



# Open Source Software Research

Prof. Dr. Dirk Riehle, M.B.A.

**Friedrich-Alexander-University of Erlangen-Nürnberg**

Inaugural Lecture: April 30<sup>th</sup>, 2010  
Antrittsvorlesung: 30. April 2010

This document is licensed under the Creative Commons BY-SA 3.0 license.  
© Copyright 2010 Dirk Riehle, see <http://dirkriehle.com>. Some Rights Reserved.



## Open Source

## Open Source Economics

## Open Source Engineering Research



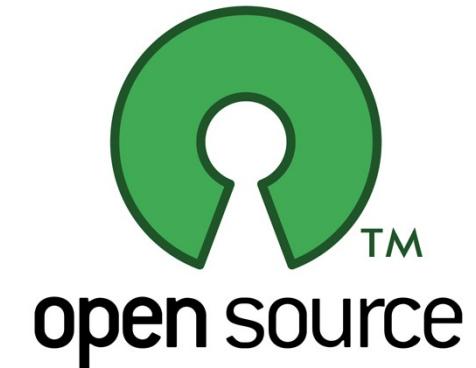


## PART I

# Open Source



- Definition of open source software
  - Software that is provided under an OSI-approved license
  - OSI = Open Source Initiative, <http://opensource.org>
  - Tried (but failed) to register the “open source” trademark
- Characteristics of an OSI-approved license
  - Source code is available and accessible
  - Modifications of code are allowed (and desired)
  - Distribution of source and binary code is unrestricted
- Free software is a variant of open source software
  - Historically, free software predates open source software
  - Free software is a subset of open source software



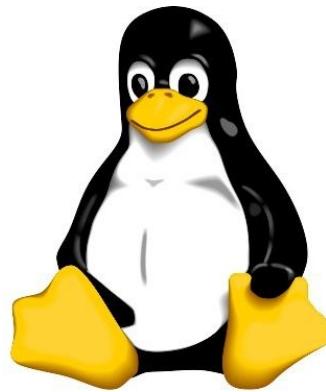
**“Open source is a development method for software that harnesses the power of distributed peer review and transparency of process. The promise of open source is better quality, higher reliability, more flexibility, lower cost, and an end to predatory vendor lock-in.”**

From <http://opensource.org/docs/osd>

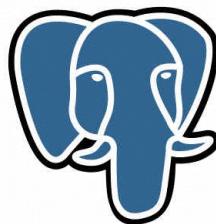
- A process: development method, marketing method to some, ...
- Of some quality: transparent, meritocratic, self-organizing, ...
- With results: better quality, more flexibility, no vendor lock-in, ...
- Frequently with a community using and developing the project



