Data Science, Empirical SE and Software Analytics: It's a **Family** Affair!

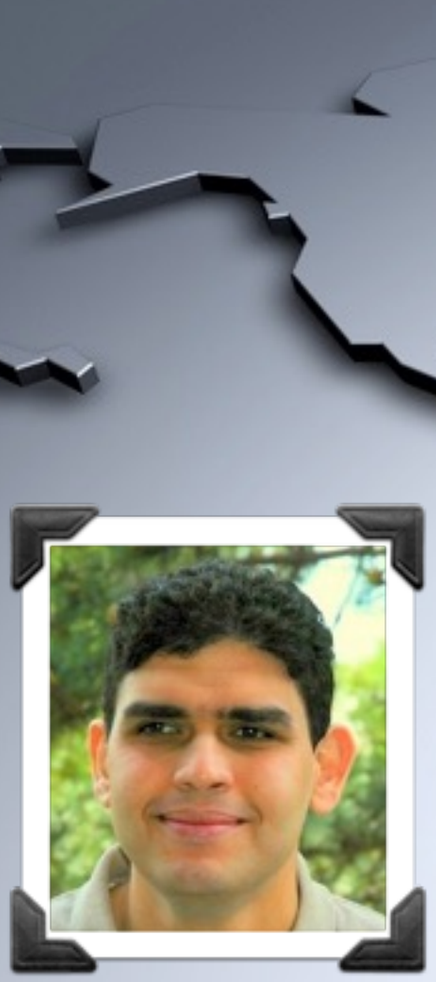
Bram Adams

http://mcis.polymtl.ca

Who is Bram Adams?

Herman Tromp
Ghent University

Wolfgang De Meuter
Vrije Universiteit Brussel



Ahmed E. Hassan
Queen's University

(Lab on Maintenance, Construction and Intelligence of Software)

POLYTECHNIQUE MONTRÉAL

LE GÉNIE EN PREMIÈRE CLASSE

Who is MCIS?

http://mcis.polymtl.ca

Who is MCIS?

Armstrong

Parisa

Doriane

Shivashree

Isabella

me

Yutaro

Javier

http://mcis.polymtl.ca

MSR 2019 Sun 26 - Mon 27 May 2019 Montreal, QC, Canada co-located with ICSE 2019

Write a Blog >>

Attending Tracks Committees Search Series

Mining Software Repositories 2019

The Mining Software Repositories (MSR) field analyzes the rich data available in software repositories to uncover interesting and actionable information about software systems and projects. The goal of this two-day conference is to advance the science and practice of MSR. The 16th International Conference on Mining Software Repositories is sponsored will be co-located with ICSE 2019 in Montréal, QC, Canada

MSR 2019 Tracks

Technical Papers | Mining Challenge | Data Showcase

MSR 2019 Sun 26 - Mon 27 May 2019 Montreal, QC, Canada co-located with ICSE 2019

Write a Blog >>

Attending Tracks Committees Search Series

Mining Software Repositories 2019

The Mining Software Repositories (MSR) field analyzes the rich data available in software repositories to uncover interesting and actionable information about software systems and projects. The goal of this two-day conference is to advance the science and practice of MSR. The 16th International Conference on Mining Software Repositories is sponsored will be co-located with ICSE 2019 in Montréal, QC, Canada

MSR 2019 Tracks

Technical Papers | Mining Challenge | Data Showcase

deadline mid January 2019

https://conf.researchr.org/home/msr-2019

Data Science, Empirical SE and Software Analytics: It's a **Family** Affair!

Bram Adams

http://mcis.polymtl.ca

Dear Bram;

As per our conversation during ICSE, we are pleased to invite you as a speaker in IASESE 2018 co-located with ESEM conference. We are planning to have a discussion about similarities, differences, and **synergies** between **empirical software engineering**, **data science**, **and** **mining software repositories** under the umbrella of "Empirical Software Engineering and Data science: Old Wine in a New Bottle". we are looking into a talk around 45 - 60 minutes and a round table discussion.

Attached please find the agenda for IASESE. Please don't hesitate to contact us in case of any questions.

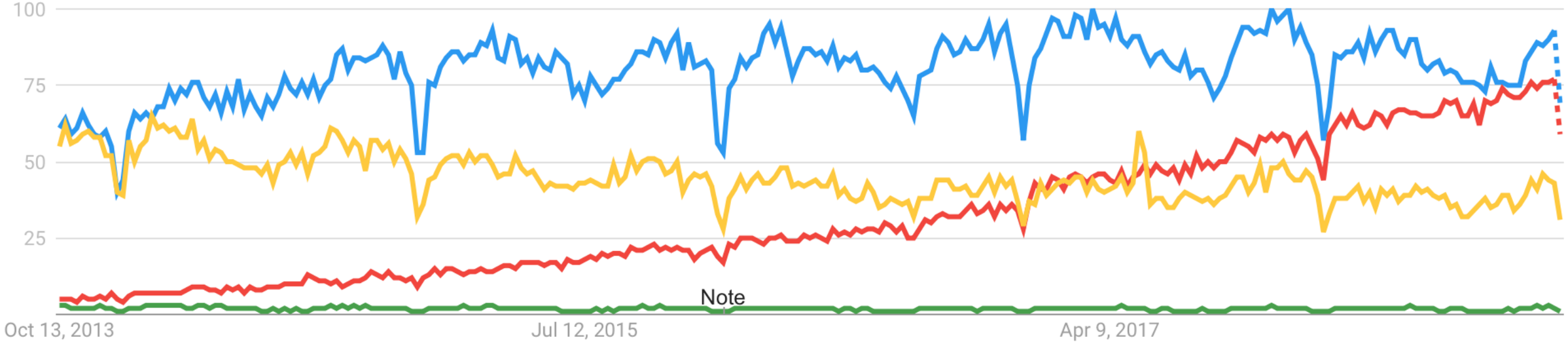
We hope to hear back from you soon;

Best Regards;
Silvia Abrahão, Maleknaz Nayebi

Data Science has Overtaken Other Major Technologies in Popularity

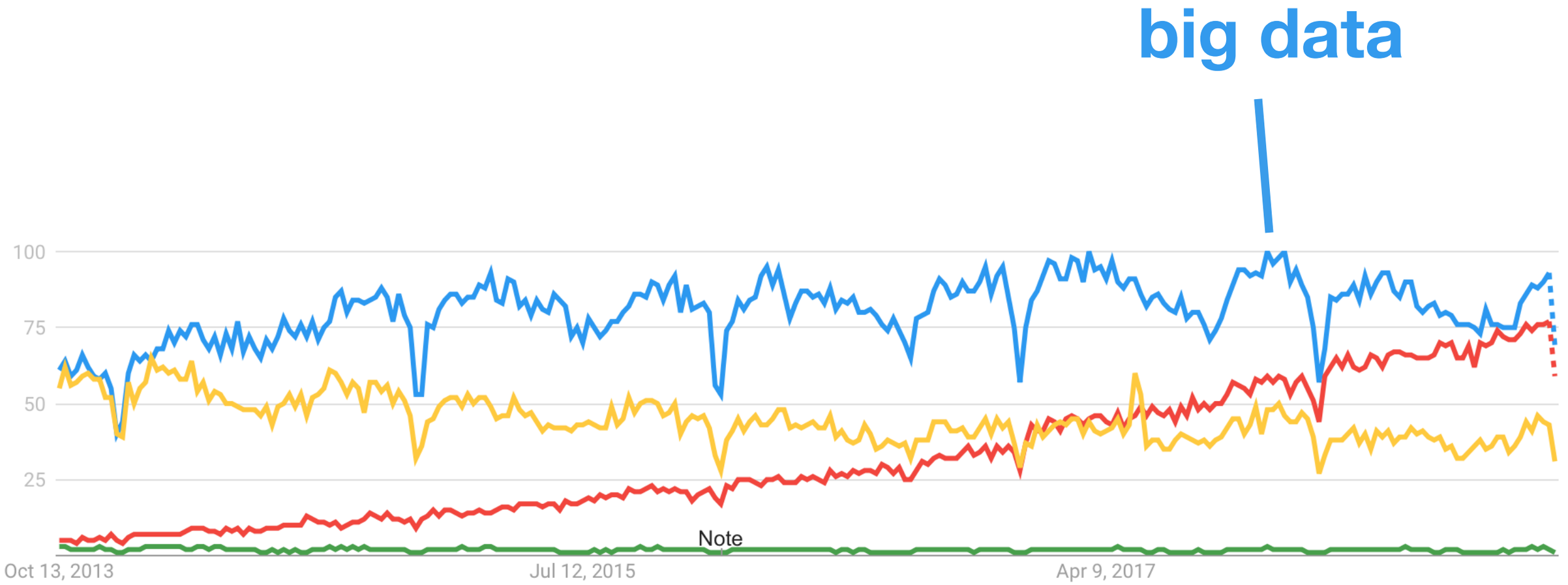
https://trends.google.com/trends/explore?date=today%205-y,today%205-y,today%205-y,today%205-y&geo=,,,&q=%22big%20data%22,%22data%20science%22,%22cloud%20computing%22,%22empirical%20research%22

Data Science has Overtaken Other Major Technologies in Popularity



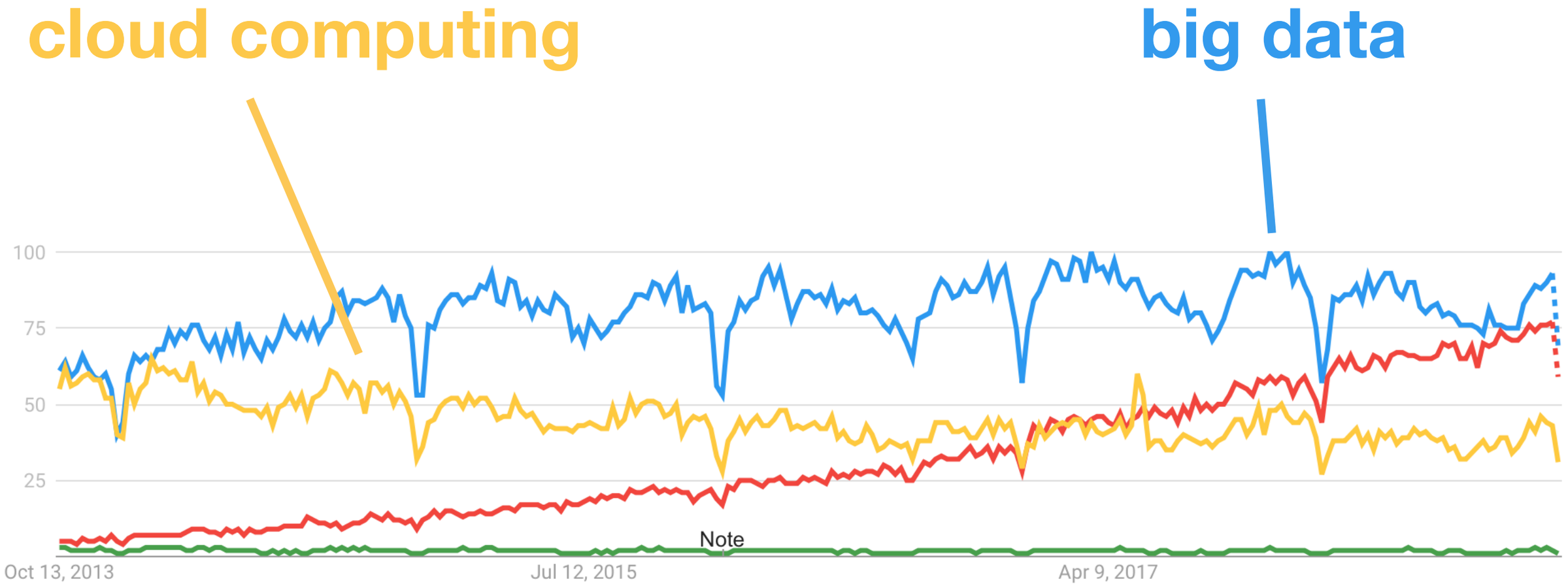
https://trends.google.com/trends/explore?date=today%205-y,today%205-y,today%205-y,today%205-y&geo=,,,&q=%22big%20data%22,%22data%20science%22,%22cloud%20computing%22,%22empirical%20research%22

Data Science has Overtaken Other Major Technologies in Popularity



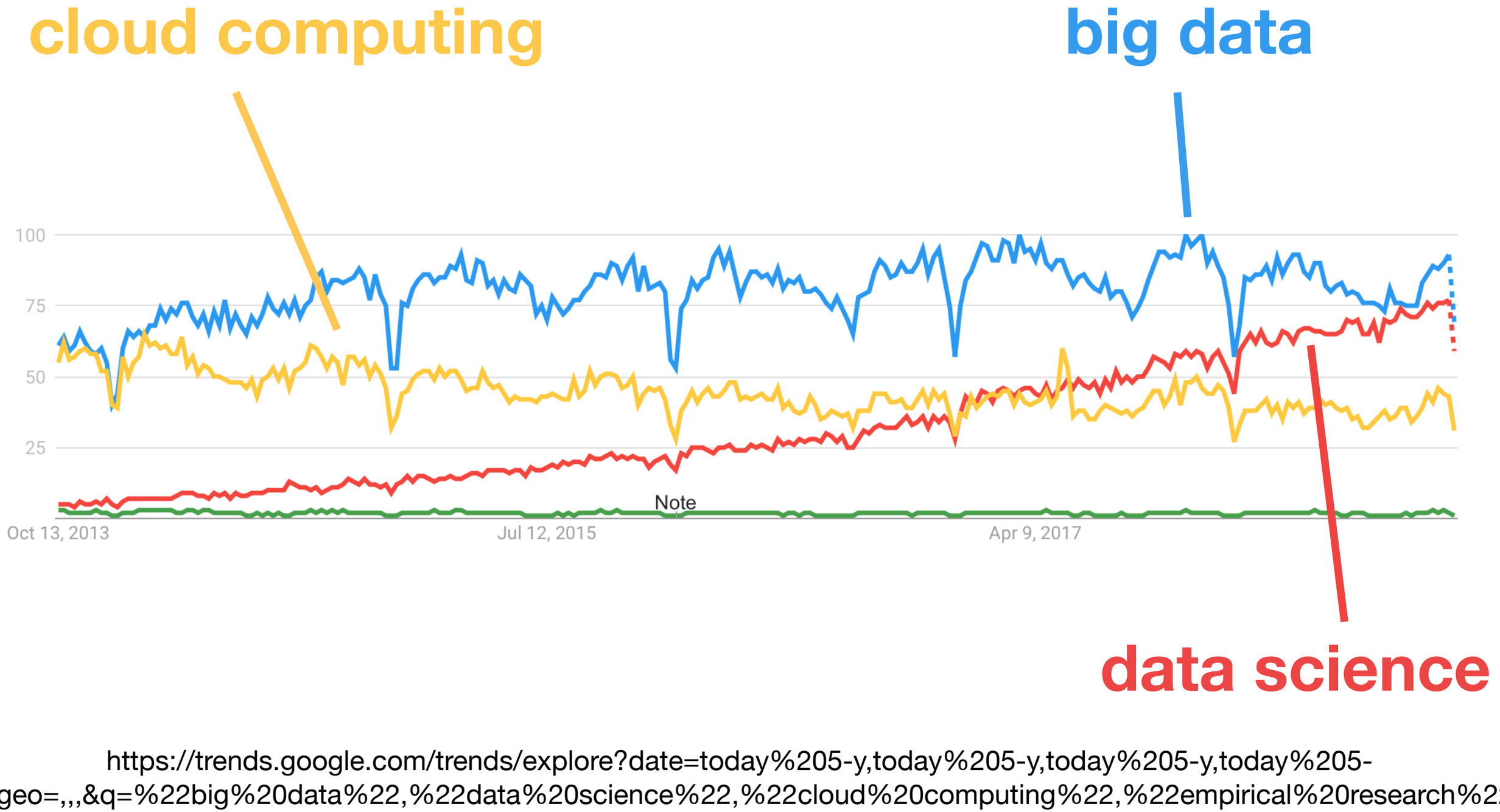
https://trends.google.com/trends/explore?date=today%205-y,today%205-y,today%205-y,today%205-y&geo=,,,&q=%22big%20data%22,%22data%20science%22,%22cloud%20computing%22,%22empirical%20research%22

Data Science has Overtaken Other Major Technologies in Popularity



https://trends.google.com/trends/explore?date=today%205-y,today%205-y,today%205-y,today%205-y&geo=,,,&q=%22big%20data%22,%22data%20science%22,%22cloud%20computing%22,%22empirical%20research%22

Data Science has Overtaken Other Major Technologies in Popularity

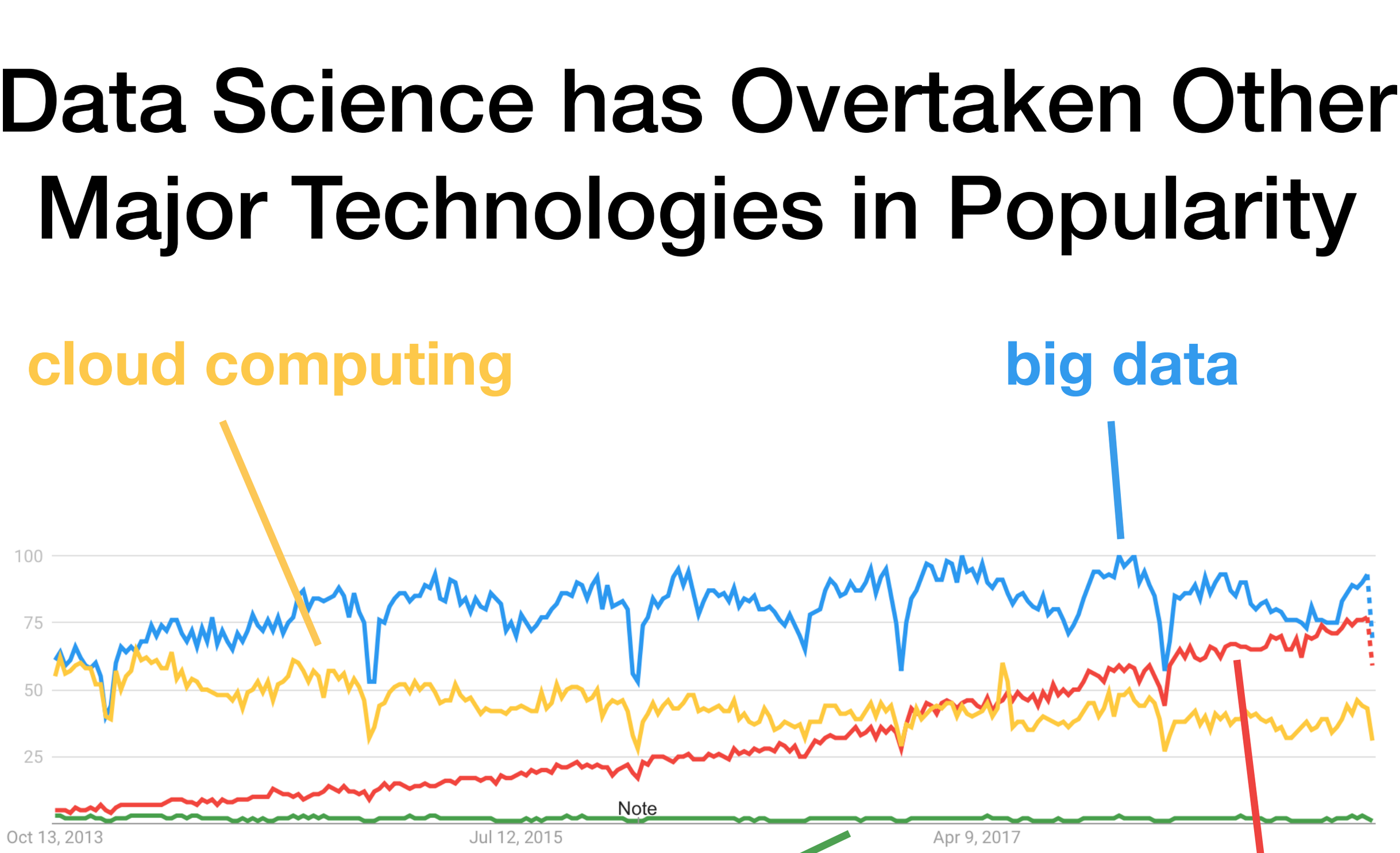


https://trends.google.com/trends/explore?date=today%205-y,today%205-y,today%205-y,today%205-y&geo=,,,&q=%22big%20data%22,%22data%20science%22,%22cloud%20computing%22,%22empirical%20research%22

Data Science has Overtaken Other Major Technologies in Popularity

cloud computing

big data



empirical research :-(

data science

https://trends.google.com/trends/explore?date=today%205-y,today%205-y,today%205-y,today%205-y&geo=,,,&q=%22big%20data%22,%22data%20science%22,%22cloud%20computing%22,%22empirical%20research%22

31,363 views | Jan 29, 2018, 02:47pm

Data Scientist Is the Best Job In America According Glassdoor's 2018 Rankings

Louis Columbus Contributor ⓘ

TWEET THIS

Data Scientist has been named the best job in America for three years running, with a median base salary of $110,000 and 4,524 job openings.

