettle-cookbook. Parameters:
  - INPUT_DIR
  - OUTPUT_DIR

Kettle-cookbook: How it works

- Kettle job scans a directory for .ktr and .kjb files creating an XML index

- XSLT is applied to XML, outputs HTML

Kettle-cookbook: Features

- Table of contents to navigate docs

- Exposes value of description fields

- Data flow Diagram

- Crosslinks

- Overviews: Variables, Connections, Fields

- Syntax highlighting (SQL, Javascript)

Kettle-cookbook: Hacking and Extending

- It's built on Kettle. Change jobs and transformations in the pdi directory to add custom processing

- Documentation generated with XSLT. Edit the kettle-report.xslt file to add custom overviews / HTML rendering

- HTML uses externalized CSS and Javacript. Hint: you'll find it in the css and js directories

- Icons in the images directory

Agenda

- Documentation
- Introducing Kettle-Cookbook
- Demonstration
- Roadmap
- Questions and answers
- Links and resources

Agenda

- Documentation
- Introducing Kettle-Cookbook
- Demonstration
- Roadmap
- Questions and answers
- Links and resources

Roadmap

- High level data flow diagrams

- Overviews (variables, connections) across ETL solution

- Replace Kettle Job with Kettle API (Benjamin Kallman)

- Dependencies / where-used list

- Not just ETL, entire Pentaho Solution (Action sequences, Mondrian Cubes, Reports)

- Data lineage

Agenda

- Documentation

- Introducing Kettle-Cookbook

- Demonstration

- Roadmap

- Questions and answers

- Links and resources

Agenda

- Documentation

- Introducing Kettle-Cookbook

- Demonstration

- Roadmap

- Questions and answers

- Links and resources

Links and resources

- Project: http://code.google.com/p/kettle-cookbook/

- Getting Started: see the project wiki

- Issues: http://code.google.com/p/kettle-cookbook/issues/list

- Downloads: https://code.google.com/p/kettle-cookbook/downloads/list

- Source: http://code.google.com/p/kettle-cookbook/source/checkout