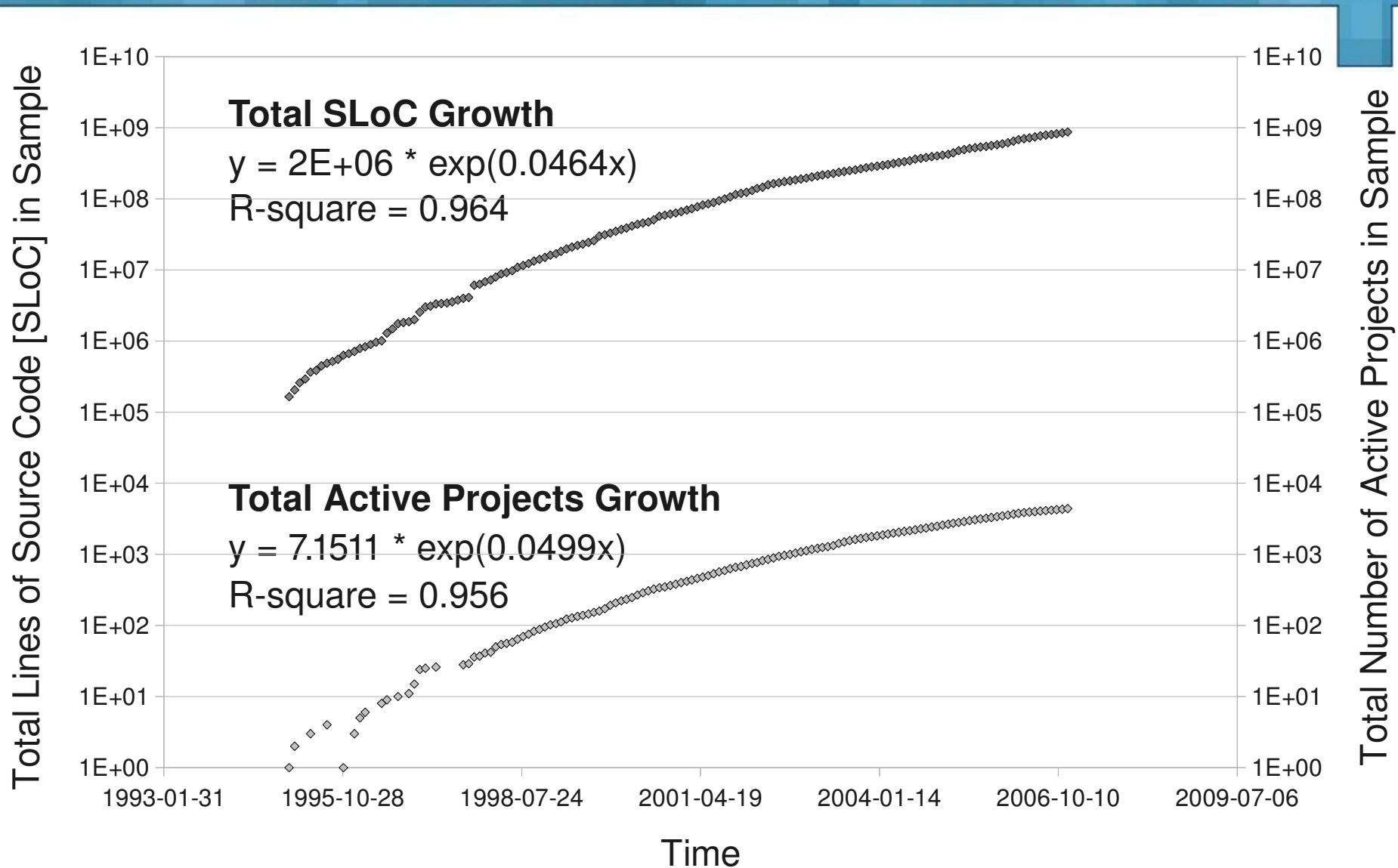
# Examples of Open Source Software



PostgreSQL



# The Total Growth of Open Source [1]



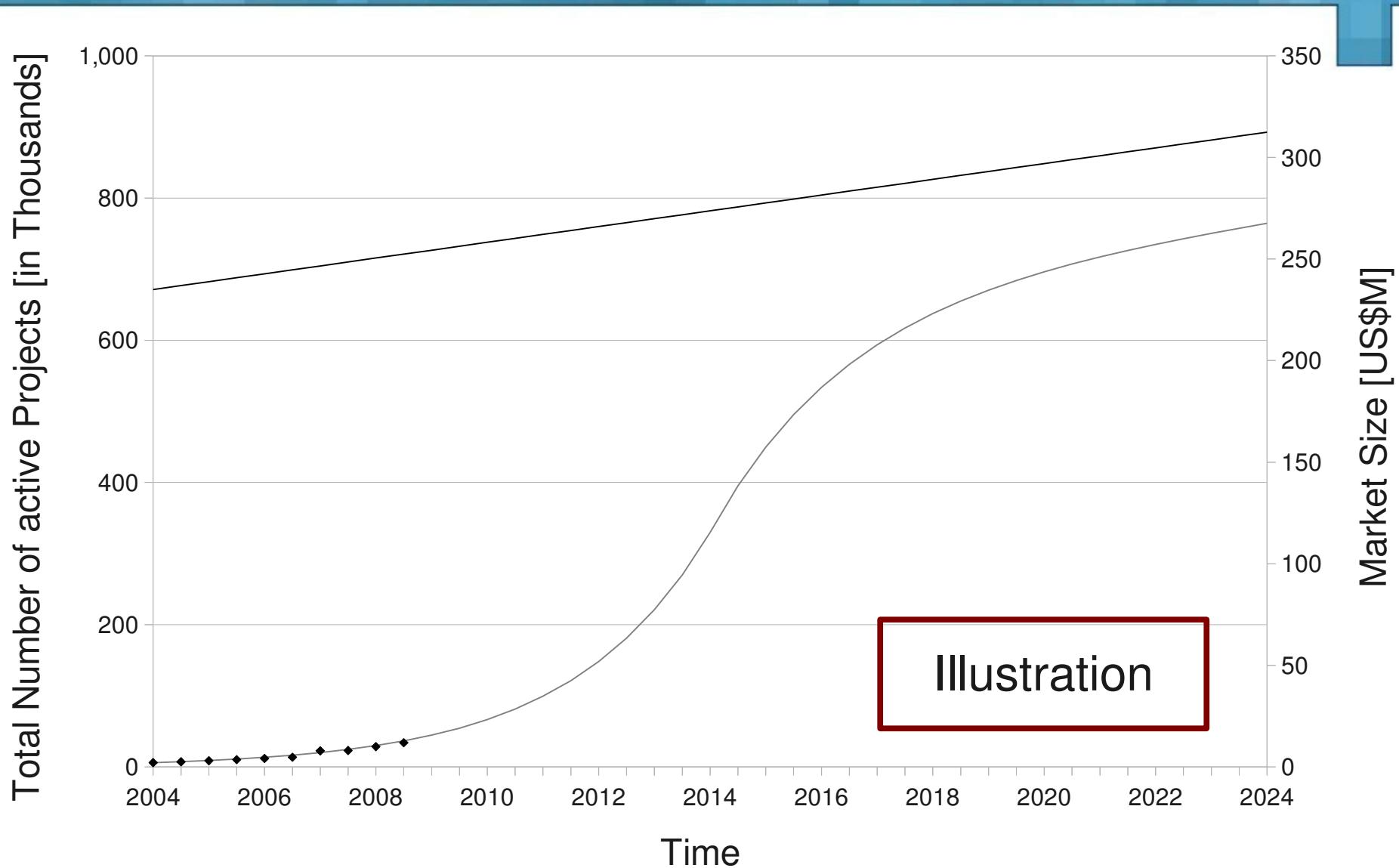
[1] Amit Deshpande, Dirk Riehle. "The Total Growth of Open Source." In Proceedings of the Fourth Conference on Open Source Systems (OSS 2008). Springer Verlag, 2008. Page 197-209.

- Commercial use of open source
  - Gartner: “By 2012, more than 90 percent of enterprises will use open source [...]” [1]
  - By and large, open source has gone mainstream, is just a product like any other
  - Today, formal open source adoption strategy elusive and TCO gains unclear
  - Open source dominates software-as-a-service and in startups
- Packaged software market
  - In 2006, open source held \$1.8B / \$235B = 0.8% of market revenue [2] [3]
  - IDC: Open source revenue will reach \$5.8B by 2011 (26% CAGR 2006-11) [2]

[1] Gartner Inc. “The State of Open Source 2008.” Gartner, 2008.

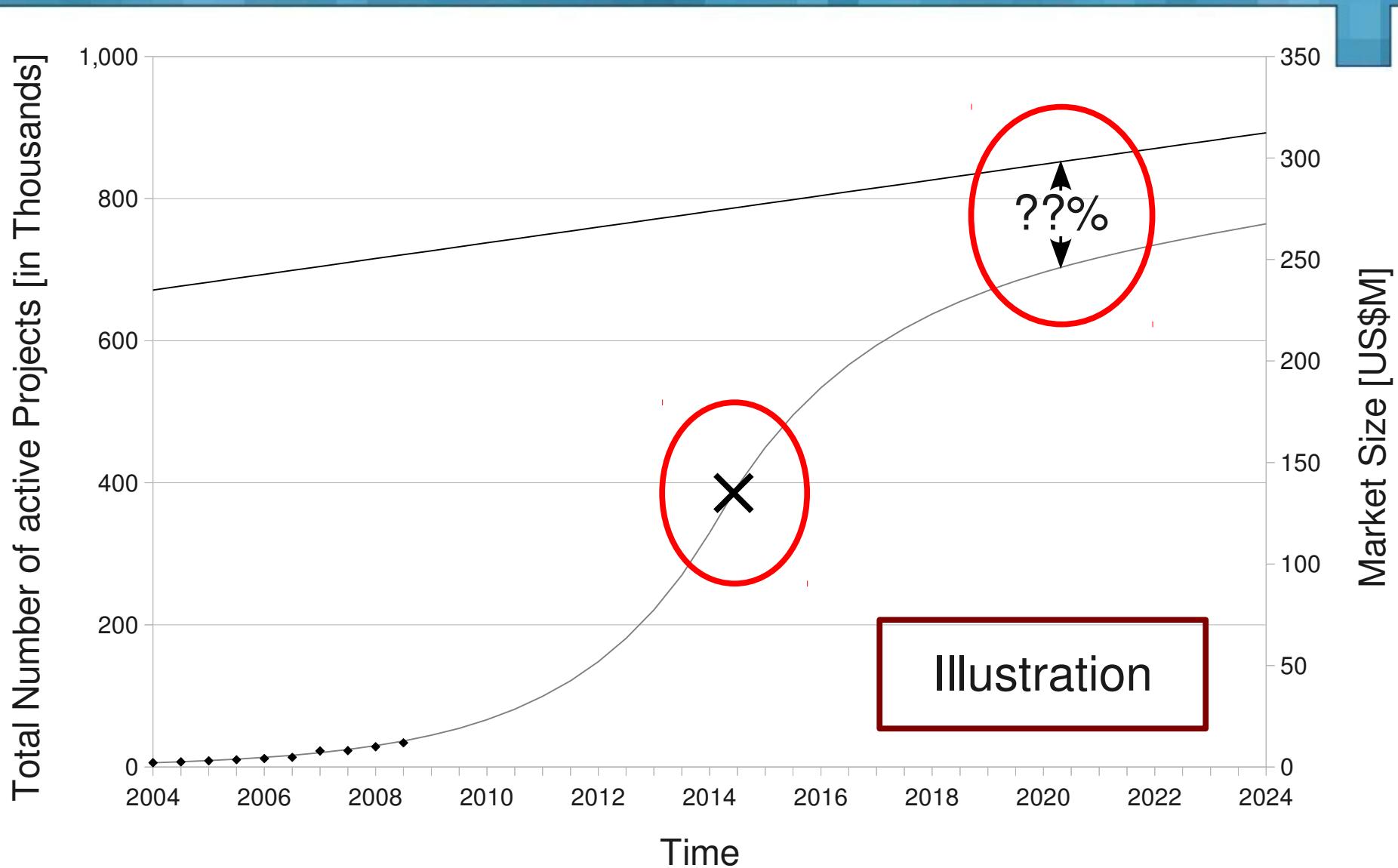
[2] IDC. “Worldwide Open Source Software Business Models 2007–2011 Forecast: A Preliminary View.” IDC, 2006.

[3] Software & Information Industry Association. “Packaged Software Industry Revenue and Growth, 2006.” SIIA, 2006.



Illustration

# Open Questions to Open Source Growth





## PART II

# Open Source Economics



## Profit Motive

## Public Welfare

## Labor Market



		Project Type	
		Single product or product line	Multi-product assembly
Ownership	Community-owned	<b>Community Open Source</b> (e.g. Linux, Apache)	<b>Community Distribution</b> (e.g. Debian)
	Single owner or dominant vendor	<b>Commercial Open Source</b> (e.g. MySQL, Alfresco)	<b>Commercial Distribution</b> (e.g. RHEL, SLES)

[1] Dirk Riehle. "The Economic Motivation of Open Source: Stakeholder Perspectives." IEEE Computer, vol. 40, no. 4 (April 2007). Page 25-32.