DevOps Engineer is the second-best job in 2018, paying a median base salary of $105,000 and 3,369 job openings.

https://www.forbes.com/sites/louiscolumbus/2018/01/29/data-scientist-is-the-best-job-in-america-according-glassdoors-2018-rankings/#c2d2ae55357e

TOP 100 BOOKS FOR DATA SCIENTISTS

If you want to become the wisest, most versatile data scientist you can be, then you'll love this page.

https://elitedatascience.com/books

12th International Symposium on Empirical Software Engineering and Measurement

October 11-12, 2018 - Oulu, Finland

Important Dates

(All dates end of day, anywhere on earth)

Full Research Papers
Abstract*: May 18, 2018
Full paper: May 25, 2018
Notification: July 6, 2018
Camera ready: July 27, 2018

Emerging Results and Vision Papers
Submission: July 1, 2018
Notification: August 14, 2018
Camera ready: August 28, 2018

Industrial Papers and Posters

The ACM/IEEE International Symposium on Empirical Softv
Engineering and Measurement (ESEM) is the premier confe
for presenting research results related to empirical softwa
engineering. ESEM provides a stimulating forum where
researchers and practitioners can present and discuss rec
research results on a wide range of topics, in addition to
exchanging ideas, experiences and challenging problems.

The 12th edition of ESEM will be held in Oulu, Finland, from
October 11th to 12th, 2018 as part of the Empirical Softwa
Engineering International Week with several co-located ev
including

- Annual ISERN Meeting (October 8th to 9th)
- MeGSuS Workshop (October 9th)

12th International Symposium on
Software Data Science and Measurement

October 11-12, 2018 - Oulu, Finland

SoDaSciM 2018

ESEIW Overview

ESEM Home | Organization | Call for Papers | Submission | Registration | Keynotes | Program | Venue | Trav

Important Dates

(All dates end of day, anywhere on eart

Full Research Papers
Abstract*: May 18, 2018
Full paper: May 25, 2018
Notification: July 6, 2018
Camera ready: July 27, 2018

Emerging Results and Vision Papers
Submission: July 1, 2018
Notification: August 14, 2018
Camera ready: August 28, 2018

Industrial Papers and Posters

would such a renaming make sense? why (not)?

International Symposium on Empirical Softw
nd Measurement (ESEM) is the premier confe
g research results related to empirical softwa
SEM provides a stimulating forum where
and practitioners can present and discuss rec
lts on a wide range of topics, in addition to
eas, experiences and challenging problems.

ion of ESEM will be held in Oulu, Finland, from
to 12th, 2018 as part of the Empirical Softwa
Engineering International Week with several co-located ev
including

- Annual ISERN Meeting (October 8th to 9th)
- MeGSuS Workshop (October 9th)

Part I: Data Science

Data science

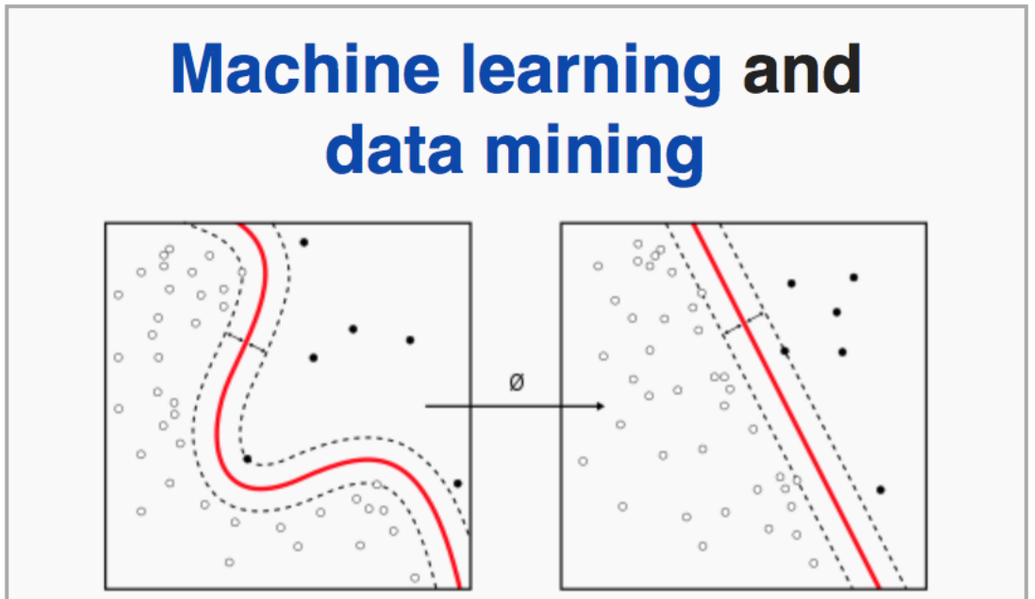
From Wikipedia, the free encyclopedia

Not to be confused with information science.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured,[1][2] similar to data mining.

Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data.[3] It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.

Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the



Data science

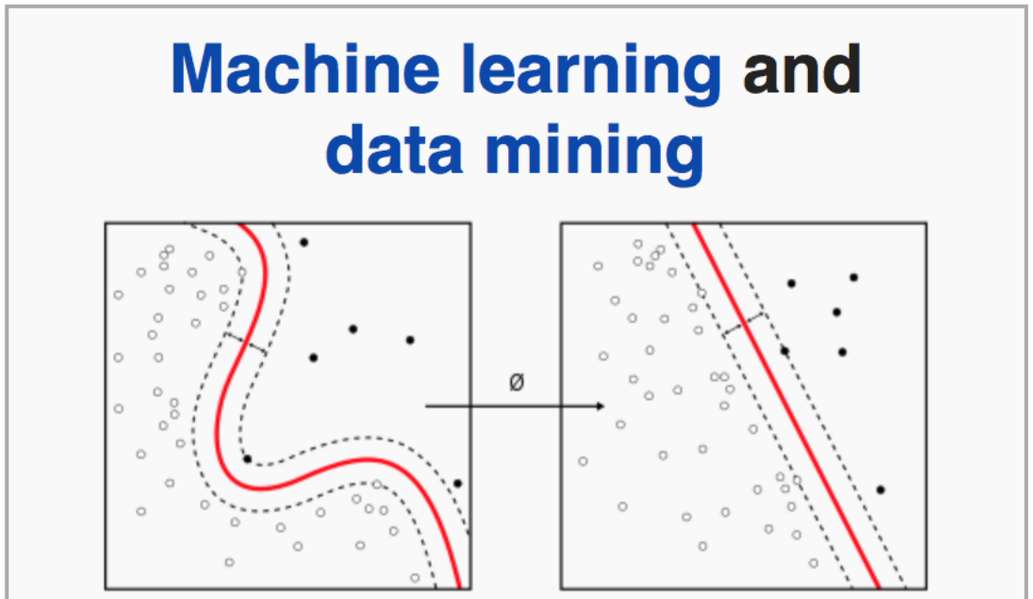
From Wikipedia, the free encyclopedia

Not to be confused with information science.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured,[1][2] similar to data mining.

Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data.[3] It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.

Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the



Data science

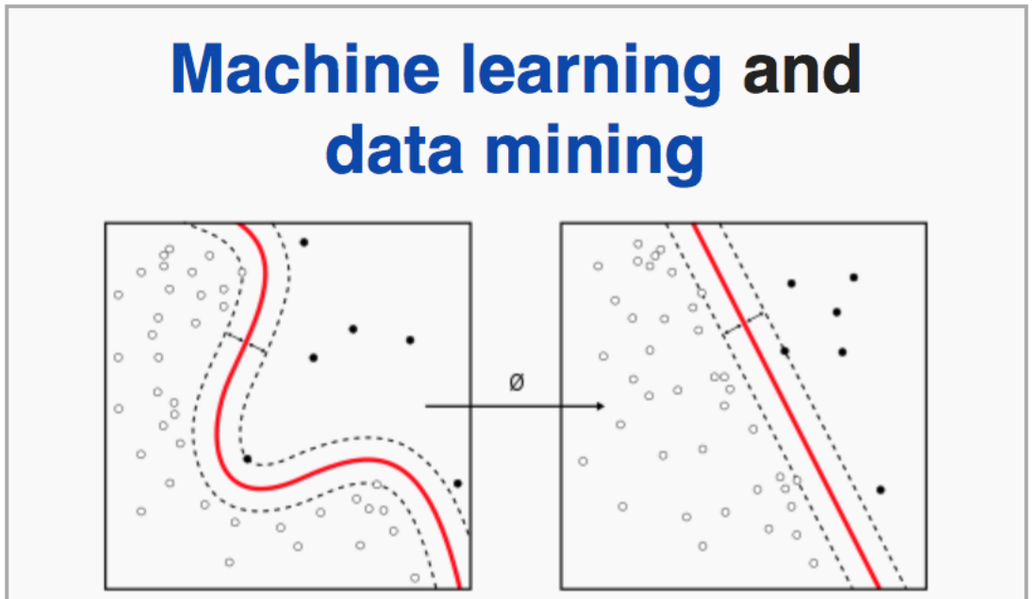
From Wikipedia, the free encyclopedia

Not to be confused with information science.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured,[1][2] similar to data mining.

Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data.[3] It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.

Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the



Data science

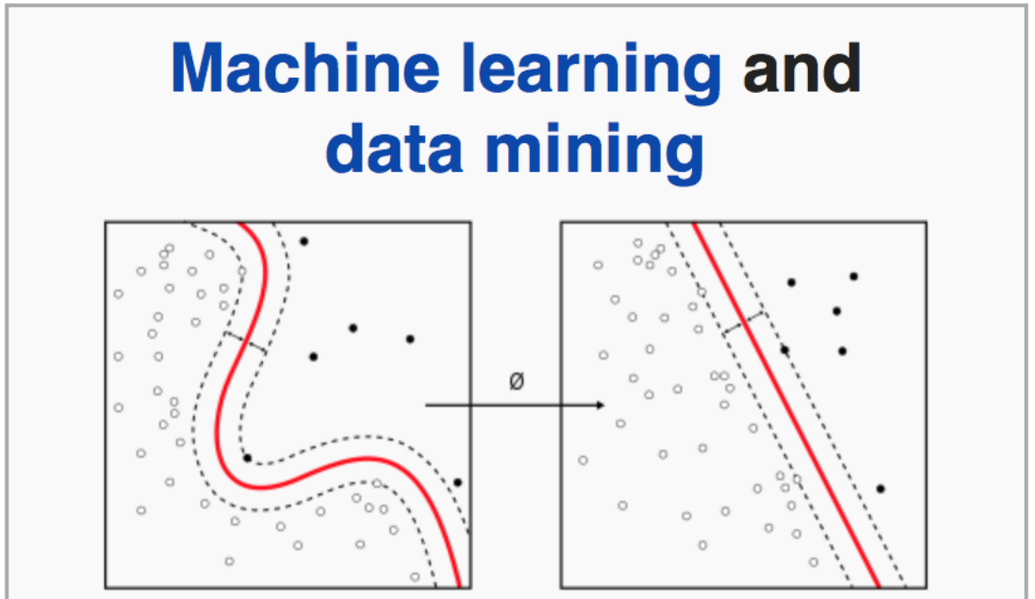
From Wikipedia, the free encyclopedia

Not to be confused with information science.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured,[1][2] similar to data mining.

Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data.[3] It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.

Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the



https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#539fcc0c55cf

I have come to feel that my central interest is in data analysis... **Data analysis**, and the parts of statistics which adhere to it, **must...take on the characteristics of** science rather than those of mathematics... data analysis is intrinsically an **empirical science**

[John Tukey, 1962]

https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#539fcc0c55cf

John Tukey, “The Future of Data Analysis”, Ann. Math. Statist., 33(1), pp. 1-67, 1962.

https://en.wikipedia.org/w/index.php?curid=17099473

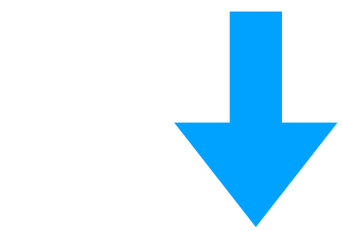
Science?

Science?

testable hypothesis

Science?

testable hypothesis



validation

Science?

testable hypothesis

validation

validated
hypothesis

Science?

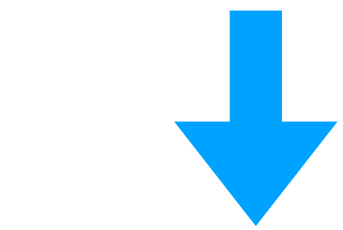
testable hypothesis

validation

validated
hypothesis

Science?

testable hypothesis

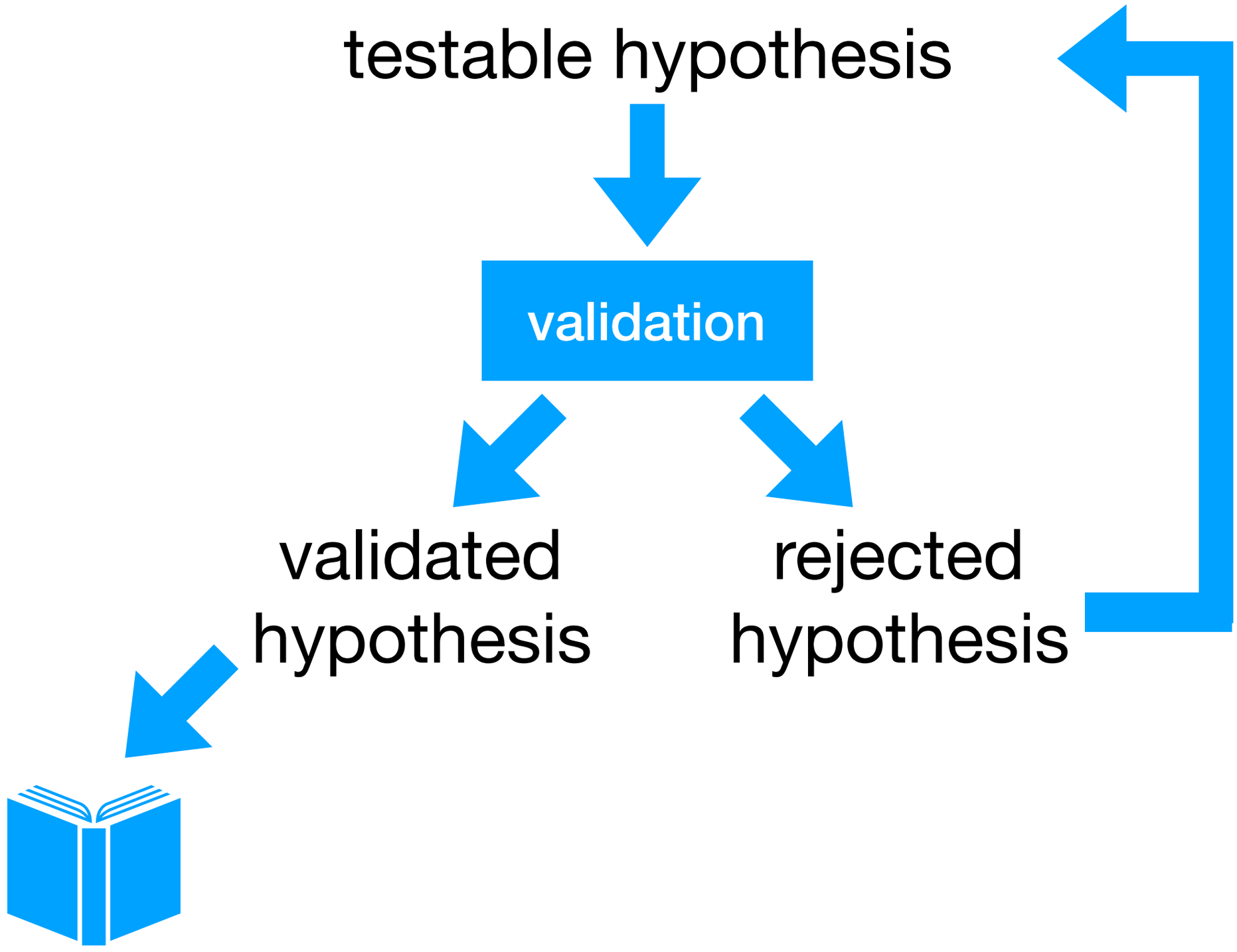


validation

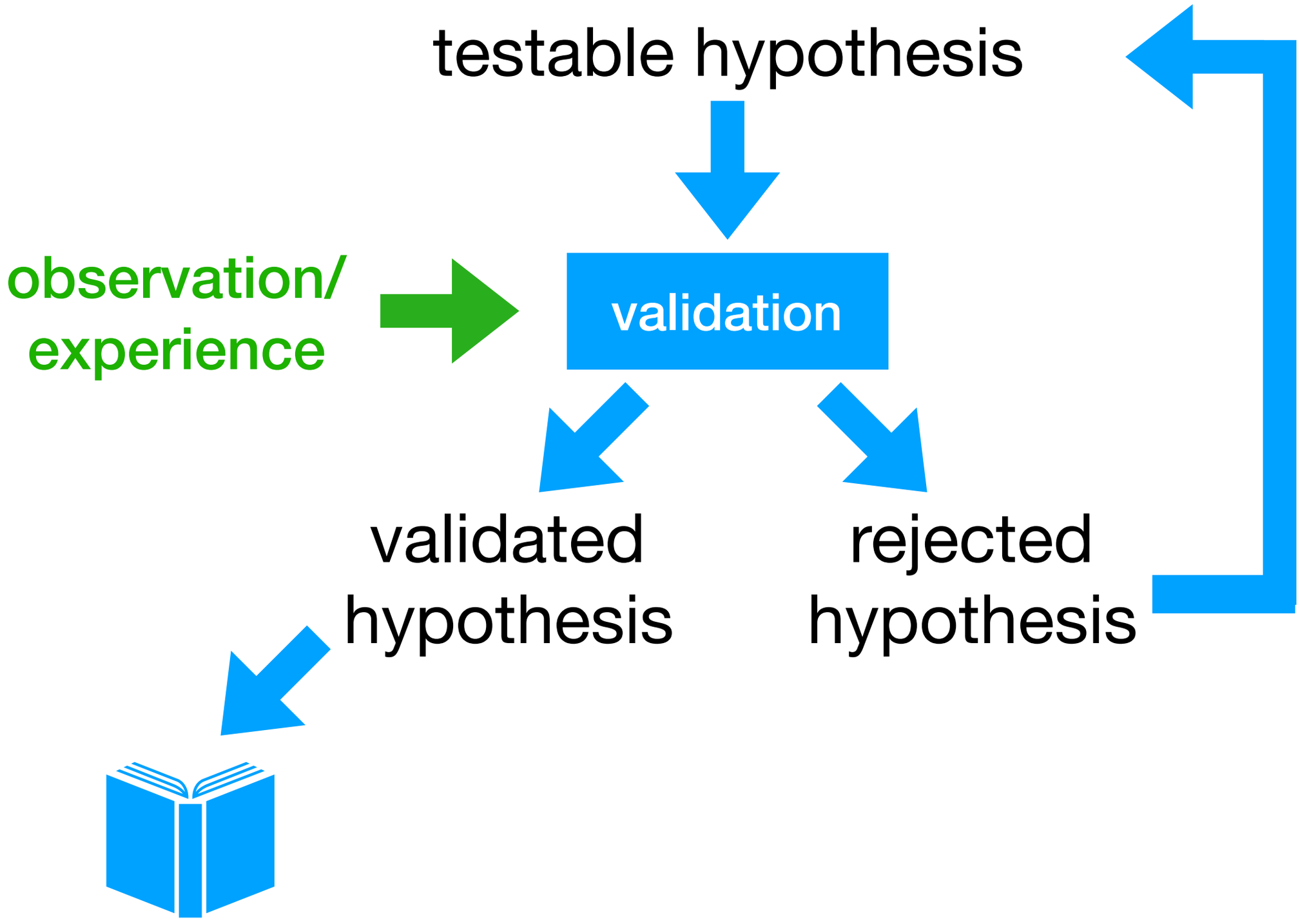
validated hypothesis

rejected hypothesis

Science?


testable hypothesis
validation
validated hypothesis
rejected hypothesis

Empirical Science?