1991: Linux project started

1998: Open Source Initiative founded

## Traditional Community Open Source

1999: Apache Software Foundation founded

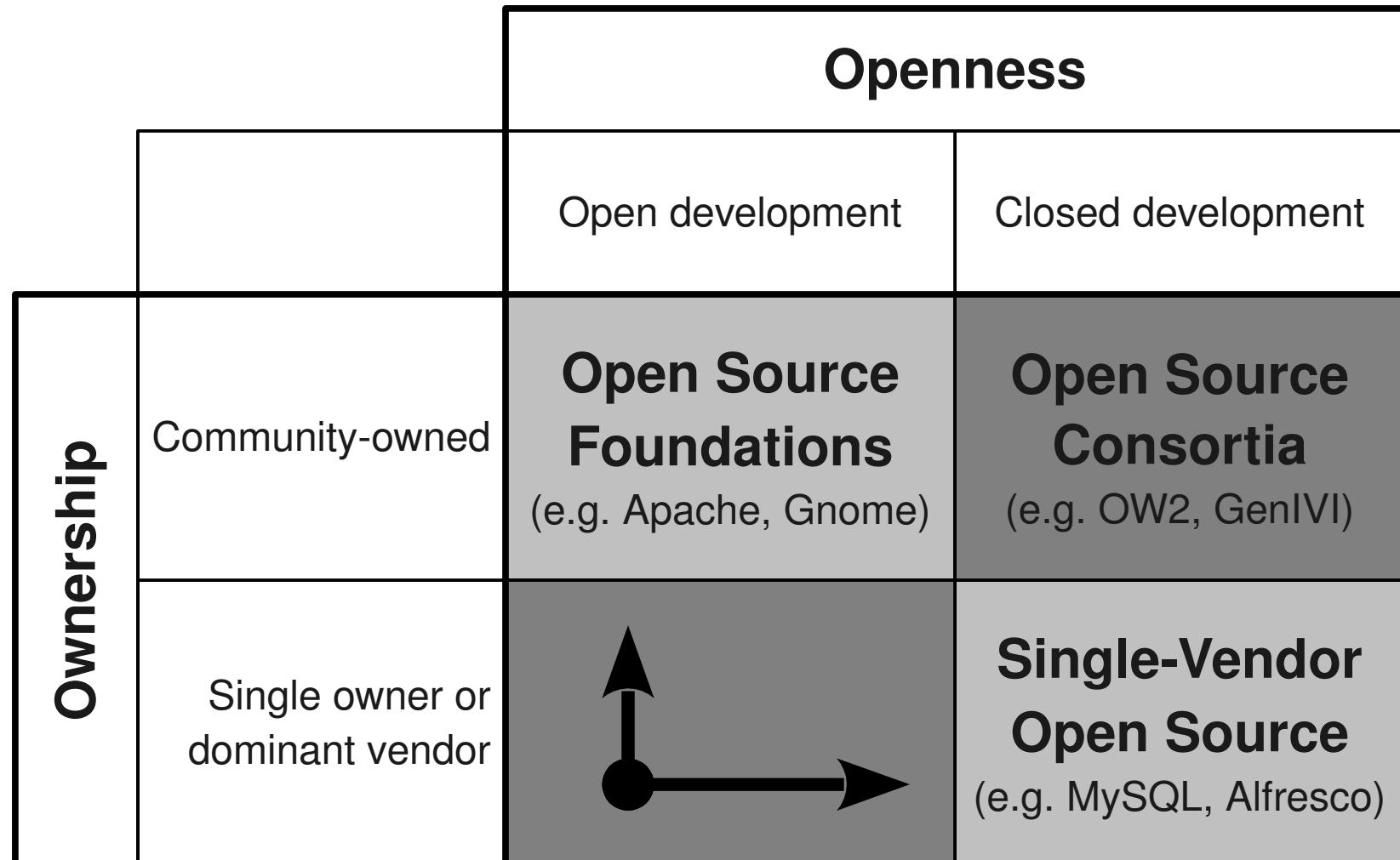
2004: Eclipse Foundation founded

## Managed Community Open Source

1995: MySQL AB founded

2001: MySQL AB funded

## Single Vendor (“Commercial”) Open Source

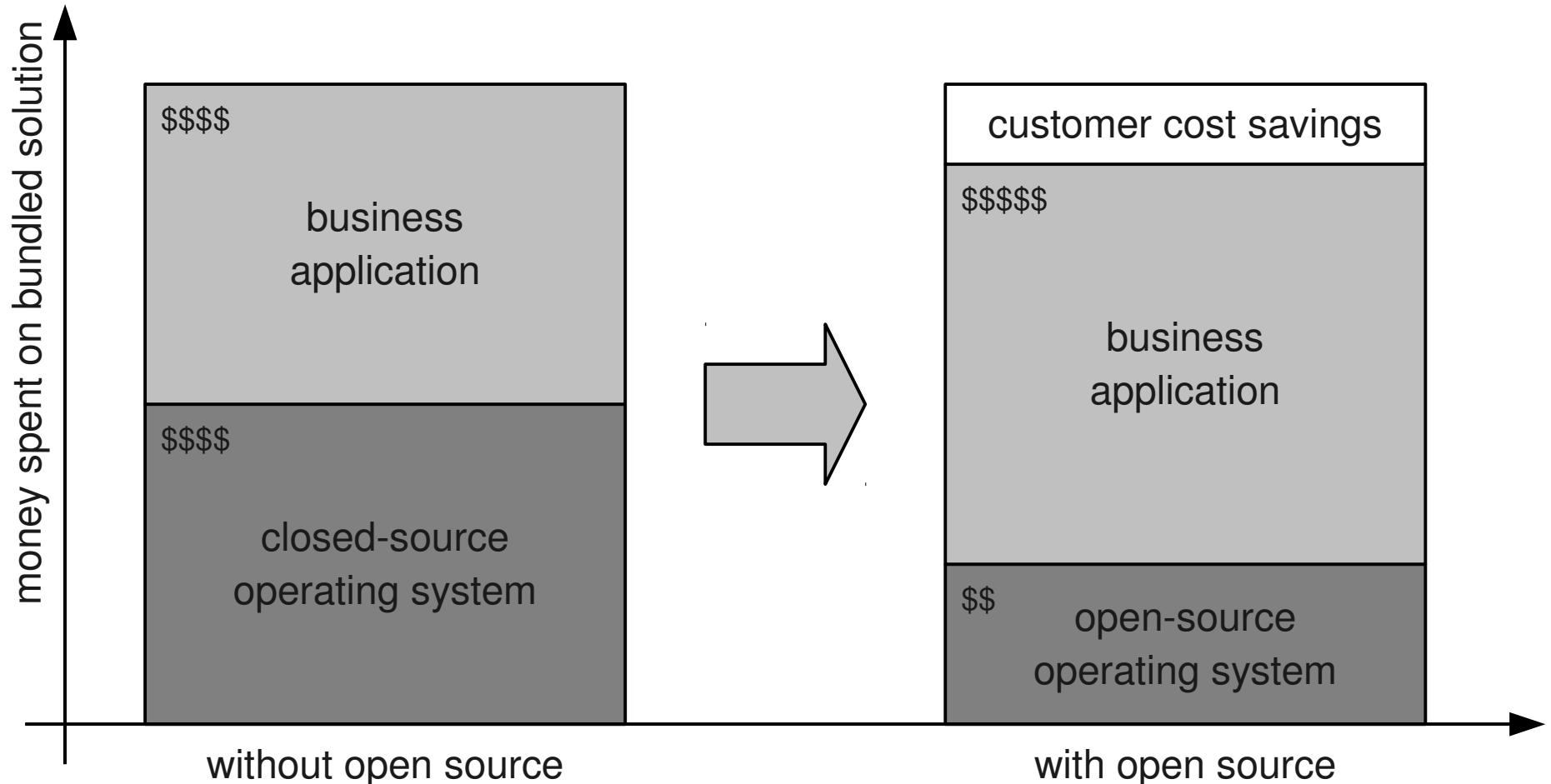


**Higher Profits per Sale**

**More Sales in a Given Market**

**Larger Addressable Market**

[1] Dirk Riehle. "The Economic Case for Open Source Foundations." IEEE Computer, vol. 43, no. 1 (January 2010). Page 86-90.



Single-Vendor Open Source	Community Open Source
Single proprietor	Community of owners
Multiple licenses	Single license
Feature differentiated	No functionality withheld
Venture-capital backed	Cross-subsidized
Significant direct revenues	Minimal direct revenues
Asymmetric community	Community of equals

All things being equal, single-vendor open source

can **go to market faster**

with a **superior product**

at **lower operational cost**

and **sell more easily**

than possible for traditional closed source software firms.

[1] Dirk Riehle. "The Commercial Open Source Business Model." Information Systems and e-Business Management. Springer Verlag, 2010. To appear.

What Position Affords	Value to Employer
Validated technical abilities (developer)	Reduced hiring risk
Deeper insight, more leverage (committer)	Better product quality
Community visibility, reputation (committer)	Higher reputation, more sales
Strategic alignment with project (committer)	Lower costs, more predictability

**Higher Salary**

**More Job Security**

**Higher Job Versatility**

**Richer Job Experience**





## PART III

# Open Source Engineering Research



Egalitarian

Meritocratic

Self-Organizing

[1] Dirk Riehle, John Ellenberger, Tamir Menahem, Boris Mikhailovski, Yuri Natchetoi, Barak Naveh, Thomas Odenwald. "Open Collaboration within Corporations Using Software Forges." IEEE Software, vol. 26, no. 2 (March/April 2009). Page 52-58.

## Open Collaboration

- Egalitarian
  - Open for contribution
  - Everyone can contribute
- Meritocratic
  - Public discussion process
  - Decisions based on merit
- Self-organizing
  - People find their own process
  - People find their best project

## Traditional Work

- Hierarchical
  - Closed and hidden silos
  - Assigned to project
- Status-oriented
  - Public and private discussions
  - Hierarchical status decides
- Assigned tasks
  - Prescribed process
  - Prescribed jobs



<b>1,000-100,000</b>	<b>User (Installations)</b>	<b>Uses the software Helps other users</b>
<b>10-100</b>	<b>Contributor</b>	<b>Provides feedback Writes code</b>
<b>1</b>	<b>Committer</b>	<b>Assures code quality Leads project</b>



# Analytics

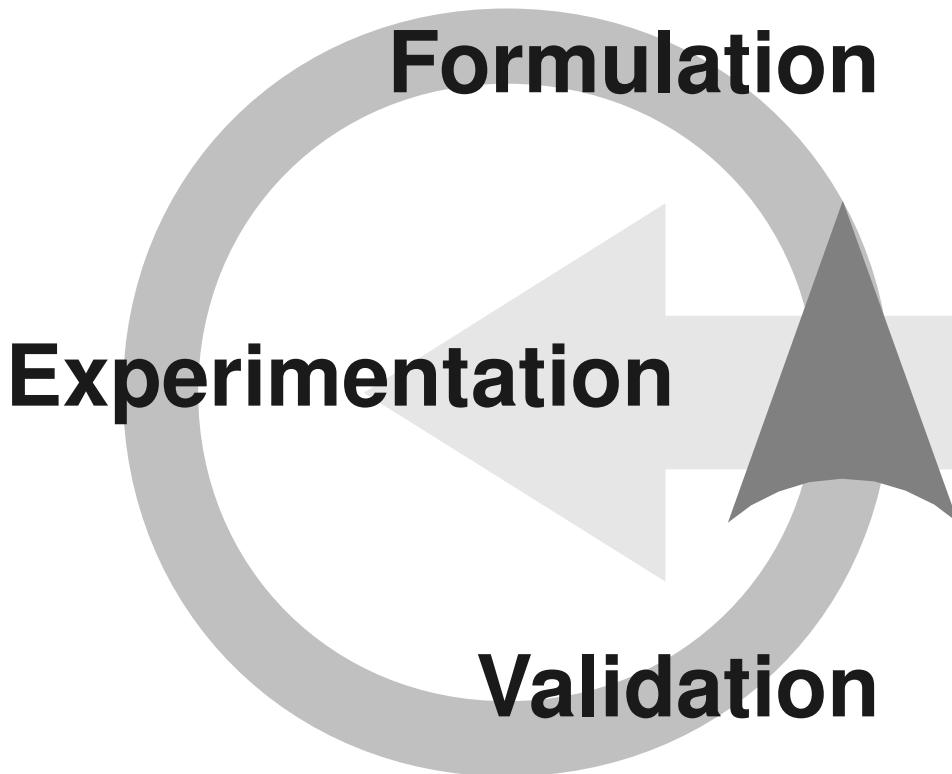
(what is)

# Innovation

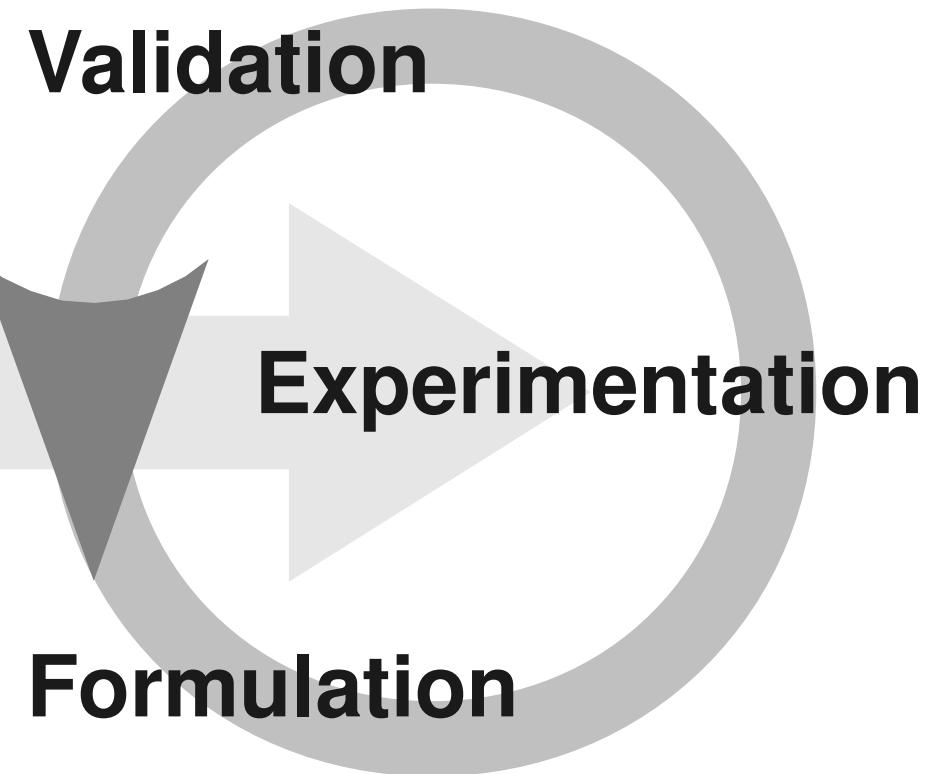
(what could be)