1989: KDD Workshop & Process

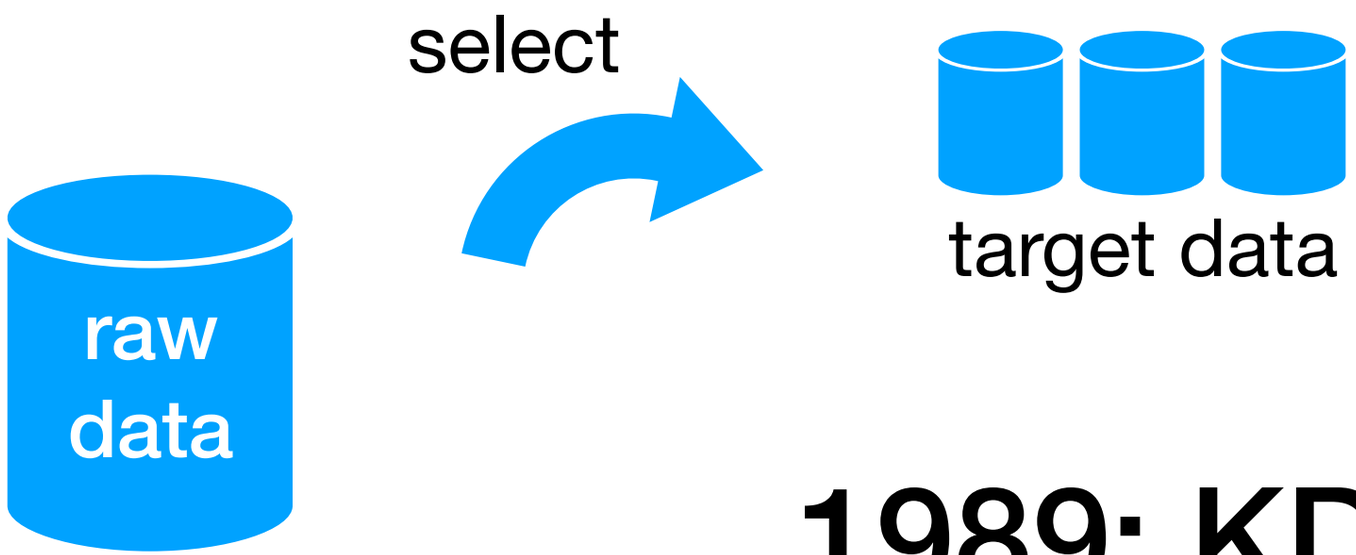
Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34

1989: KDD Workshop & Process

Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34

1989: KDD Workshop & Process

Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34

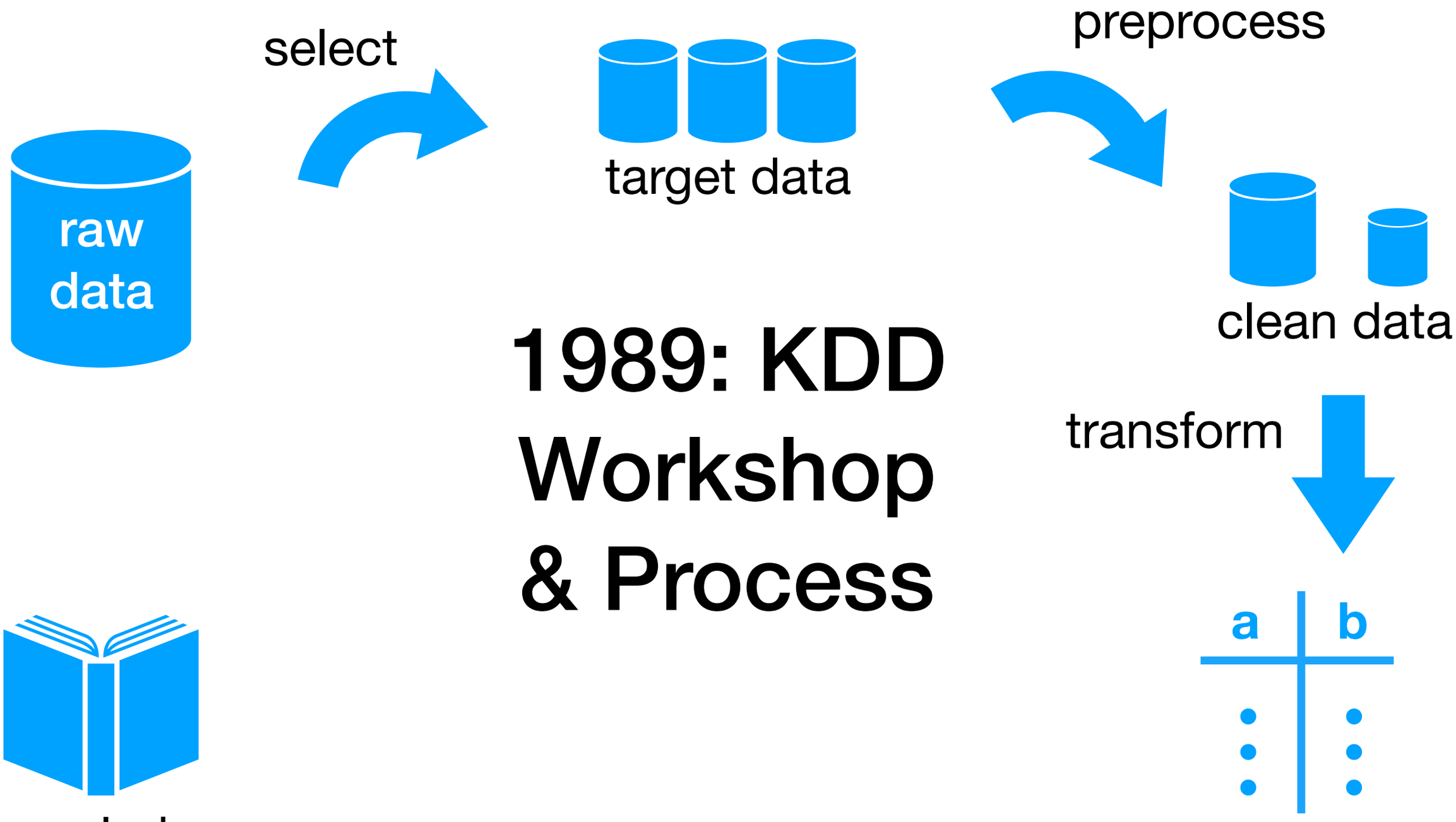


1989: KDD Workshop & Process

Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34

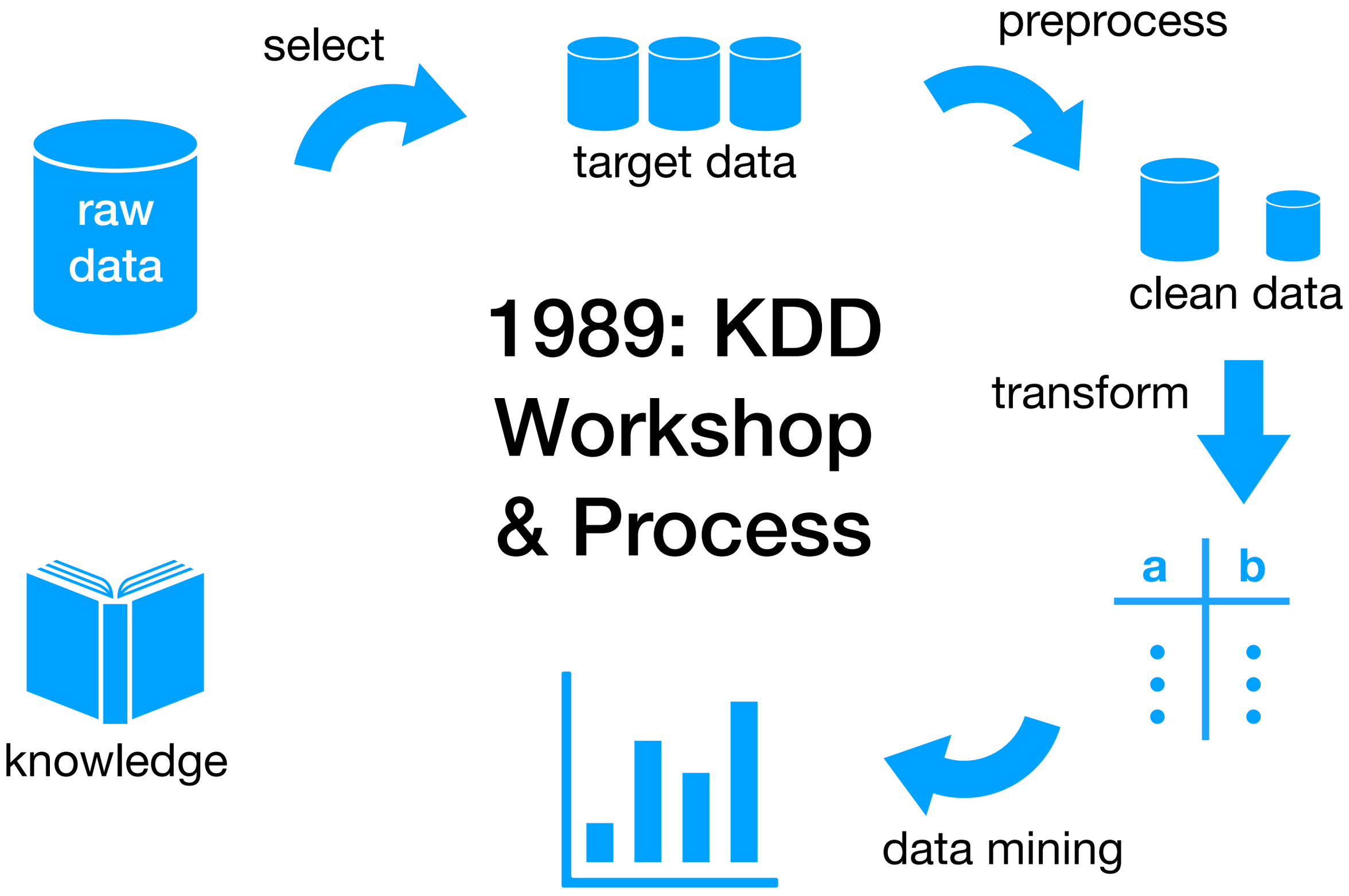
1989: KDD Workshop & Process

Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34



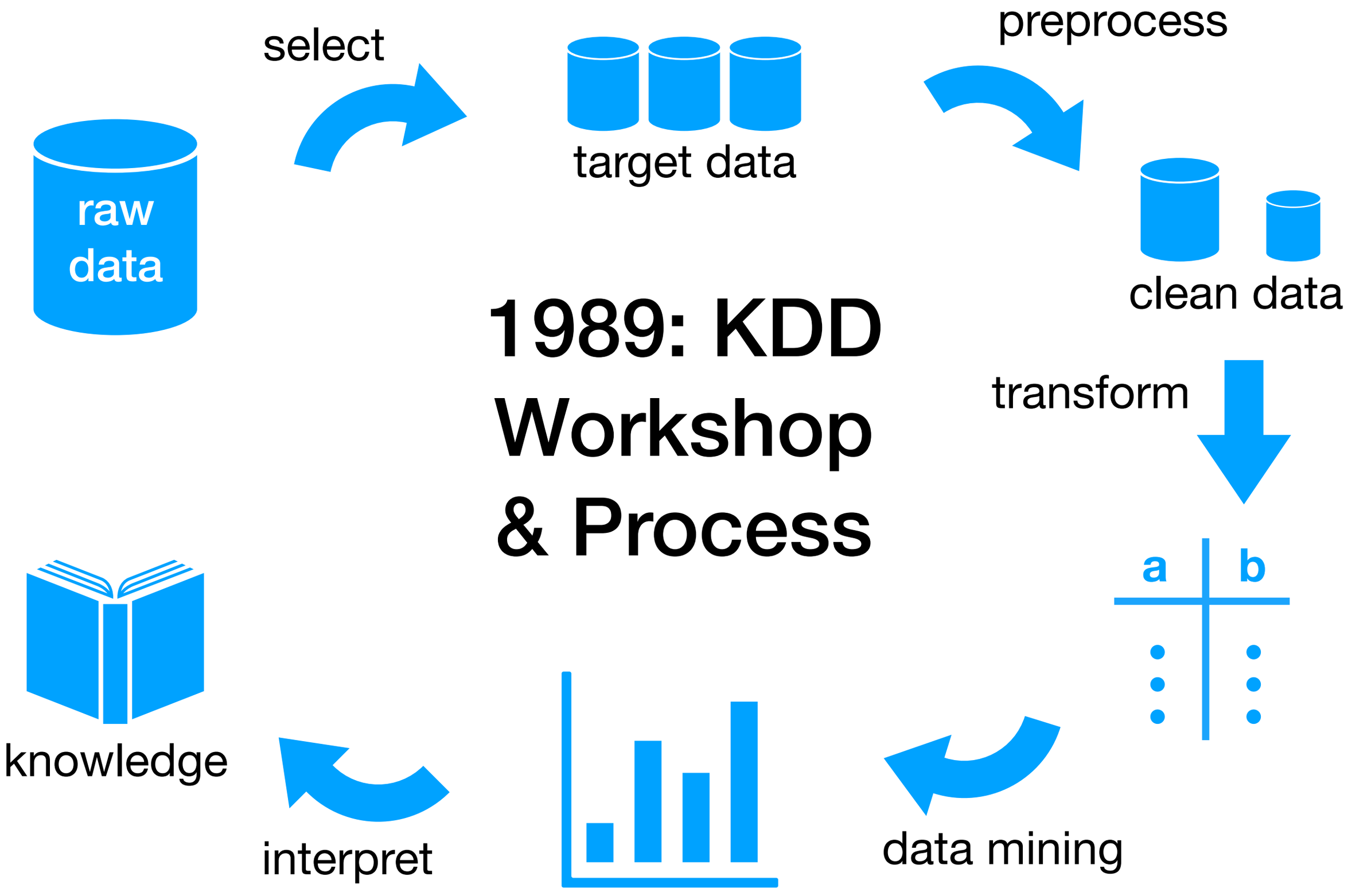
1989: KDD Workshop & Process

Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34



1989: KDD Workshop & Process

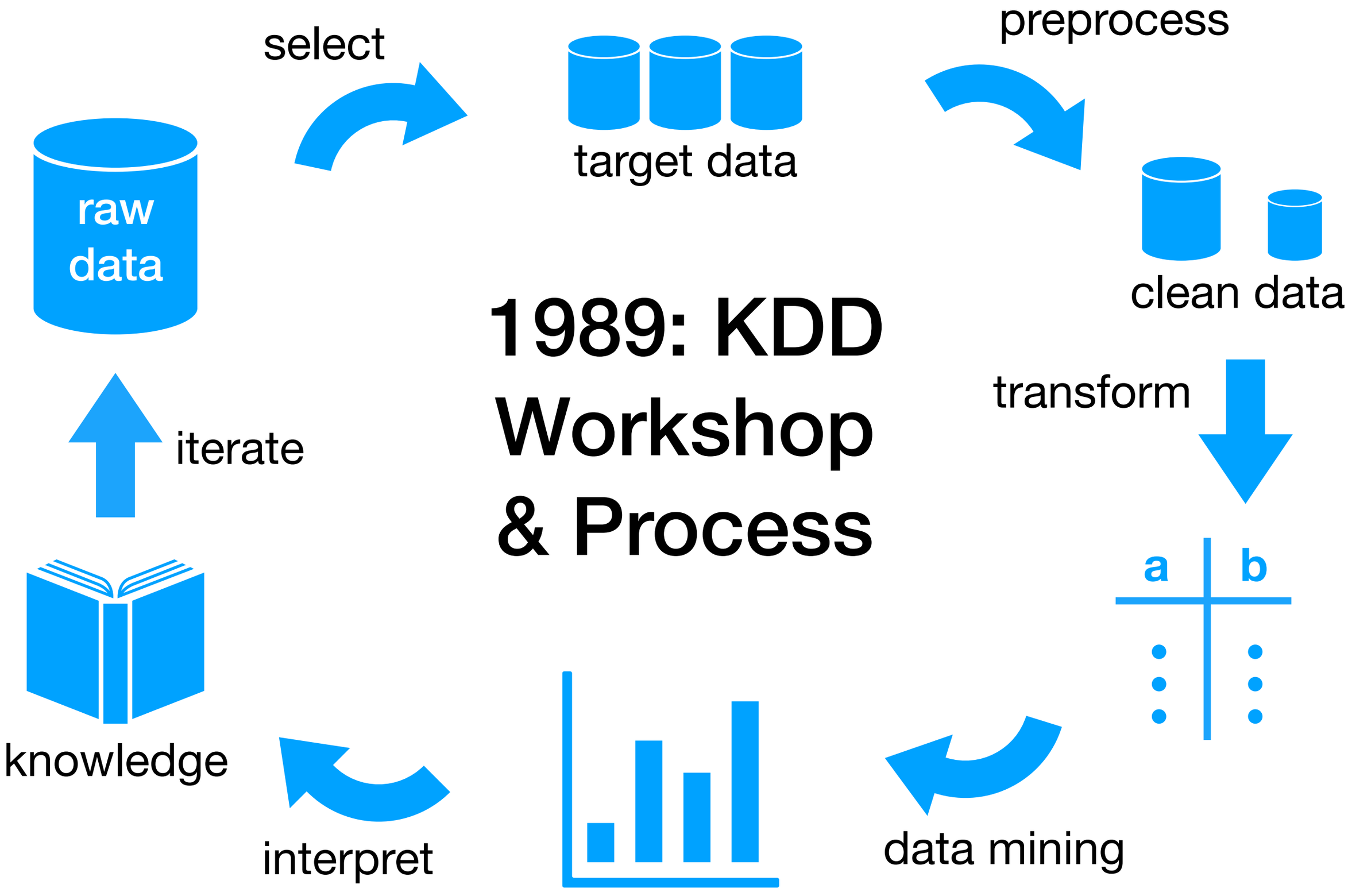
Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34



1989: KDD Workshop & Process

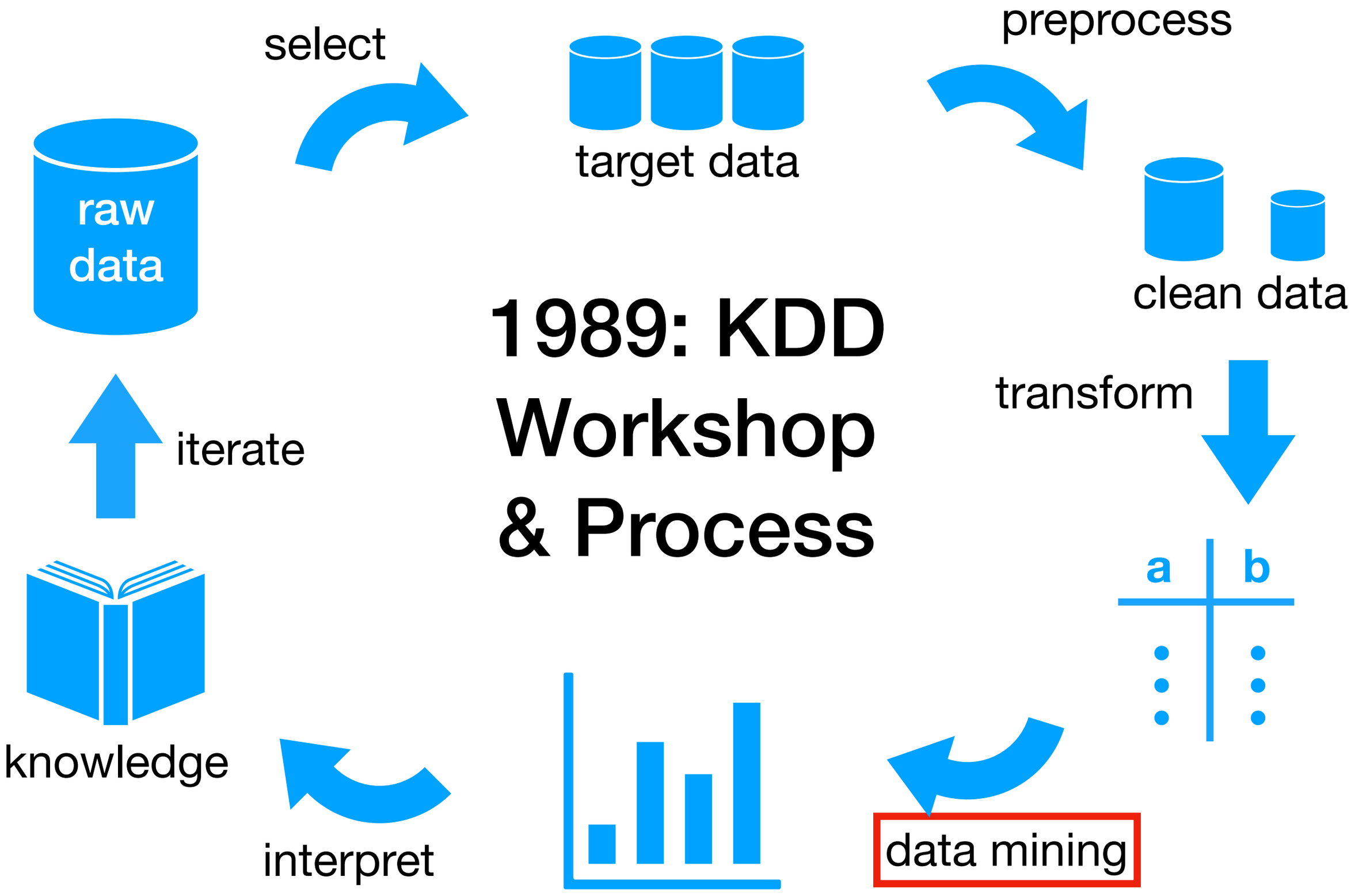
Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34

1989: KDD Workshop & Process

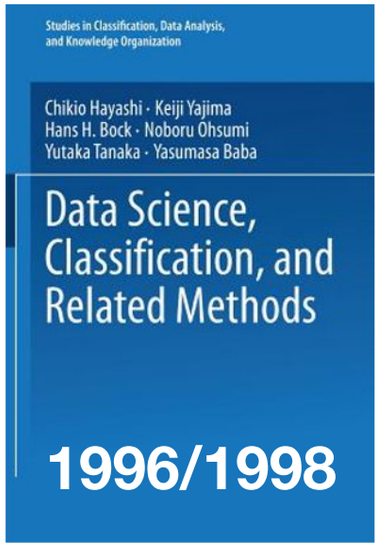


Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34

1989: KDD Workshop & Process



Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in “Advances in Knowledge Discovery and Data Mining”, 1996, pp.1-34



What is Data Science ?*
Fundamental Concepts and a Heuristic Example

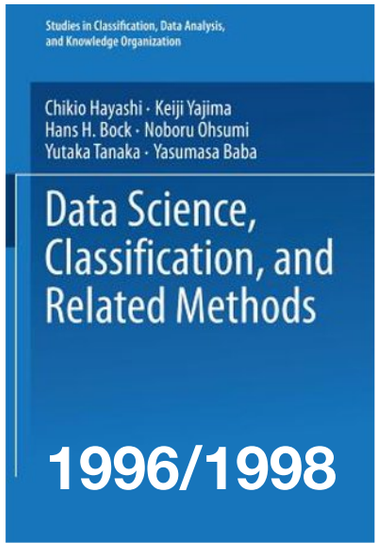
Chikio Hayashi

The Institute of Statistical Mathematics
Sakuragaoka, Birijian 304
15-8 Sakuragaoka, Shibuya-ku
Tokyo 150, Japan

Summary: Data Science is not only a synthetic concept to unify statistics, data analysis and their related methods but also comprises its results. It includes three phases, design for data, collection of data, and analysis on data. Fundamental concepts and various methods based on it are discussed with a heuristic example.

1. Introduction:

Statistics and data analysis have developed in their realms separately and contributed to the development of science, showing their unique properties. The ideas and various methods of statistics were very useful, well known and solved many problems. Mathematical statistics succeeded it and developed new frontiers with the idea of statistical inference. Thus the application of these view points brought us many useful results.
However, the development of mathematical statistics, has devoted itself only to the problems of statistical inference, an apparent rise of precision of statistical models, and to the pursuit of exactness and mathematical refinement, so mathematical statistics have been prone to be removed from reality.



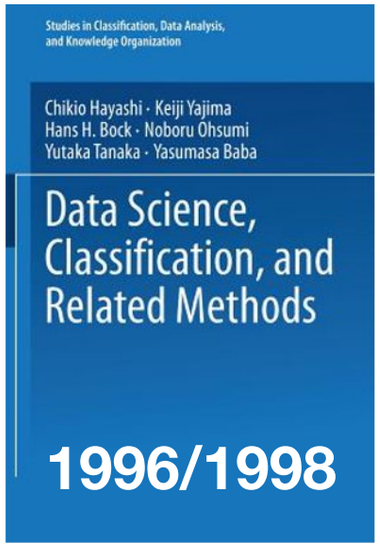
What is Data Science ?*
Fundamental Concepts and a Heuristic Example

Chikio Hayashi

The Institute of Statistical Mathematics
Sakuragaoka, Birijian 304
15-8 Sakuragaoka, Shibuya-ku
Tokyo 150, Japan

Data Science is not only a synthetic concept to **unify statistics, data analysis and their related methods** but also comprises its **results**. It includes three phases, **design** for data, **collection** of data, and **analysis** on data.

Statistics and data analysis have developed in their realms separately and contributed to the development of science, showing their unique properties. The ideas and various methods of statistics were very useful, well known and solved many problems. Mathematical statistics succeeded it and developed new frontiers with the idea of statistical inference. Thus the application of these view points brought us many useful results.
However, the development of mathematical statistics, has devoted itself only to the problems of statistical inference, an apparent rise of precision of statistical models, and to the pursuit of exactness and mathematical refinement, so mathematical statistics have been prone to be removed from reality.



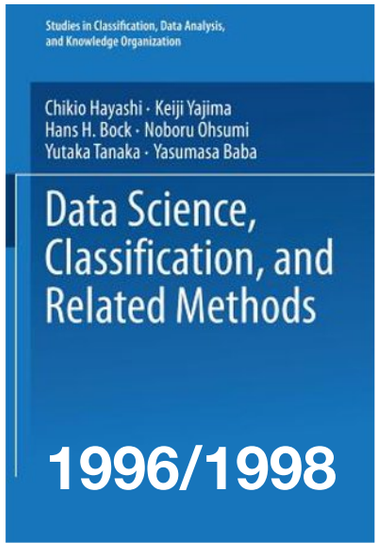
What is Data Science ?*
Fundamental Concepts and a Heuristic Example

Chikio Hayashi

The Institute of Statistical Mathematics
Sakuragaoka, Birijian 304
15-8 Sakuragaoka, Shibuya-ku
Tokyo 150, Japan

Data Science is not only a synthetic concept to **unify statistics, data analysis and their related methods** but also comprises its **results**. It includes three phases, **design** for data, **collection** of data, and **analysis** on data.

Statistics and data analysis have developed in their realms separately and contributed to the development of science, showing their unique properties. The ideas and various methods of statistics were very useful, well known and solved many problems. Mathematical statistics succeeded it and developed new frontiers with the idea of statistical inference. Thus the application of these view points brought us many useful results.
However, the development of mathematical statistics, has devoted itself only to the problems of statistical inference, an apparent rise of precision of statistical models, and to the pursuit of exactness and mathematical refinement, so mathematical statistics have been prone to be removed from reality.



What is Data Science ?*
Fundamental Concepts and a Heuristic Example

Chikio Hayashi

The Institute of Statistical Mathematics
Sakuragaoka, Birijian 304
15-8 Sakuragaoka, Shibuya-ku
Tokyo 150, Japan

Data Science is not only a synthetic concept to **unify statistics, data analysis and their related methods** but also comprises its **results**. It includes three phases, **design** for data, **collection** of data, and **analysis** on data.

Statistics and data analysis have developed in their realms separately and contributed to the development of science, showing their unique properties. The ideas and various methods of statistics were very useful, well known and solved many problems. Mathematical statistics succeeded it and developed new frontiers with the idea of statistical inference. Thus the application of these view points brought us many useful results.
However, the development of mathematical statistics, has devoted itself only to the problems of statistical inference, an apparent rise of precision of statistical models, and to the pursuit of exactness and mathematical refinement, so mathematical statistics have been prone to be removed from reality.

2010: OSEMN

Process for Data Science

Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

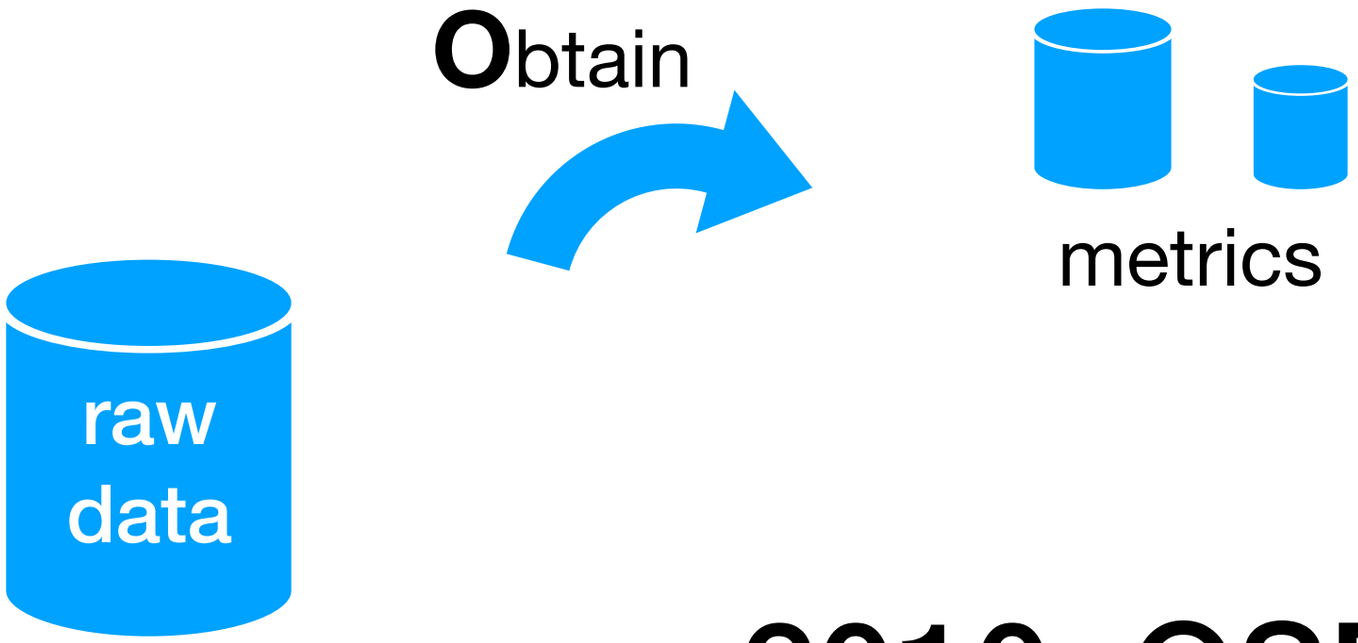
2010: OSEMN Process for Data Science

Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

2010: OSEMN Process for Data Science

knowledge

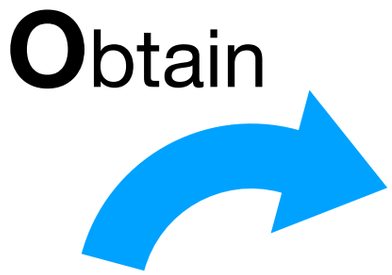
Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

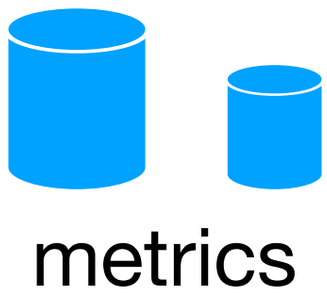


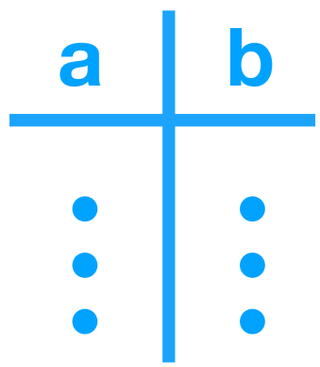
2010: OSEMN

Process for Data Science

Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/





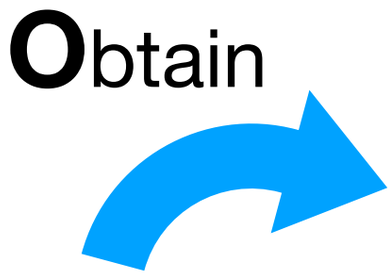


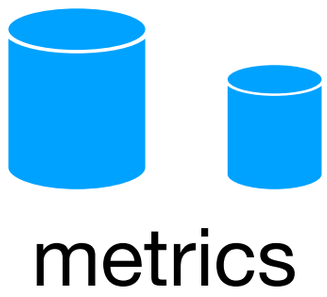
2010: OSEMN

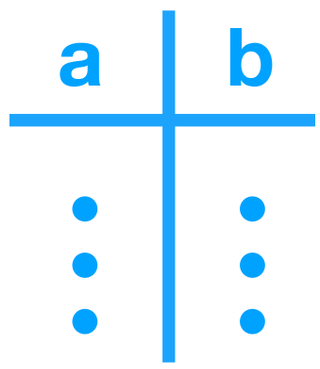
Process for Data Science

knowledge

Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/



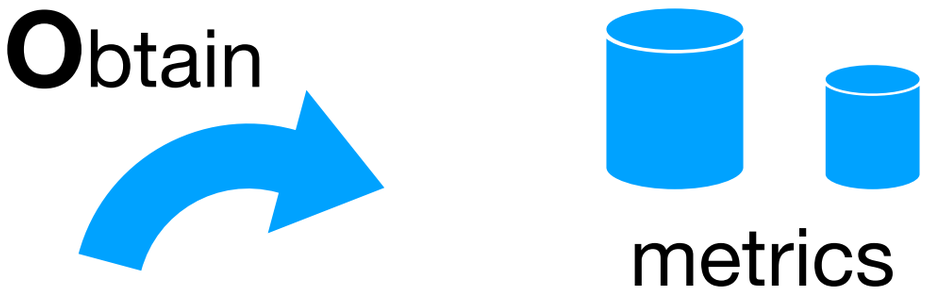


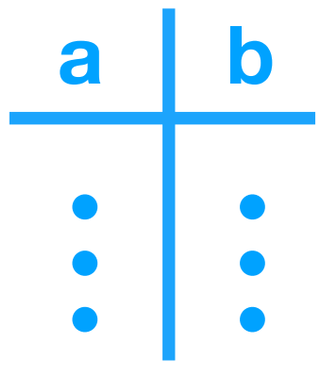


2010: OSEMN Process for Data Science

knowledge

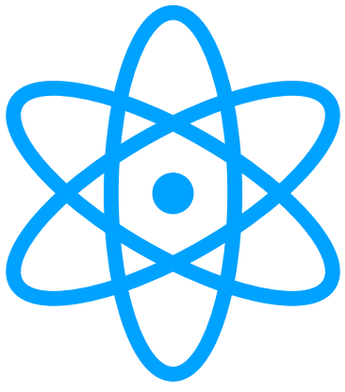
Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