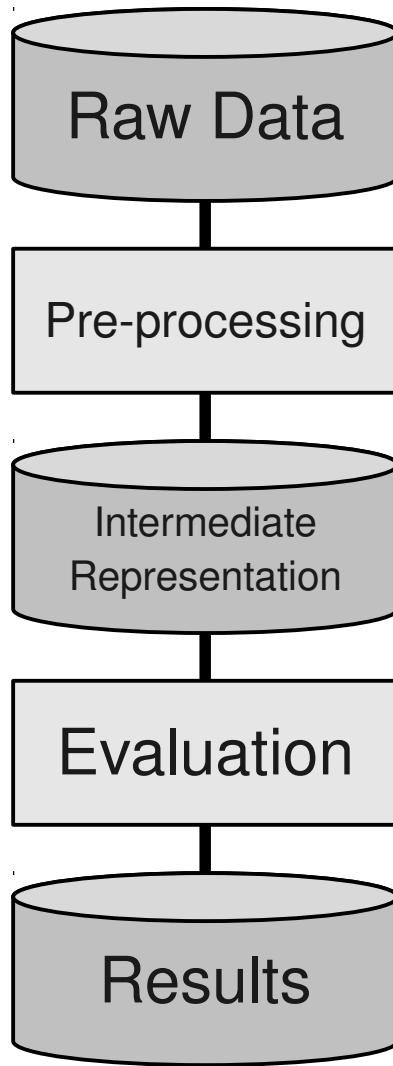


## Analytics



## Innovation

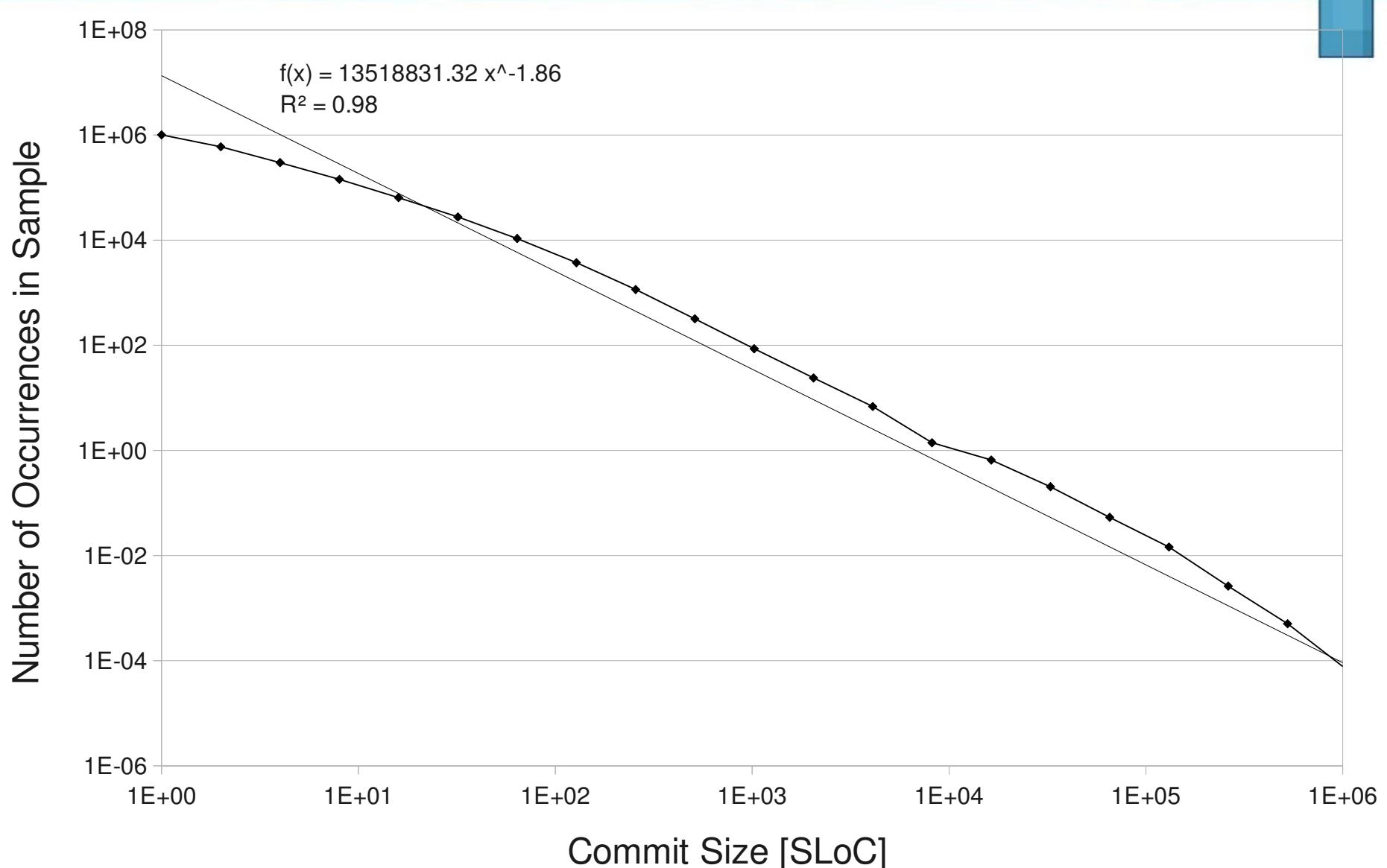




- Raw data source
  - Local database (ohloh.net, crawled sources)
  - Web services access (ohloh.net, sf.net, others)
- Pre-processing
  - Database queries using SQL and SQL scripts
  - Uses Java for computationally heavyweight filters
- Intermediate representation
  - Output of pre-processing stage for specific tasks
  - Aggregation speeds up analytical processing
- Analytical processing
  - Mines data for insight, hypothesis testing
  - Basic processing, Java programming, R-project
- Analysis output and results
  - Results of processing: averages, distributions, correlations
  - Presented as models, tables, graphs, charts, etc.