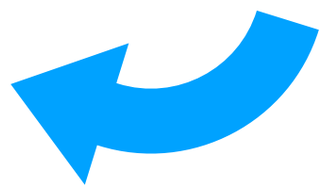




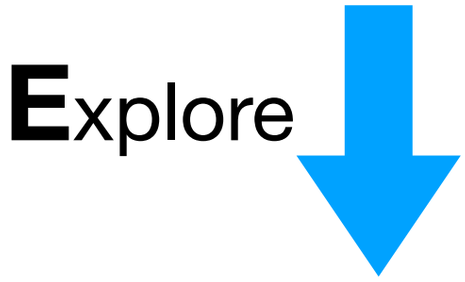
2010: OSEMN Process for Data Science

knowledge

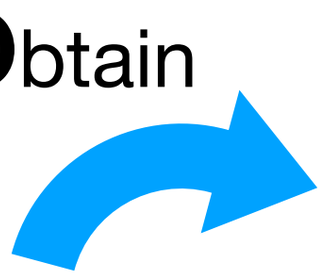


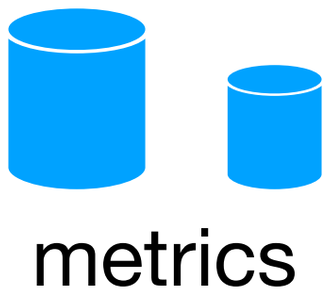


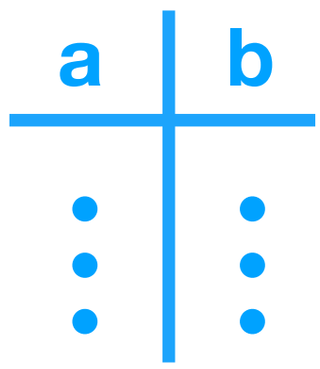
Model



Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

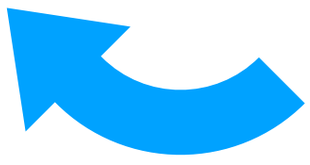




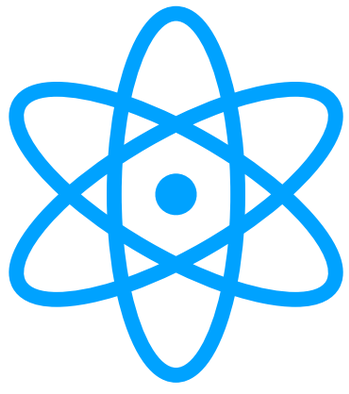


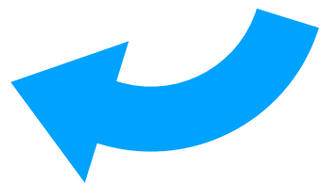
2010: OSEMN Process for Data Science

knowledge

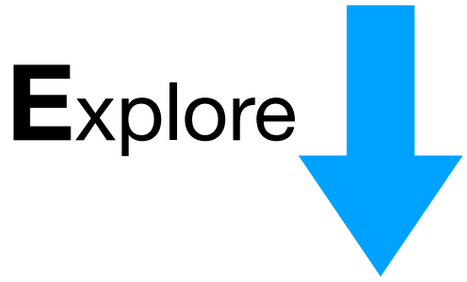


iNterpret





Model



Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

2010: OSEMN Process for Data Science

Obtain

metrics

Scrub

raw data

iterate

knowledge

iN**N**terpret

Model

Explore

a | b

2010: OSEMN Process for Data Science

Obtain

metrics

Scrub

raw data

iterate

knowledge

iN**N**terpret

Model

Explore

a | b

Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

2010: OSEMN Process for Data Science

Obtain

metrics

Scrub

raw data

iterate

knowledge

iN**N**terpret

Model

Explore

a | b

Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

Obtain

metrics

Scrub

raw data

iterate

Explore

knowledge

i**N**terpret

Model

Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

Obtain

metrics

Scrub

raw data

Scrapy

pandas

scikits learn

HDFS

a | b

iterate

WEKA The University of Waikato

Explore

K

theano

Perl

APACHE Spark™

hadoop

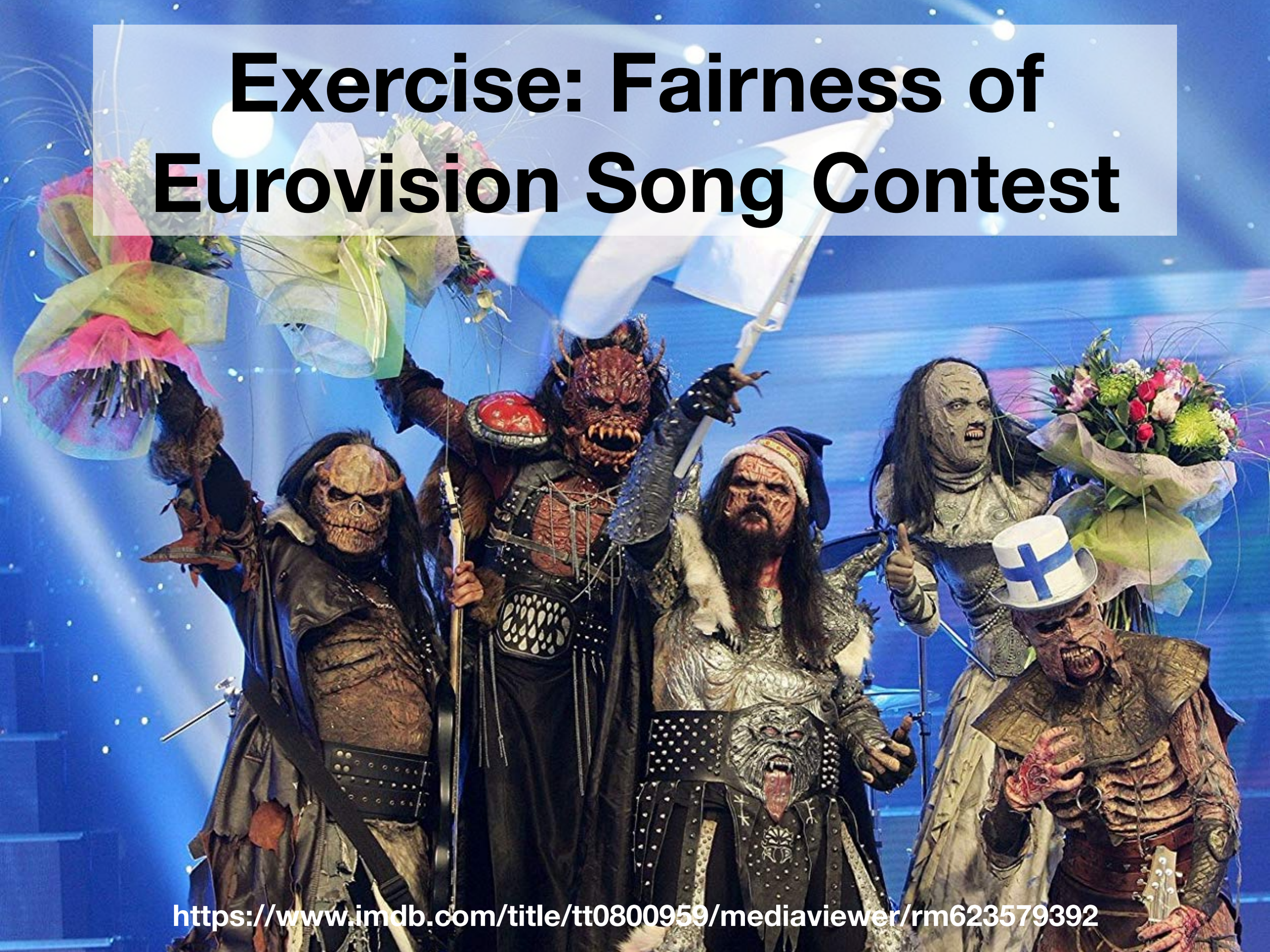
knowledge

iN**t**erpret

Model

Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

Exercise: Fairness of Eurovision Song Contest



https://www.imdb.com/title/tt0800959/mediaviewer/rm623579392

2010: OSEMN Process for Data Science

Obtain

metrics

Scrub

raw data

iterate

Explore

knowledge

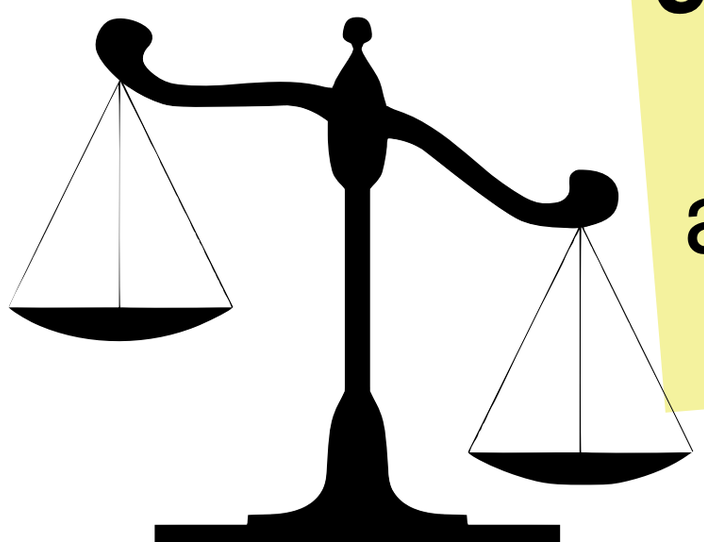
iN**N**terpret

Model

Hilary Mason & Chris Wiggins, http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

Why (Not) Data Science?

Why (Not) Data Science?

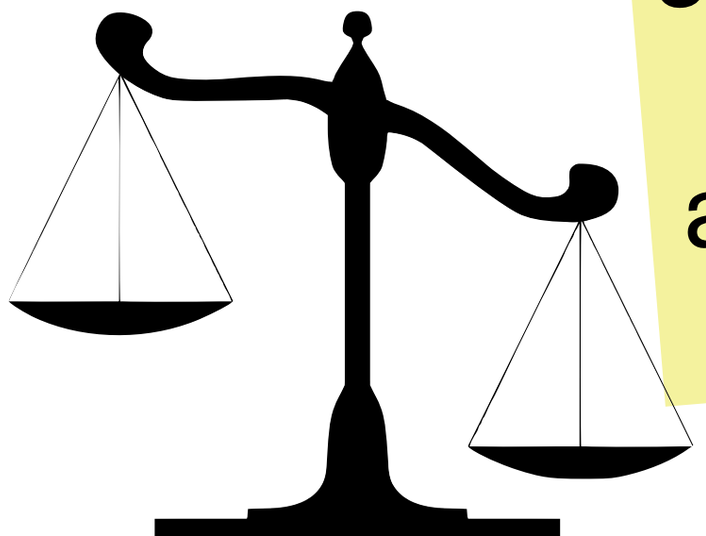


customize software tools and data analysis techniques to problem at hand

Why (Not) Data Science?

exploit explosion in computing power (cloud!) to **scale** with volume and velocity of available data sources

customize software tools and data analysis techniques to problem at hand



Why (Not) Data Science?

lacking **guidance** on the planning/problem formulation side (due to focus on exploratory/ bottom-up research)

exploit explosion in computing power (cloud!) to **scale** with volume and velocity of available data sources

customize software tools and data analysis techniques to problem at hand

Part II: Empirical SE

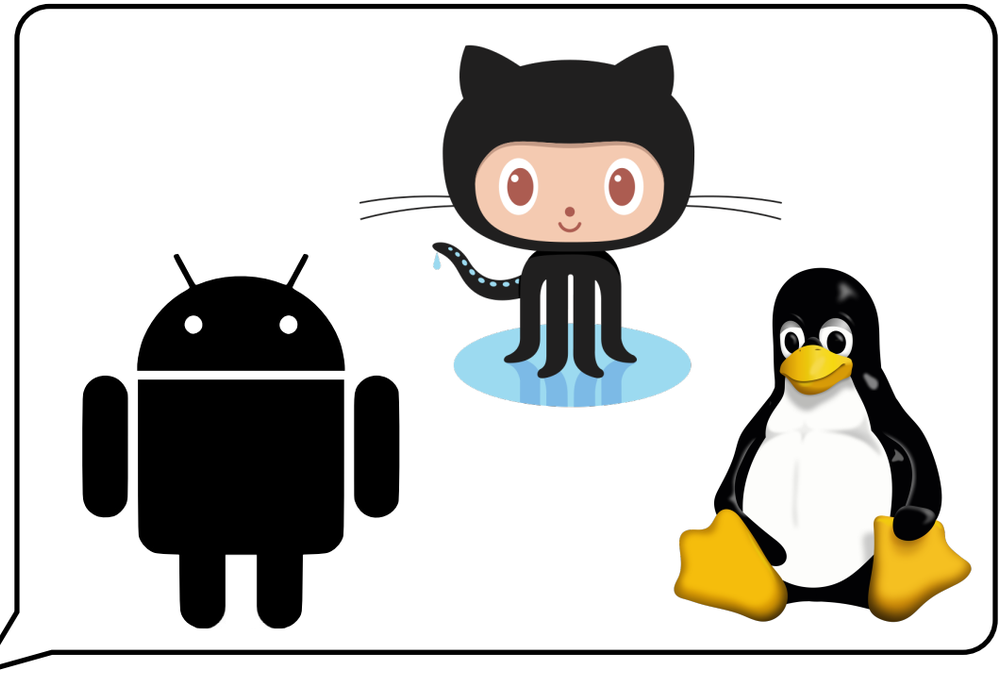
Where does “Raw Data” Come from?

Where does “Raw Data” Come from?

any data domain

Where does “Raw Data” Come from?

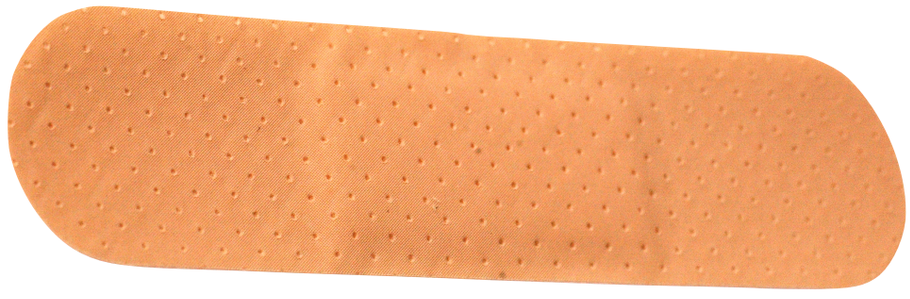
any data domain



SE Development Process

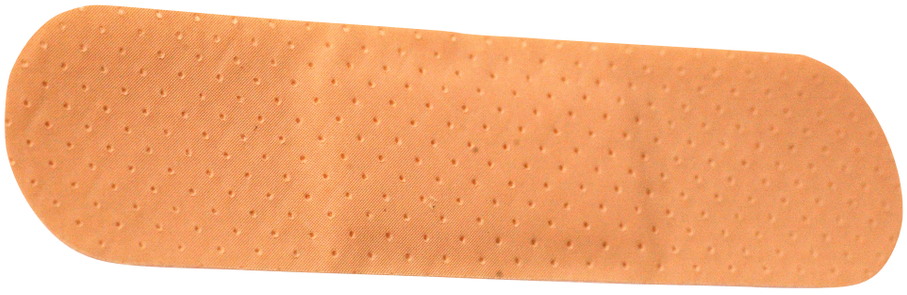
… and How does it Relate to High-Level Questions?

… and How does it Relate to High-Level Questions?

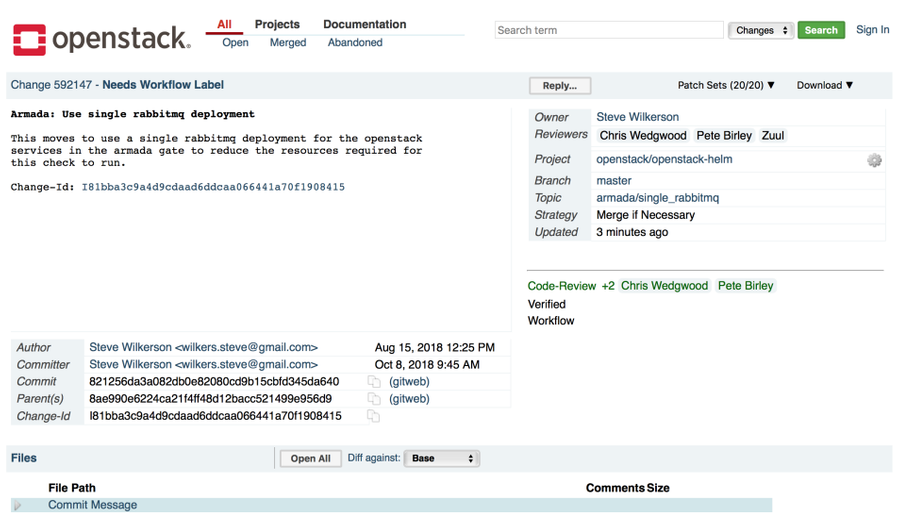


What kinds of patches are more likely to be accepted by OSS projects?

… and How does it Relate to High-Level Questions?

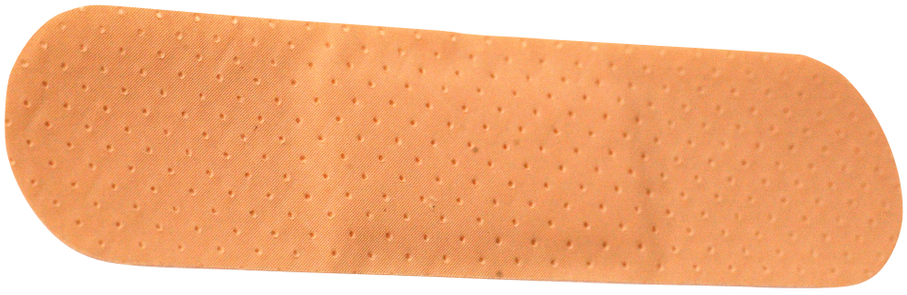


What kinds of patches are more likely to be accepted by OSS projects?

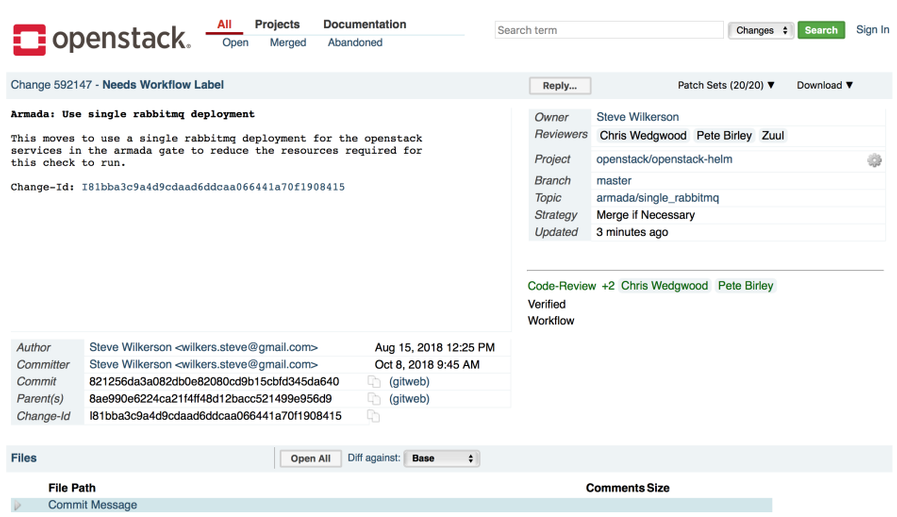


Does our new dashboard for code review improve reviewers' effectiveness?

… and How does it Relate to High-Level Questions?



What kinds of patches are more likely to be accepted by OSS projects?



Does our new dashboard for code review improve reviewers' effectiveness?

`Commit.`

Why do developers prefer short commit messages?

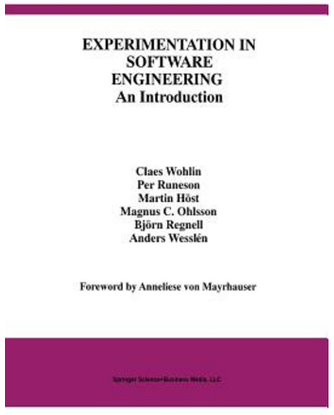
These Questions All Require Empirical Evidence

- Why? SE development process is managed by **humans**, hence **no** formal laws, rules, etc. exist that "**prove**" the answer to these questions

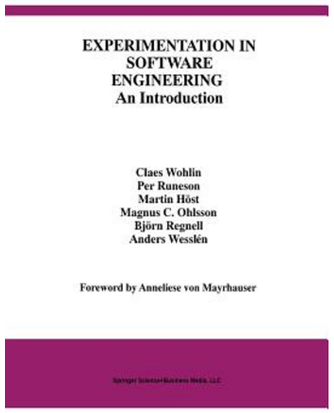
- How? through "a posteriori" knowledge "**observed**" or "**experienced**" through:

  - quantitative analysis: **how large** is effect of given manipulation or activity?

  - qualitative analysis: **why** does something happen, from perspective/point of view of subjects?

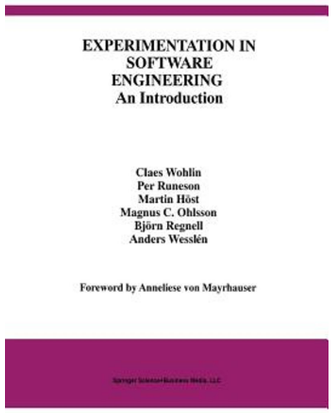


Common Strategies to Obtain Empirical Evidence



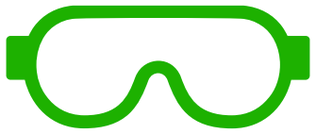
Common Strategies to Obtain Empirical Evidence

case study (observational study, no control)

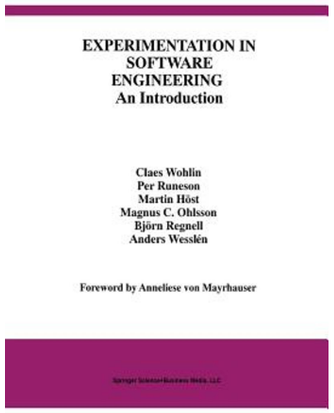


Common Strategies to Obtain Empirical Evidence

case study (observational study, no control)

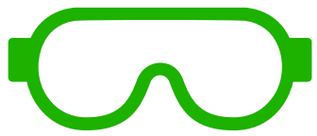


experiment (assigning treatments, with control)

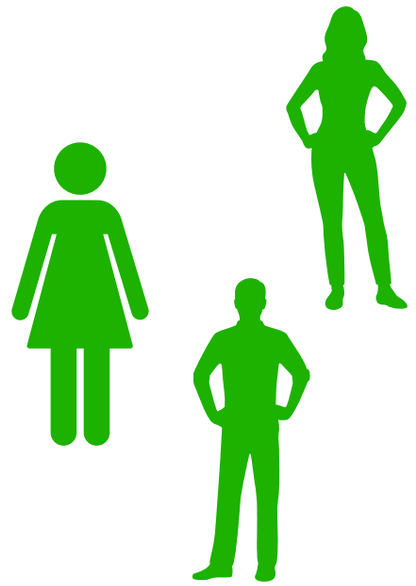


Common Strategies to Obtain Empirical Evidence

case study (observational study, no control)

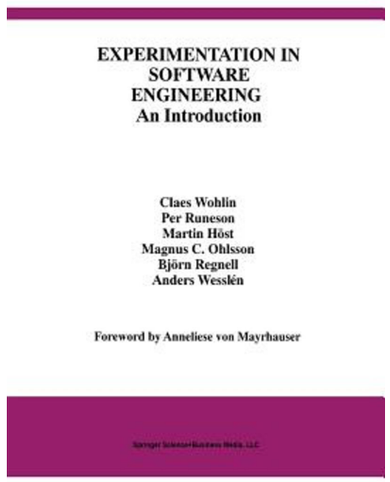


experiment (assigning treatments, with control)



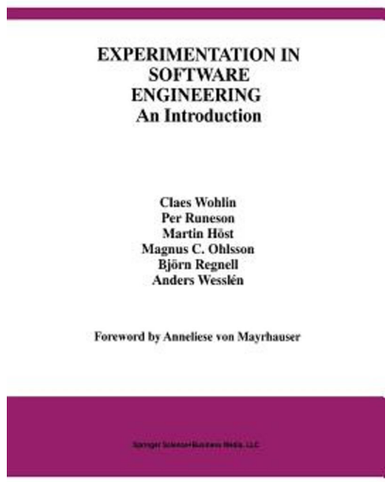
survey (interviews/ questionnaire)

Traditional Empirical SE Process



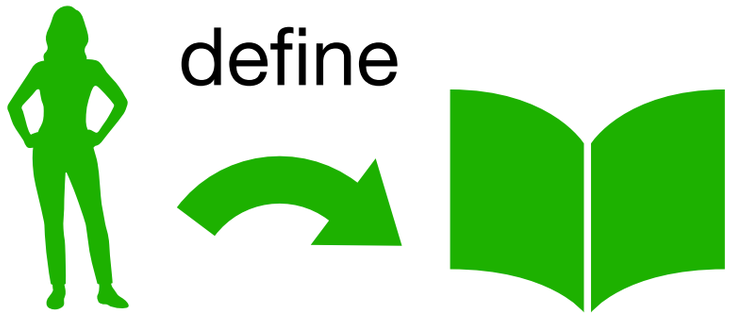
Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

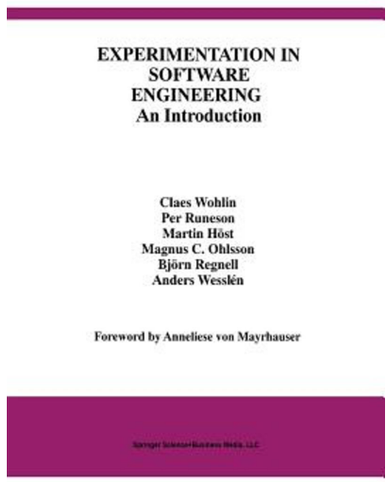
Traditional Empirical SE Process



Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

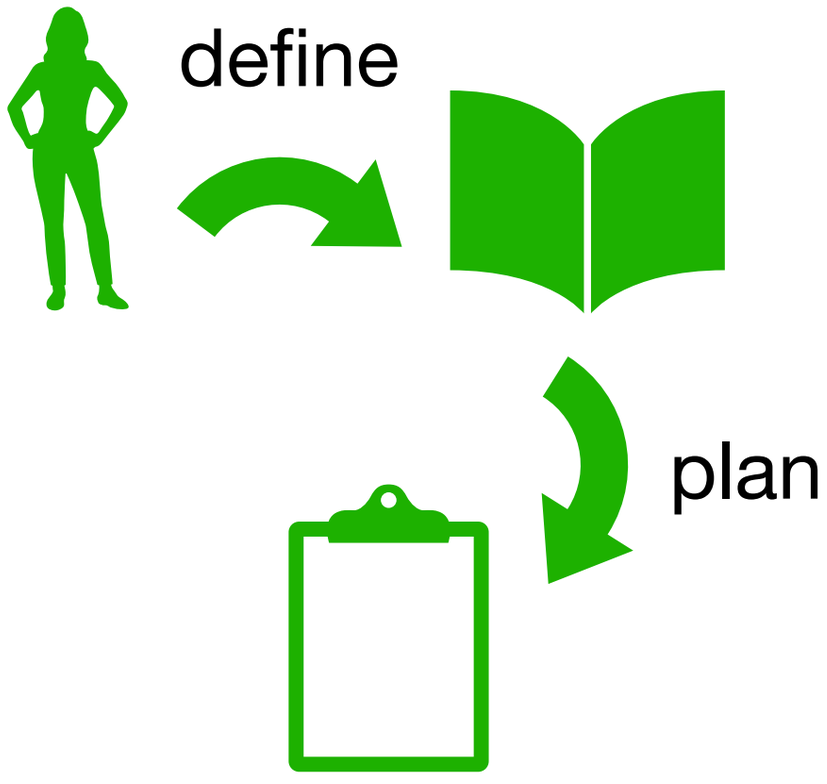
Traditional Empirical SE Process

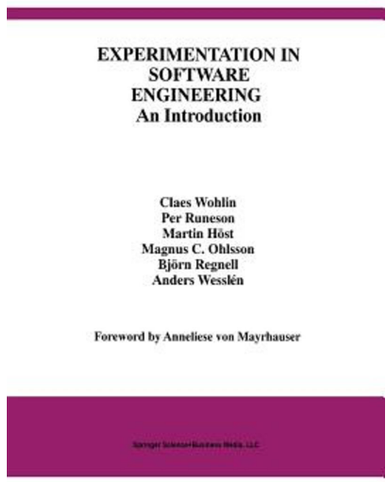




Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

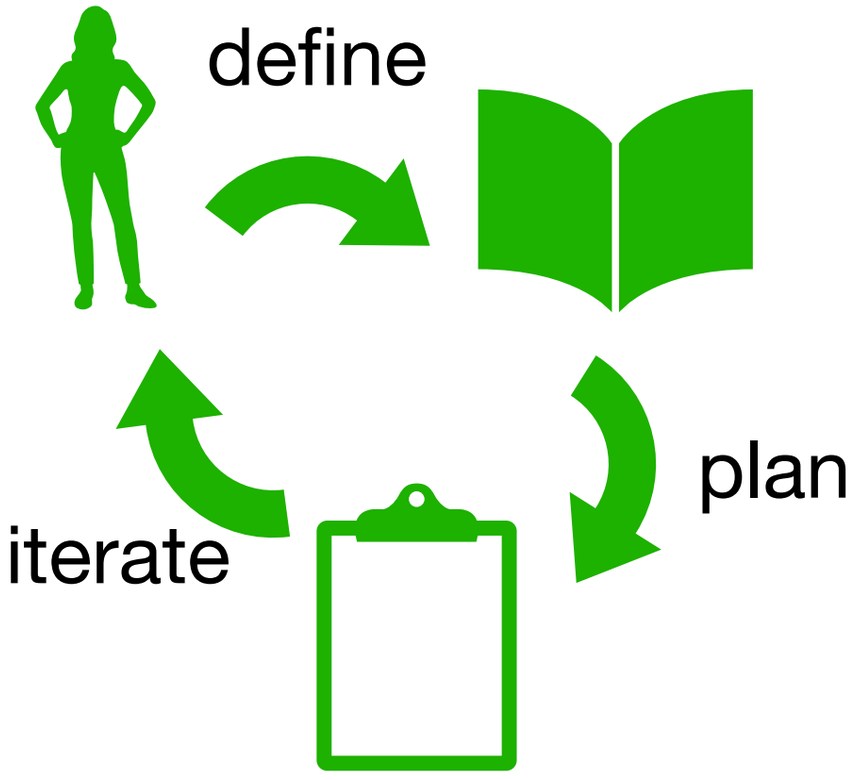
Traditional Empirical SE Process

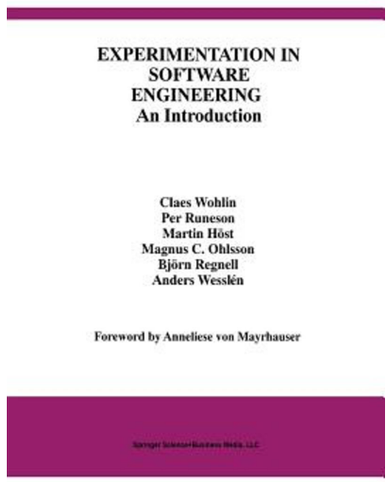




Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

Traditional Empirical SE Process



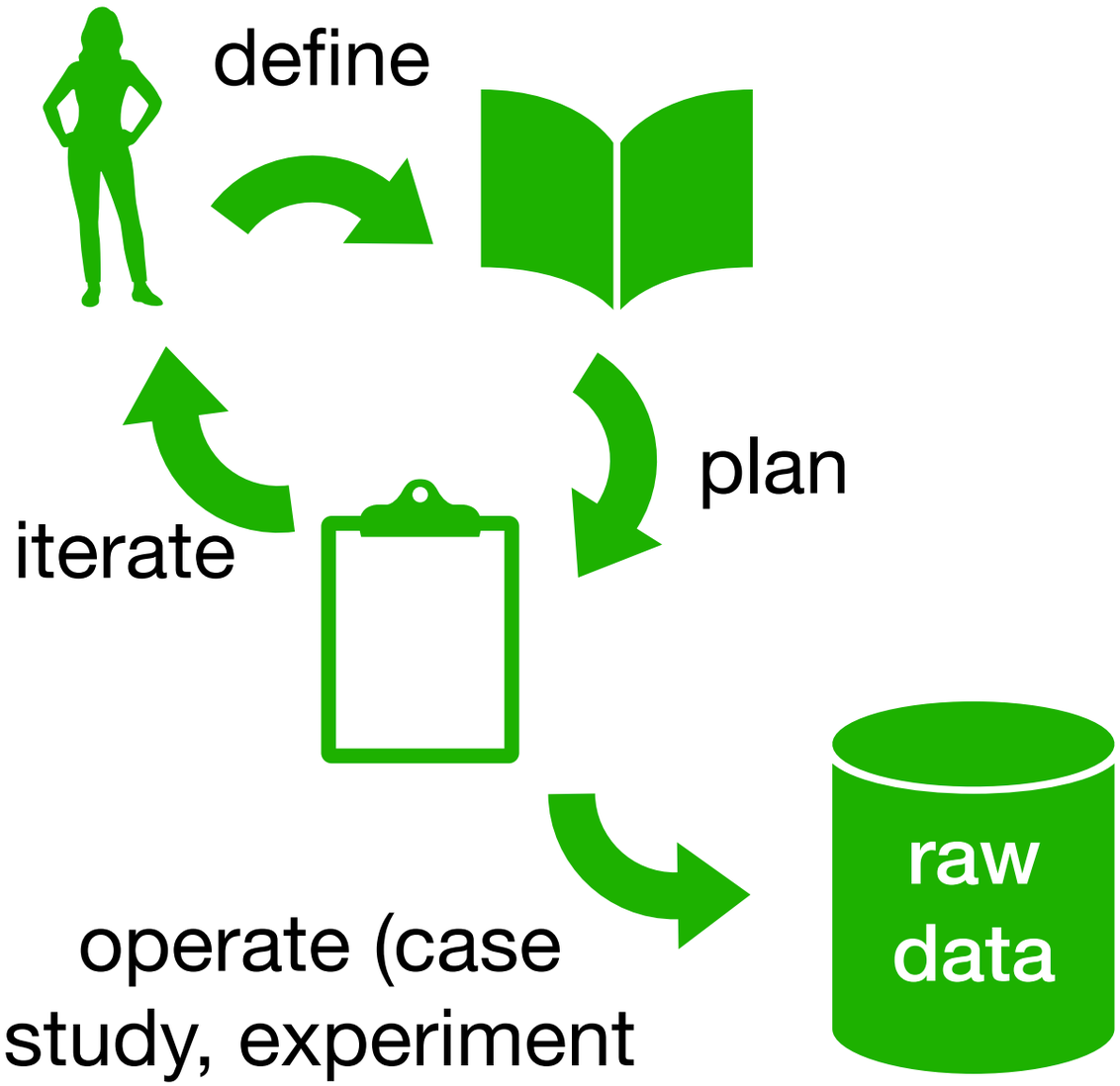

EXPERIMENTATION IN SOFTWARE ENGINEERING
An Introduction

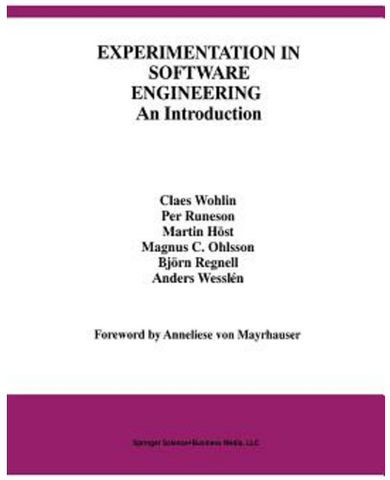
Claes Wohlin
Per Runeson
Martin Höst
Magnus C. Ohlsson
Björn Regnell
Anders Wesslén

Foreword by Anneliese von Mayrhauser

Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

Traditional Empirical SE Process



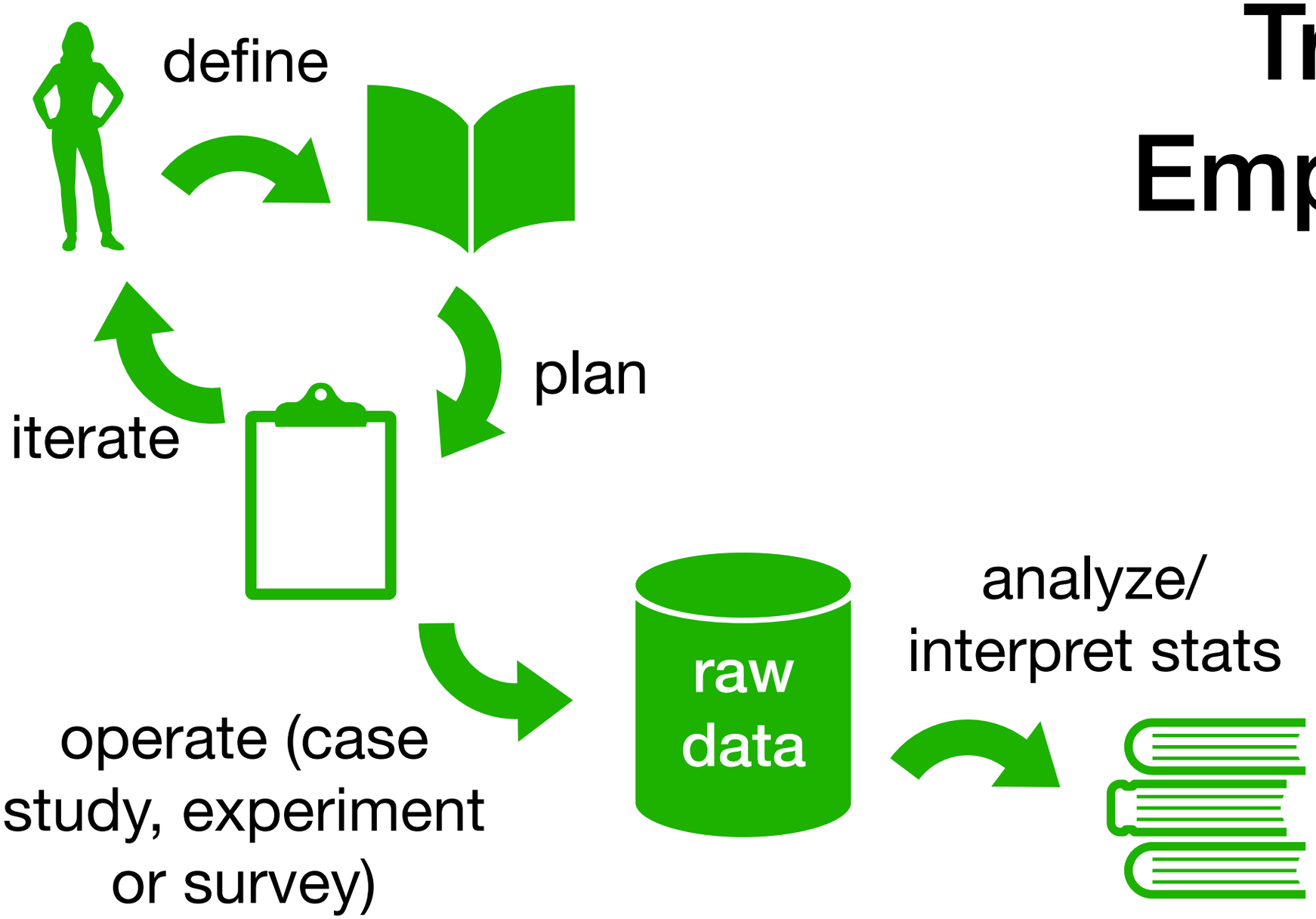

EXPERIMENTATION IN SOFTWARE ENGINEERING
An Introduction

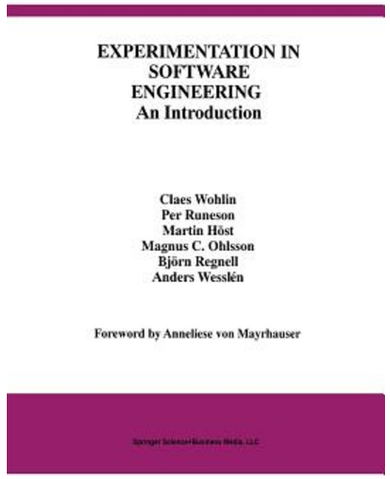
Claes Wohlin
Per Runeson
Martin Höst
Magnus C. Ohlsson
Björn Regnell
Anders Wesslén