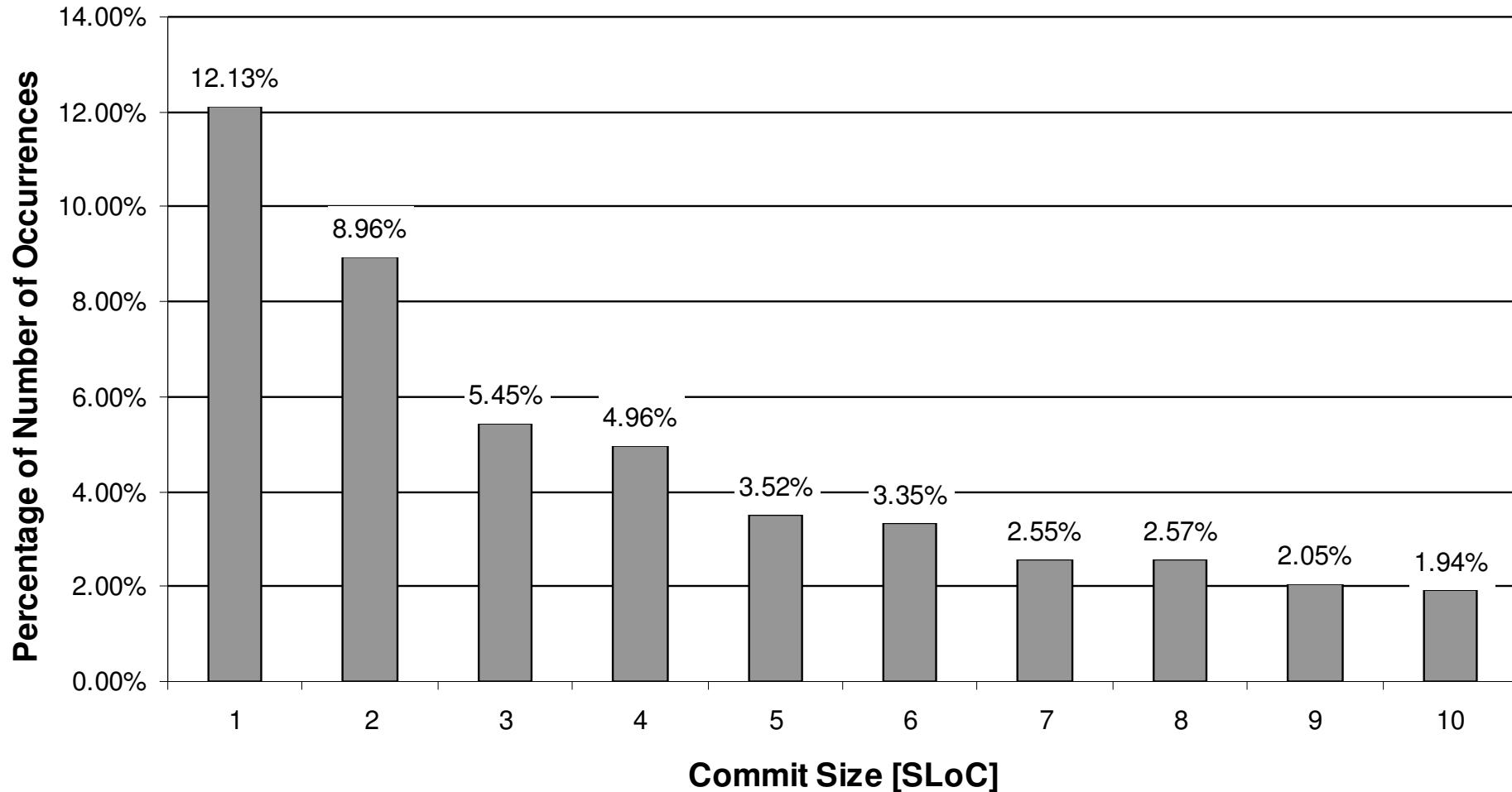


# The Commit Size Distribution of Open Source [1]



[1] Oliver Arafat, Dirk Riehle. "The Commit Size Distribution of Open Source Software." In Proceedings of the 42nd Hawaiian International Conference on System Sciences (HICSS 42). IEEE Press, 2009. Page 1-8.

# The Dominance of Small Commits



**Smaller Window Sizes  
for Merge and Code Review Tools**

**Time Series View of Commits  
rather than Update Snapshots**

**Finer-grain Execution of  
Regression Test Suites**

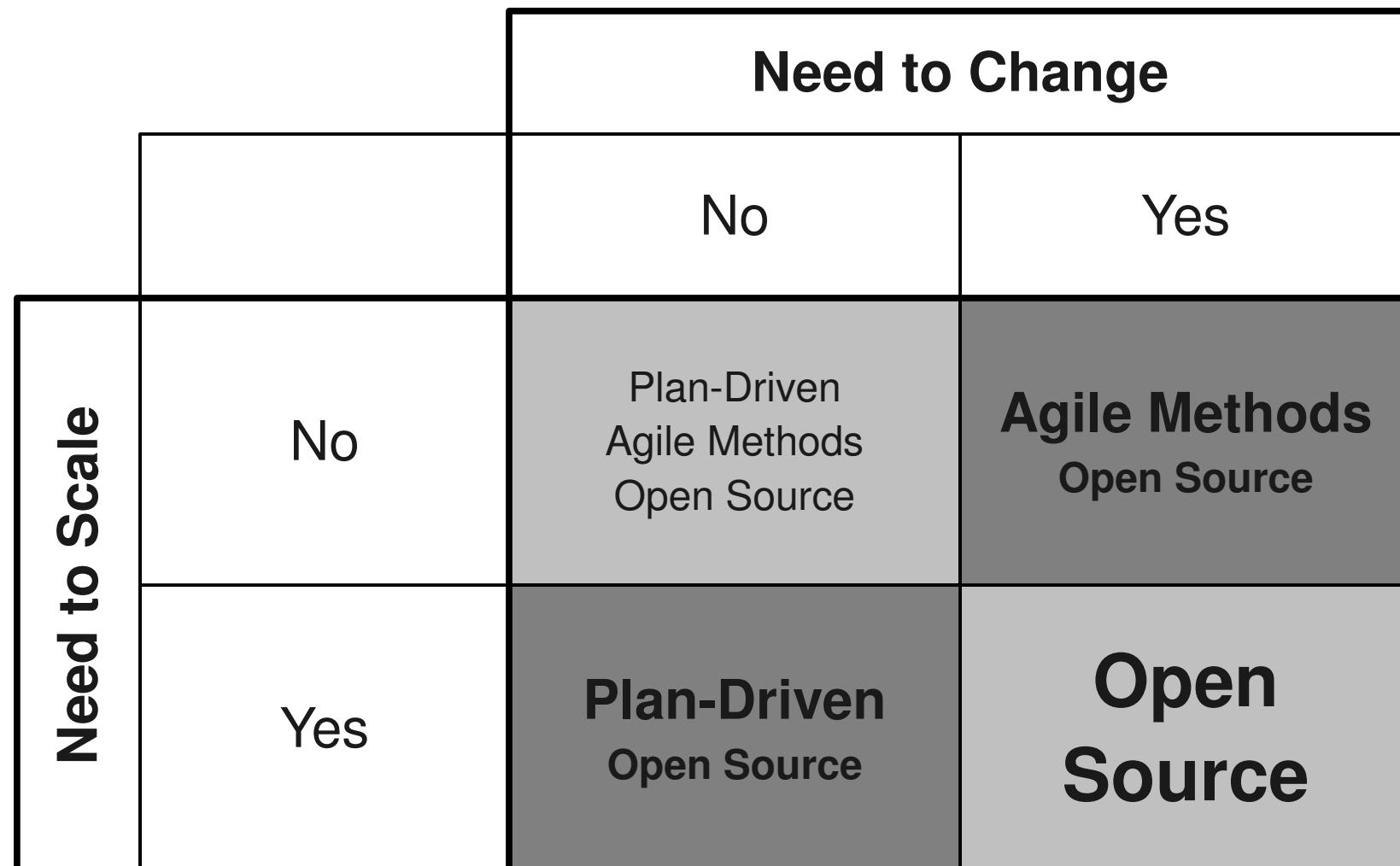


## Plan-driven Methods

## Agile Methods

## Open Source





## Agile Methods

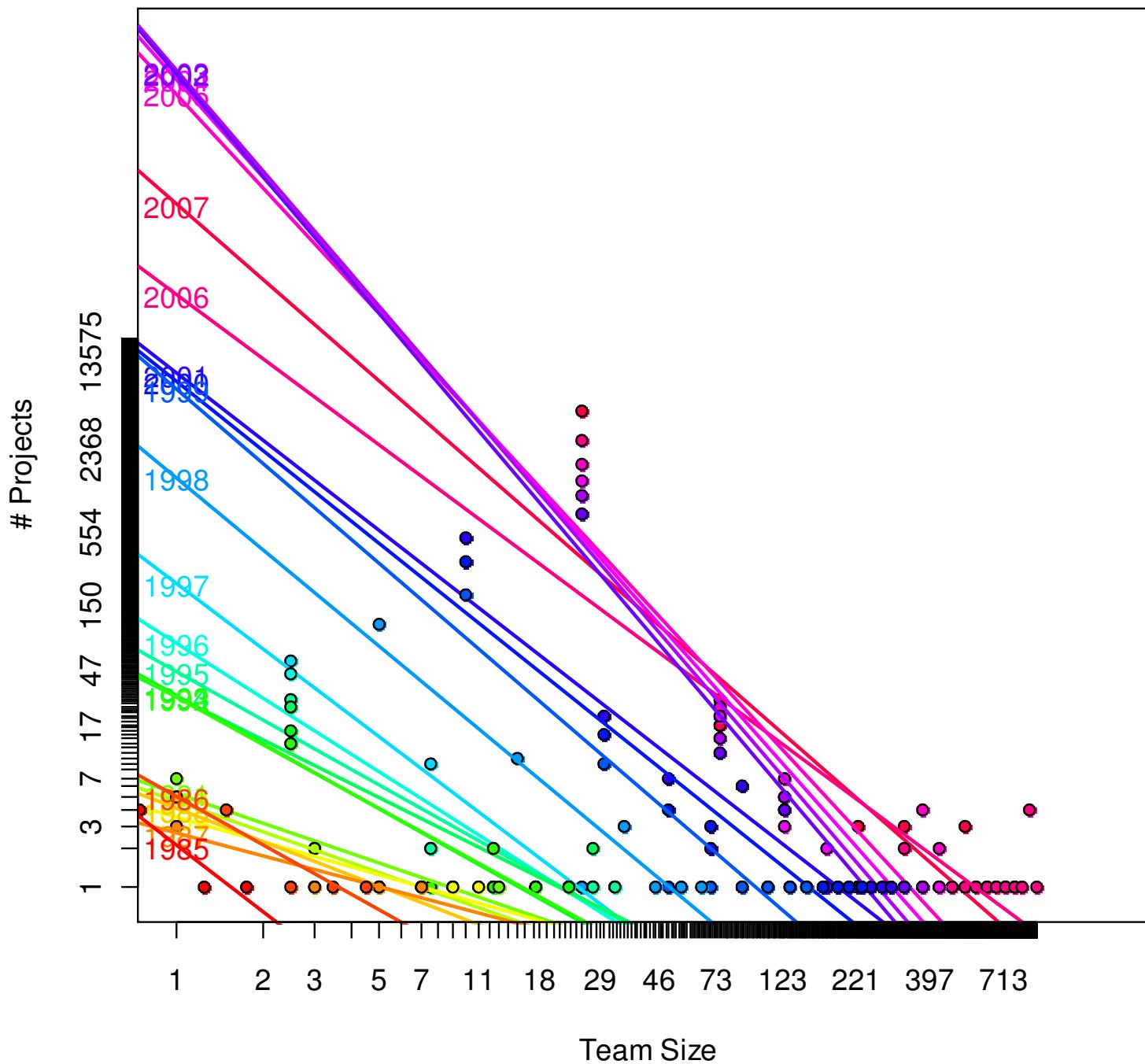
- Co-location
  - Stand-up meetings
  - Short communication paths
- Code speaks for itself
  - Self-explanatory code
  - Refactor, don't comment
- Continuous integration
  - Small focussed commits
  - Frequent commits