Foreword by Anneliese von Mayrhauser

Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

Traditional Empirical SE Process



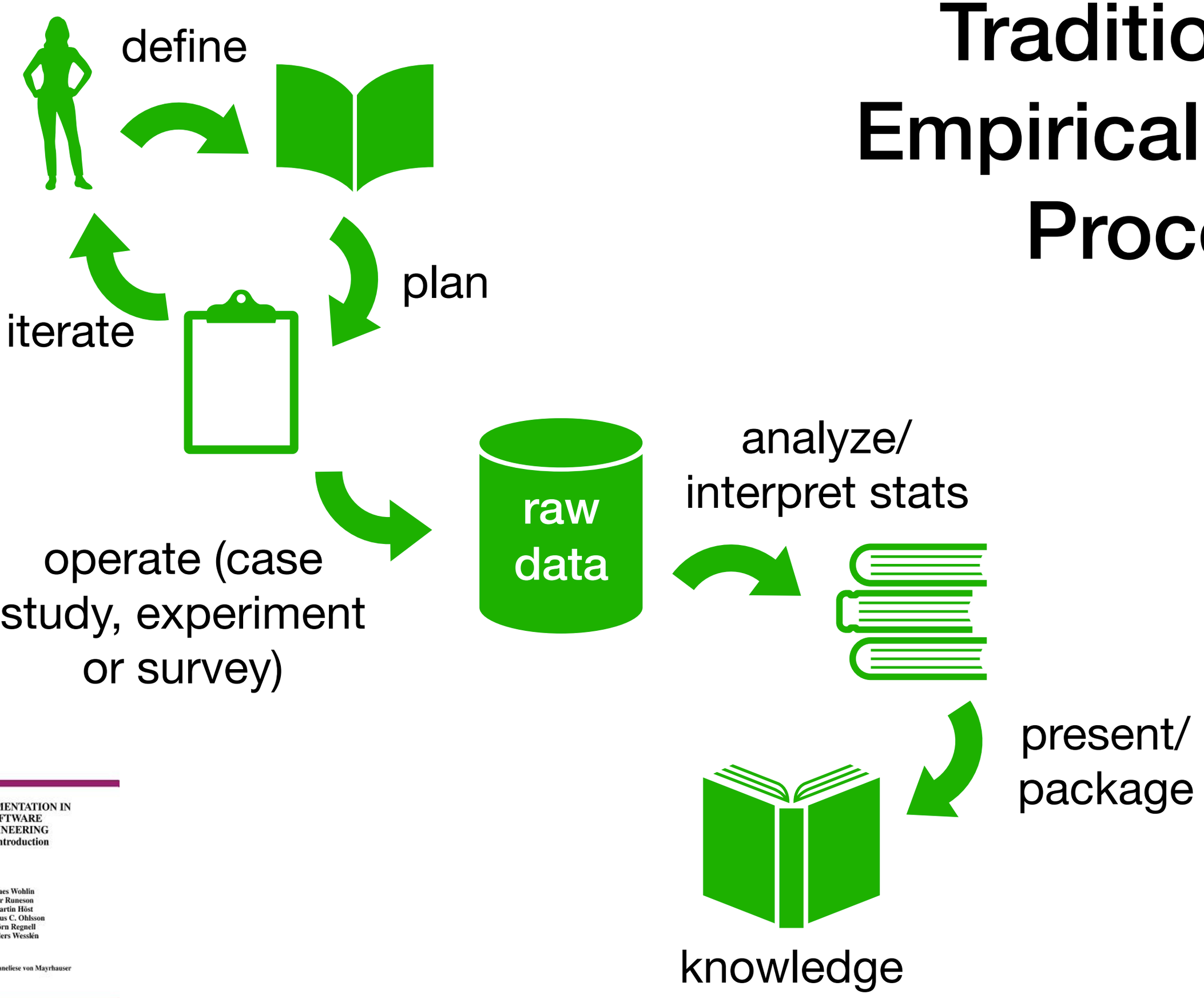

EXPERIMENTATION IN SOFTWARE ENGINEERING
An Introduction

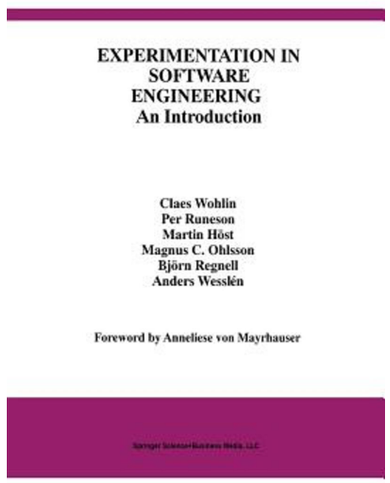
Claes Wohlin
Per Runeson
Martin Höst
Magnus C. Ohlsson
Björn Regnell
Anders Wesslén

Foreword by Anneliese von Mayrhauser

Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

Traditional Empirical SE Process



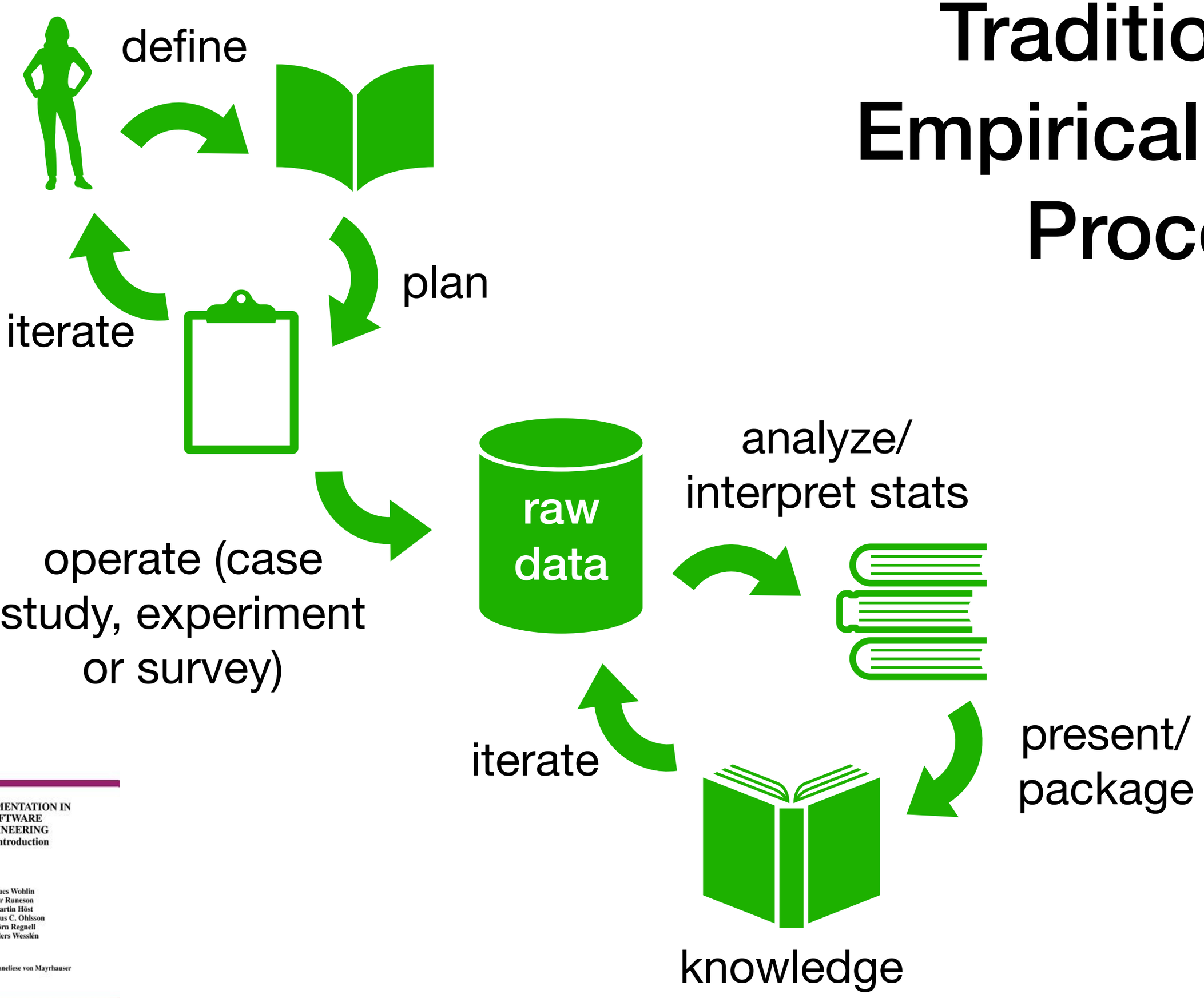

EXPERIMENTATION IN SOFTWARE ENGINEERING
An Introduction

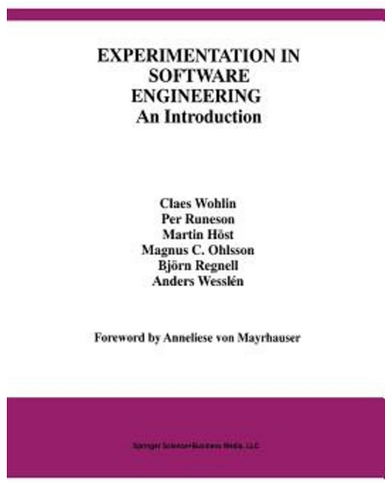
Claes Wohlin
Per Runeson
Martin Höst
Magnus C. Ohlsson
Björn Regnell
Anders Wesslén

Foreword by Anneliese von Mayrhauser

Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

Traditional Empirical SE Process




EXPERIMENTATION IN SOFTWARE ENGINEERING
An Introduction

Claes Wohlin
Per Runeson
Martin Höst
Magnus C. Ohlsson
Björn Regnell
Anders Wesslén

Foreword by Anneliese von Mayrhauser

Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

Traditional Empirical SE Process

define

plan

iterate

operate (case study, experiment or survey)

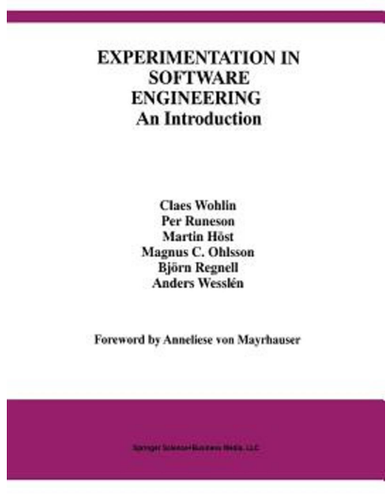
raw data

analyze/ interpret stats

present/ package

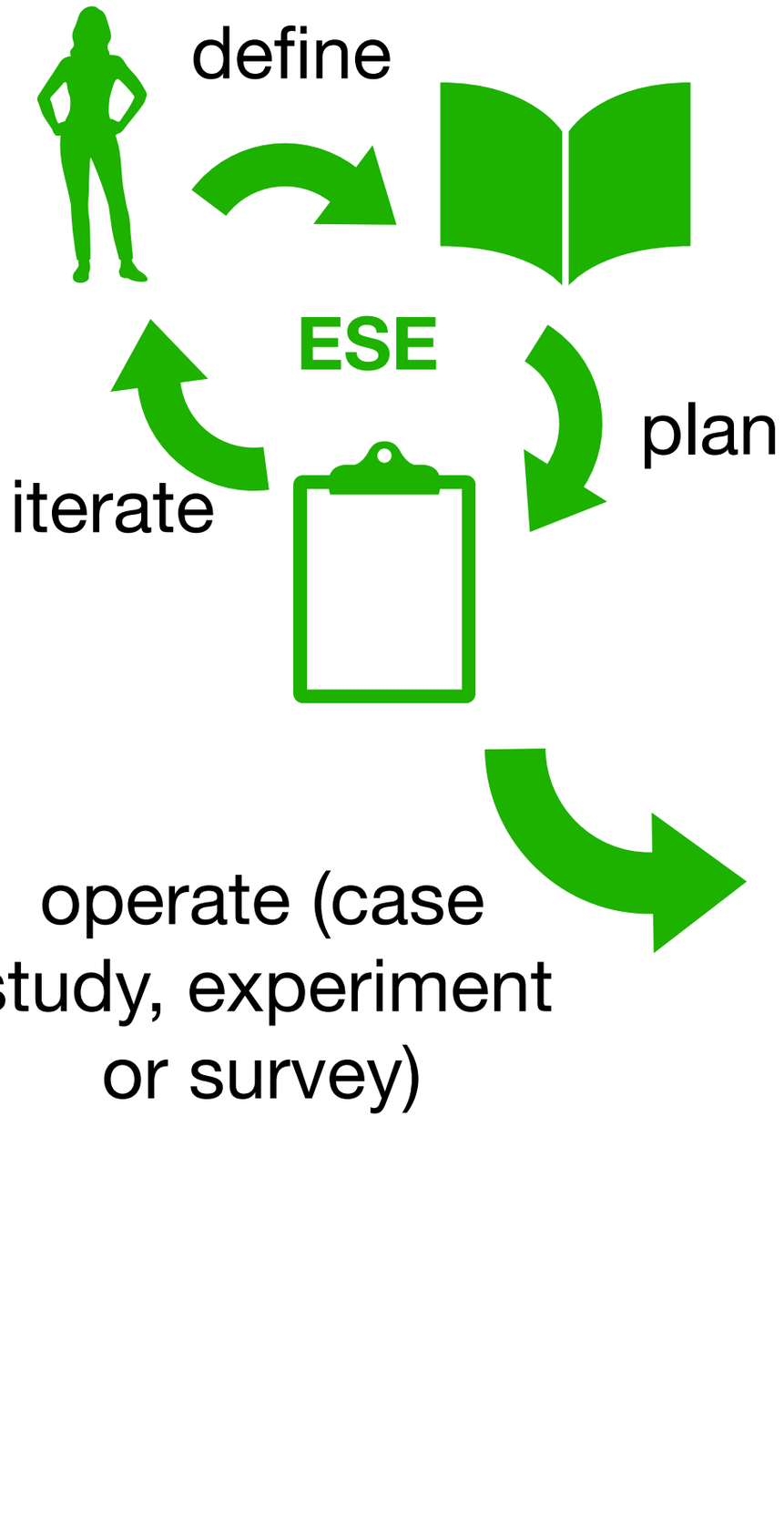
iterate

knowledge

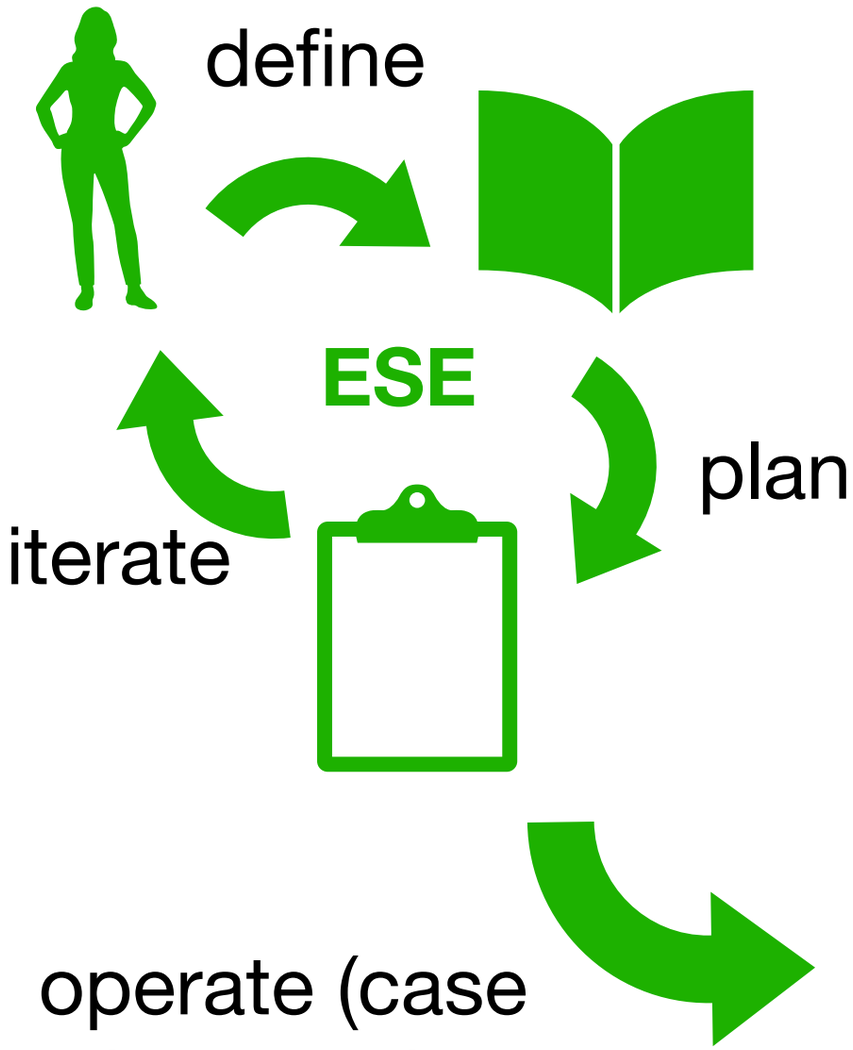


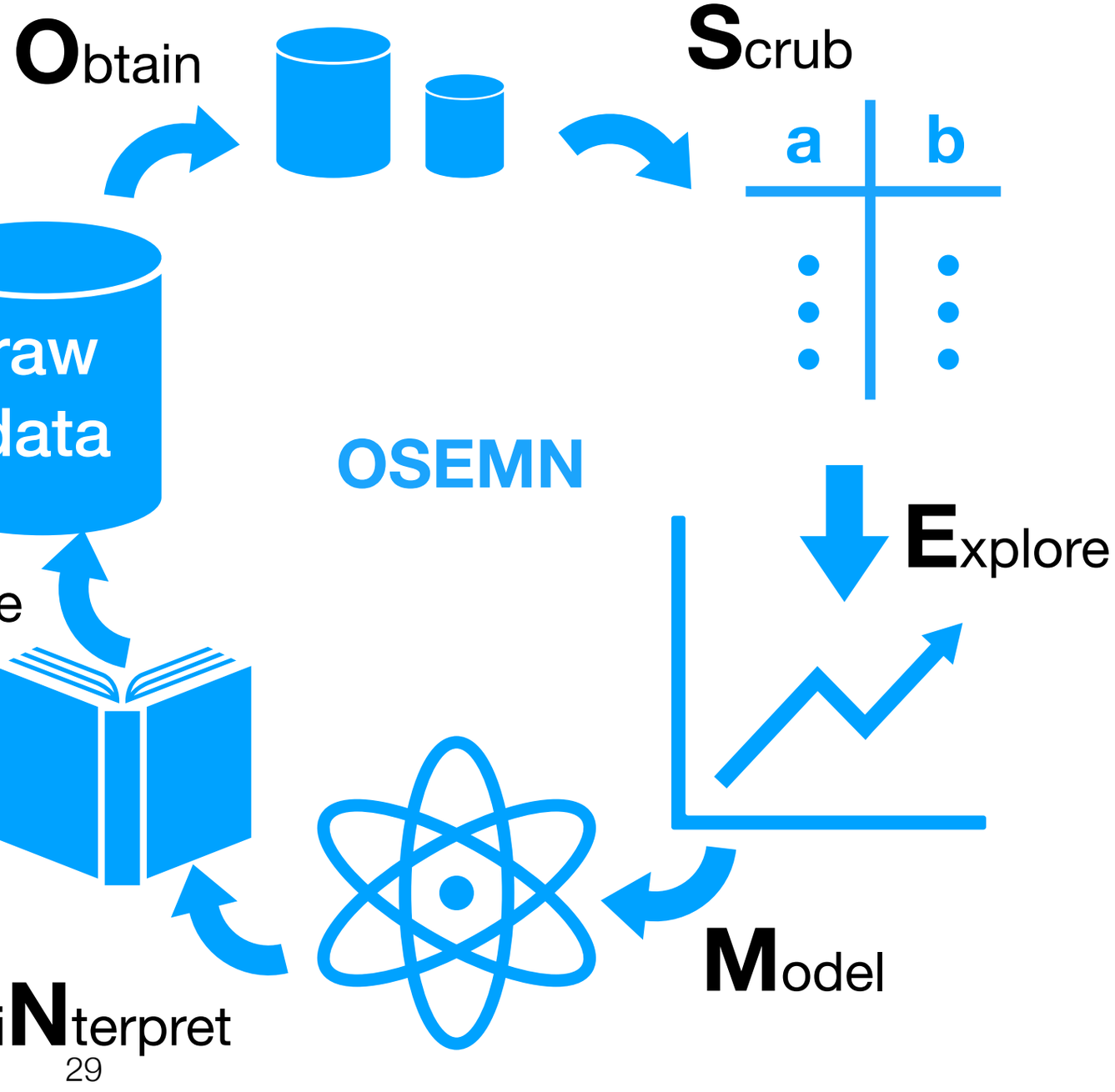
Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