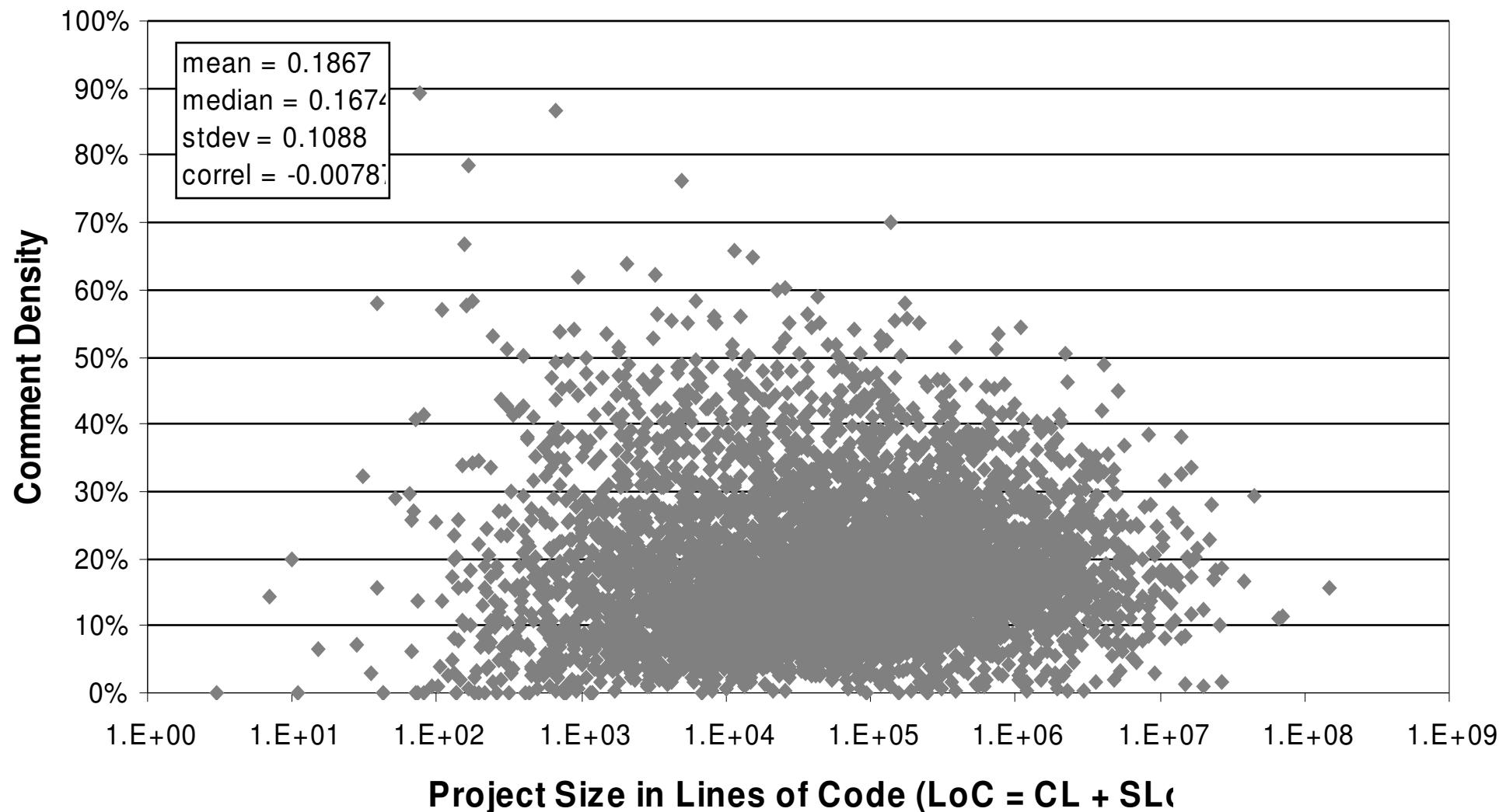
## Open Source

- Distributed development
  - Lengthy email discussions
  - Across all time zones
- Frequent commenting
  - About 20% comment density
  - Frequent comment only commits
- Has not changed practices
  - Commit size constant
  - Commit frequency constant

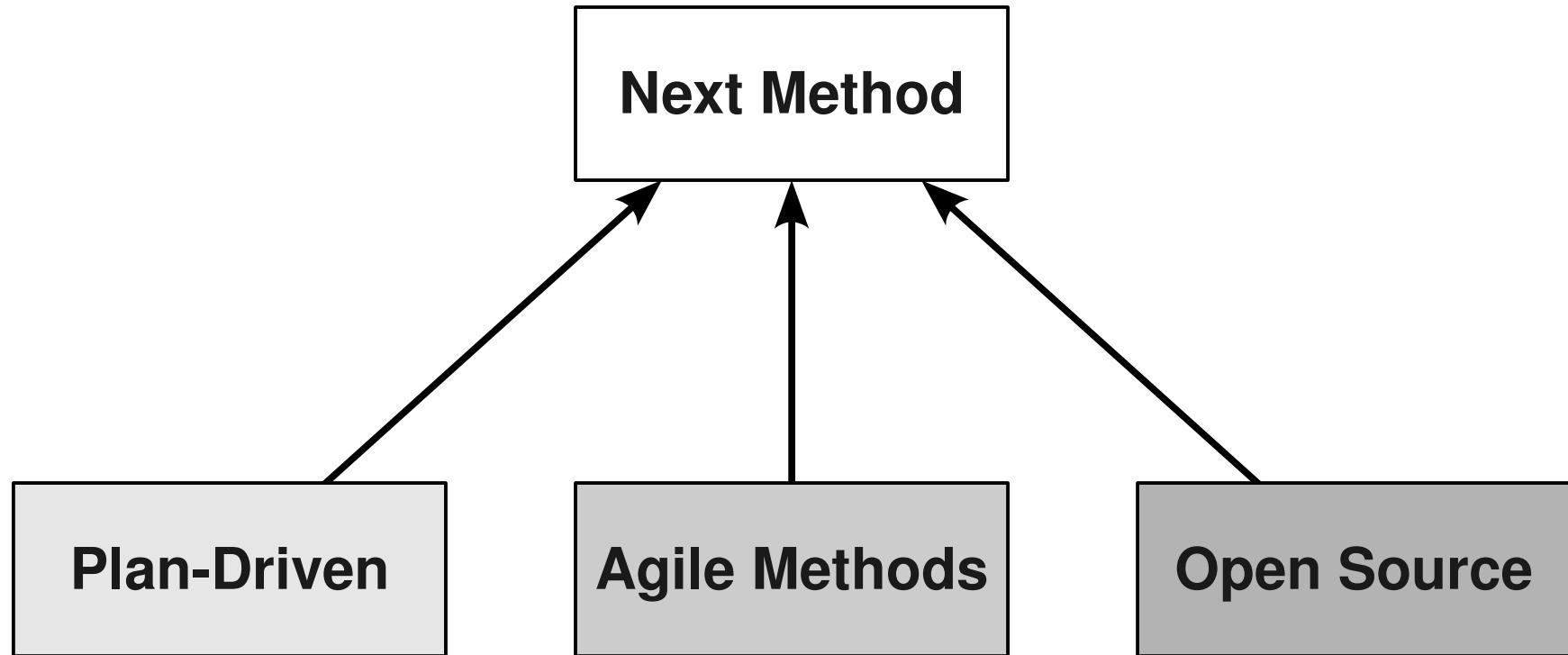


## Evolution of Team Sizes





[1] Oliver Arafat, Dirk Riehle. "The Comment Density of Open Source Software Code." In Companion to Proceedings of the 31st International Conference on Software Engineering (ICSE 2009). IEEE Press, 2009. Page 195-198. **36**



- A software forge is an
  - extensible web-based software tools platform that
  - integrates best-of-breed tools for collaborative software development
- Best known example is SourceForge.net but there are many more

The screenshot shows the SourceForge.net homepage. At the top, it displays "Registered Projects: 180,997" and "Registered Users: 1,881,080 + 250,000,000". Below this, a "Project News" section highlights "libmtx: 0.5.1 Released" (2008-07-02 11:17) and "JSMS: 4.0.3 Released!" (2008-07-02 11:17). The "Browse Software" section lists categories like Clustering, Enterprise, Multimedia, SysAdmin, Database, Financial, Networking, VolP, Desktop, Games, Security, and Development. A "Splunk Server" section is also visible at the bottom.

The screenshot shows the BerliOS Developer bug tracking system. It features a search bar for "Bugs" with options for "Require All Words" and "Search". Below the search bar, a sidebar for the "J2ME Polish" project includes links for Project Summary, Discussion Forums, Submit Bugs, Request Support, Request Features, and Project Admin. The main area displays a table of bugs with columns for Bug ID, Summary, Category, Group, Status, Date, and Assigned To. The first few rows of the table are:

Bug ID	Summary	Category	Group	Status	Date	Assigned To
012843	Using java5 causes problems with RMI	J2ME Polish Build Framework	None	Open	* 2008-Jan-04 02:54	none
012756	MIDP 2LayoutView.getNextItem() throws exception	J2ME Polish GUI	None	Open	* 2007-Dec-20 14:49	none
012753	Bug in TextField.initContent for BlackBerry	J2ME Polish GUI	None	Open	* 2007-Dec-19 18:20	none
012752	Bug within record delete in RmsStorage	J2ME Polish GUI	None	Open	* 2007-Dec-19 16:47	none





Software/Group

Search

Logout | My Account

Quick Jump To...

Home

My Stuff

Project Tree

Code Snippets

Wanted

[My Personal Page](#) | [Diary & Notes](#) | [Account Maintenance](#) | [Register Project](#)

Your personal page contains lists of bugs and tasks that you are assigned, plus a list of groups that you are a member of.

**My Assigned Items**

You have no open tracker items assigned to you.

**My Submitted Items****My ABAP Connect - Bugs**

1174 BO issue with new BAPI

1173 Failure in RFC for Linux

**TRI Code Front - Bugs**

1178 URL serialization problem with UTF-8

1177 Ruby/Java bridge issue in Java 5

1176 Serialization issue with deep copies

1175 Deep copy of AST fails on BO

**Monitored Forums****SAP Forge Admin**

discussion



help

**Monitored FileModules**

You are not monitoring any files.

**My Tasks****My ABAP Connect - Next Release**

489 Add persistent queue implementation

**My ABAP Connect - To Do**

488 Revise Linux/BAPI binding

**TRI Code Front - To Do**

490 Revamp AST implementation

491 Fix DOM serialization

492 Add XSLT handler; fix extension points

**Quick Survey**

Survey not found.

**My Bookmarks**

You currently do not have any bookmarks saved.

**My Projects**

My ABAP Connect

TRI Code Front

Zrtab Compression Library



Search the entire project

Search

Advanced search

Logout My Account

Quick Jump To...

Home

My Stuff

Project Tree

Code Snippets

Wanted

TRI Code Front

Summary

Admin

Forums

Tracker

Lists

Tasks

Docs

Surveys

News

Source

Releases

The TRI Code Front is

This project has not yet categorized itself in the Trove Software Map.

Registered: 2008-06-28 00:49

Activity Percentile: 0%

[View project activity statistics.](#)[View list of RSS feeds available for this project](#)

## Developer Info

Project Admins:

Dirk Riehle

Developers:

Omar Alonso

Philipp Hofmann

[\[View Members\]](#)[\[Request to join\]](#)

## Latest File Releases

Package	Version	Date	Notes / Monitor	Download
tricodefront	tricf 0.8	September 30, 2008		<a href="#">Download</a>

[\[View All Project Files\]](#)

## Public Areas

[Project Home Page](#)[Tracker](#)

- Bugs ( 4 open / 4 total )

Bug Tracking System

- Support ( 0 open / 0 total )

## Latest News

## TRI Code Front v0.8 released!

Dirk Riehle - 2008-09-30 05:50

[\(0 Comment\)](#) [\[Read More/Comment\]](#)

## TRI Code Front v0.7 released!

Dirk Riehle - 2008-05-20 05:50

[\(0 Comment\)](#) [\[Read More/Comment\]](#)[\[News archive\]](#)

Process Step	Forge Features
Get curious	<ul style="list-style-type: none"><li>• Project of the week</li><li>• Top 10 active projects</li></ul>
Find interesting project	<ul style="list-style-type: none"><li>• Project search</li><li>• Cross-linked projects</li></ul>
Understand project	<ul style="list-style-type: none"><li>• Open forums and mailing lists</li><li>• Documented code and wikis</li></ul>
Engage with project	<ul style="list-style-type: none"><li>• One click to forum reply</li><li>• Easy to install and run</li></ul>
Stay with project	<ul style="list-style-type: none"><li>• Email updates and conversation</li><li>• Forge account/persona, reputation</li></ul>



## PART IV

# Conclusions



- Open Source is changing the software industry
  - Continued exponential growth is eating into closed source
  - Has penetrated all parts of software and user scenarios
- Open Source is a sustainable phenomenon
  - It is economically rational for software vendors
  - It does not depend on volunteer work alone any longer
- Open Source is great for software engineering research
  - It is public and hence analyzable software development
  - It is novel and different from approaches like plan-driven or agile



This document is licensed under the Creative Commons license BY-SA 3.0,  
see <http://creativecommons.org/licenses/by-sa/3.0/>

The document's original author and copyright holder is Dirk Riehle. The author can be reached at [dirk@riehle.org](mailto:dirk@riehle.org). The document can be found at <http://dirkriehle.com>.

You are allowed to share and create derivatives of this document, even for commercial purposes, but you must keep this attribution page included in the document and you must show it as part of your performance. The title slide must visibly bear the following two lines of text:

This document is licensed under the Creative Commons BY-SA 3.0 license.  
© Copyright 2010 Dirk Riehle, see <http://dirkriehle.com>. Some Rights Reserved.

You may add your own copyright notice for the portions of the document you added.





# Questions? Feedback!

<http://osr.cs.fau.de> - [dirk.riehle@cs.fau.de](mailto:dirk.riehle@cs.fau.de)

(Auf Deutsch: [dirk.riehle@informatik.uni-erlangen.de](mailto:dirk.riehle@informatik.uni-erlangen.de))

<http://dirkriehle.com> - [@dirkriehle](https://twitter.com/dirkriehle)