Enhanced Empirical SE Process



Enhanced Empirical SE Process

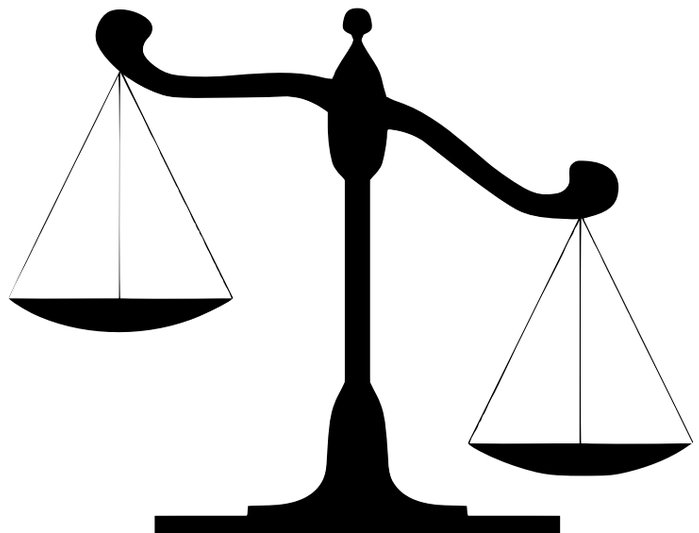




empirical SE

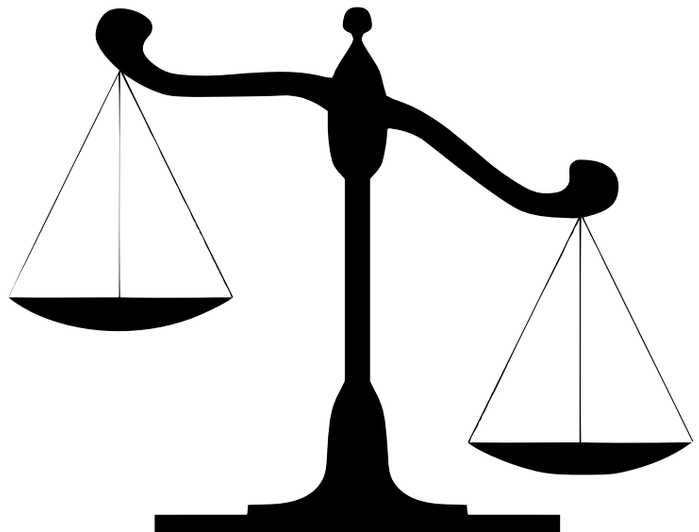
data science

Why this Hybrid Process?



Why this Hybrid Process?

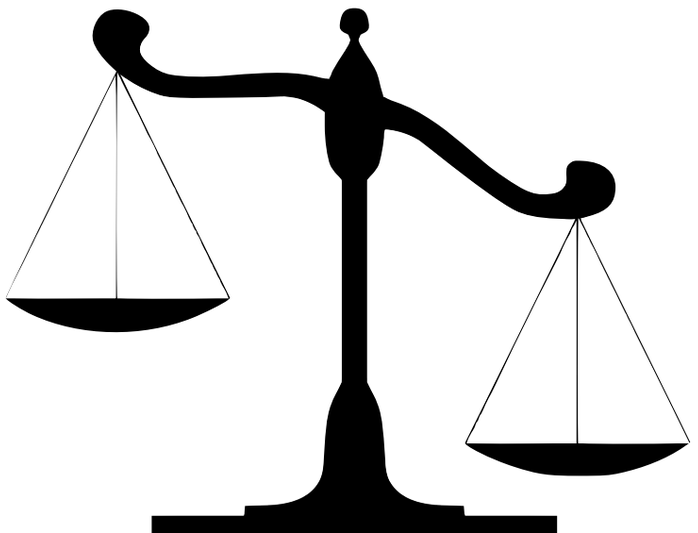
strong **empirical** definition/planning/ operation **base**



Why this Hybrid Process?

strong **empirical** definition/planning/ operation **base**

exploit explosion in computing power (cloud!) to **scale** with volume and velocity of available data sources



Why this Hybrid Process?

strong **empirical** definition/planning/ operation **base**

exploit explosion in computing power (cloud!) to **scale** with volume and velocity of available data sources

customize software tools and data analysis techniques to problem at hand

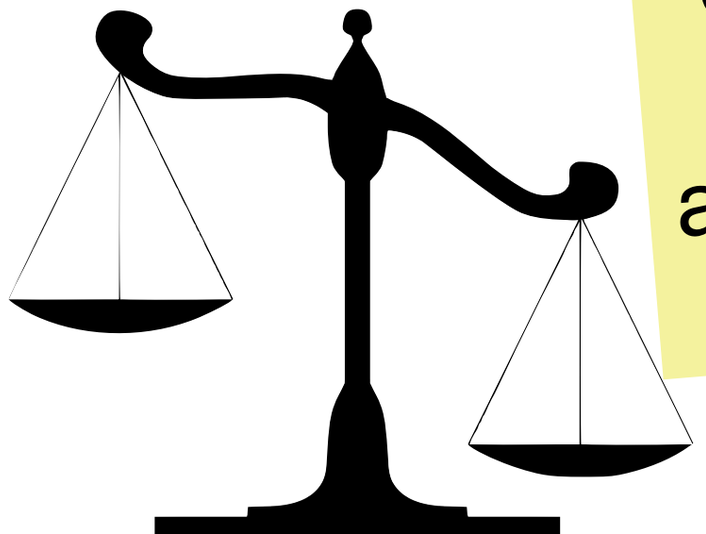
Why this Hybrid Process?

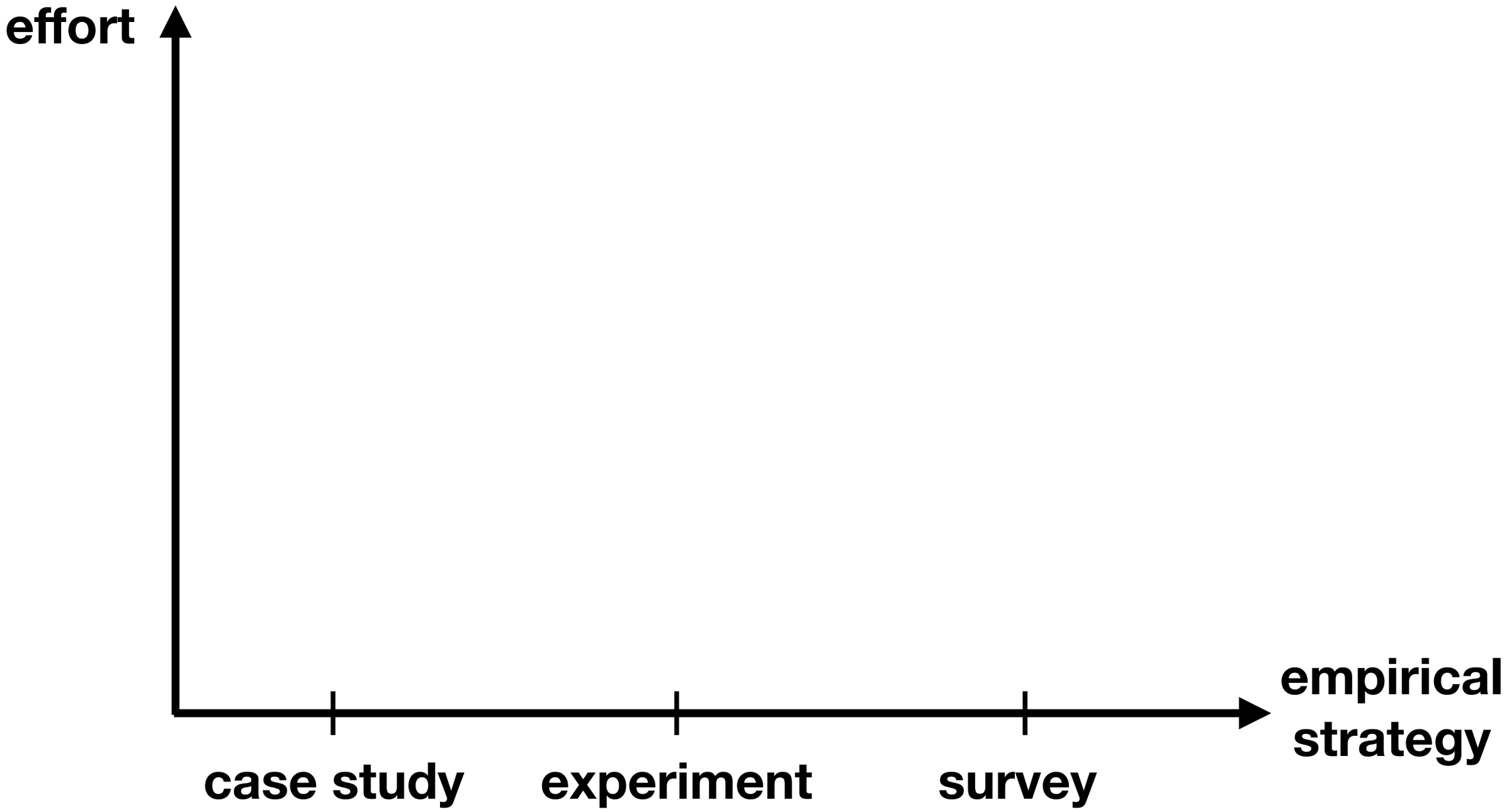
strong **empirical** definition/planning/ operation **base**

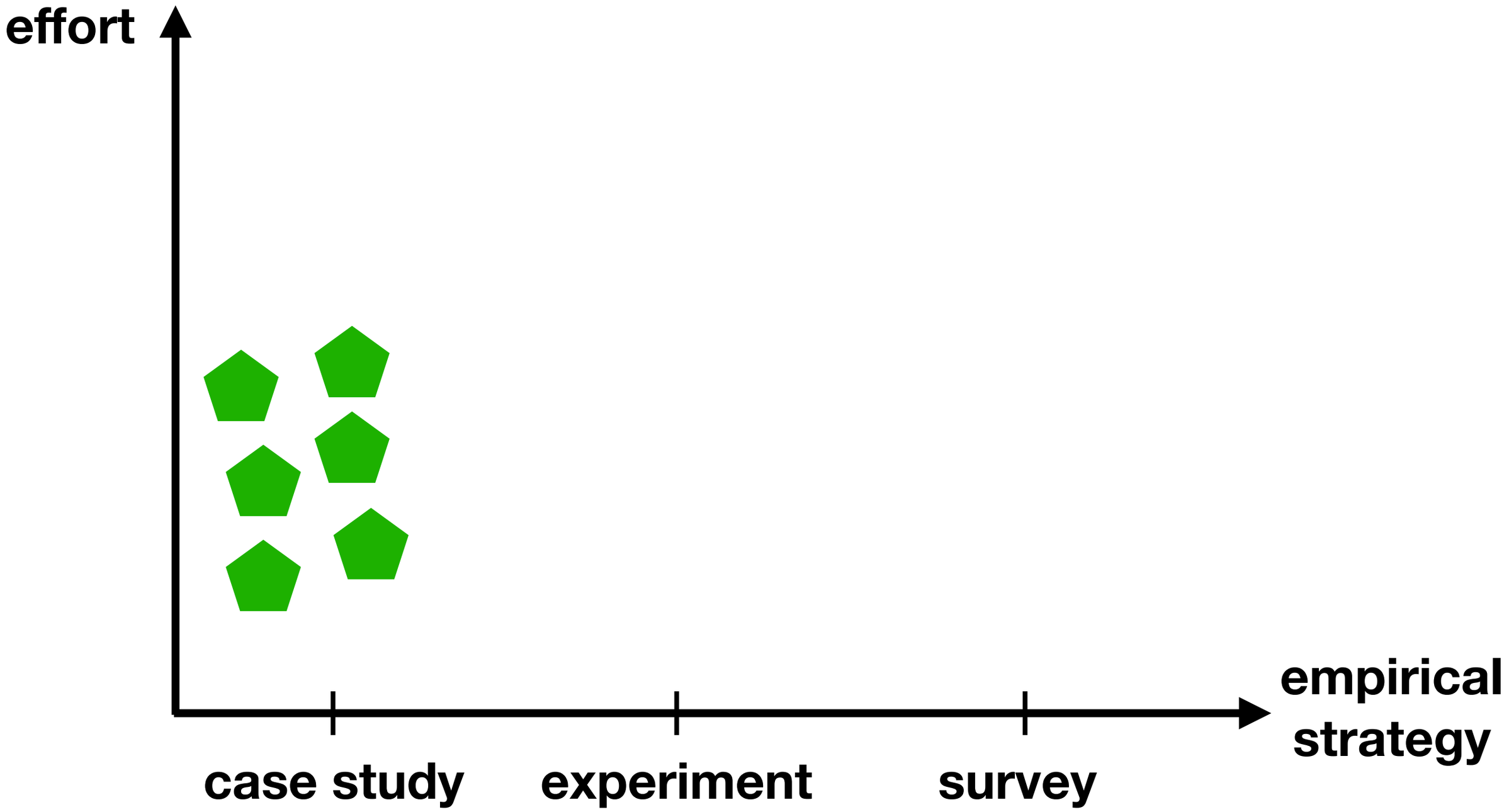
exploit explosion in computing power (cloud!) to **scale** with volume and velocity of available data sources

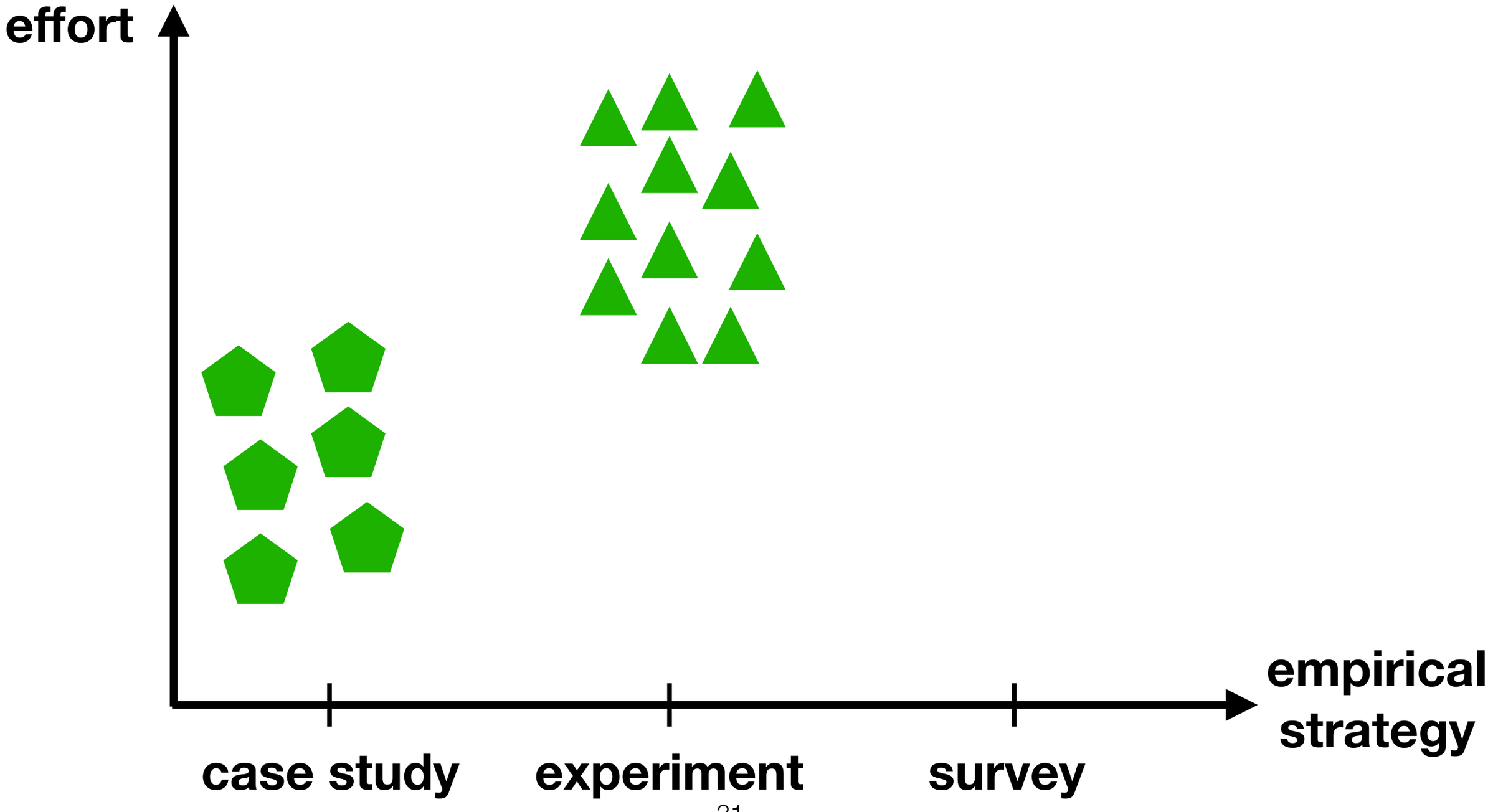
experiments, surveys and even case studies require large **effort**, depending on available data sources

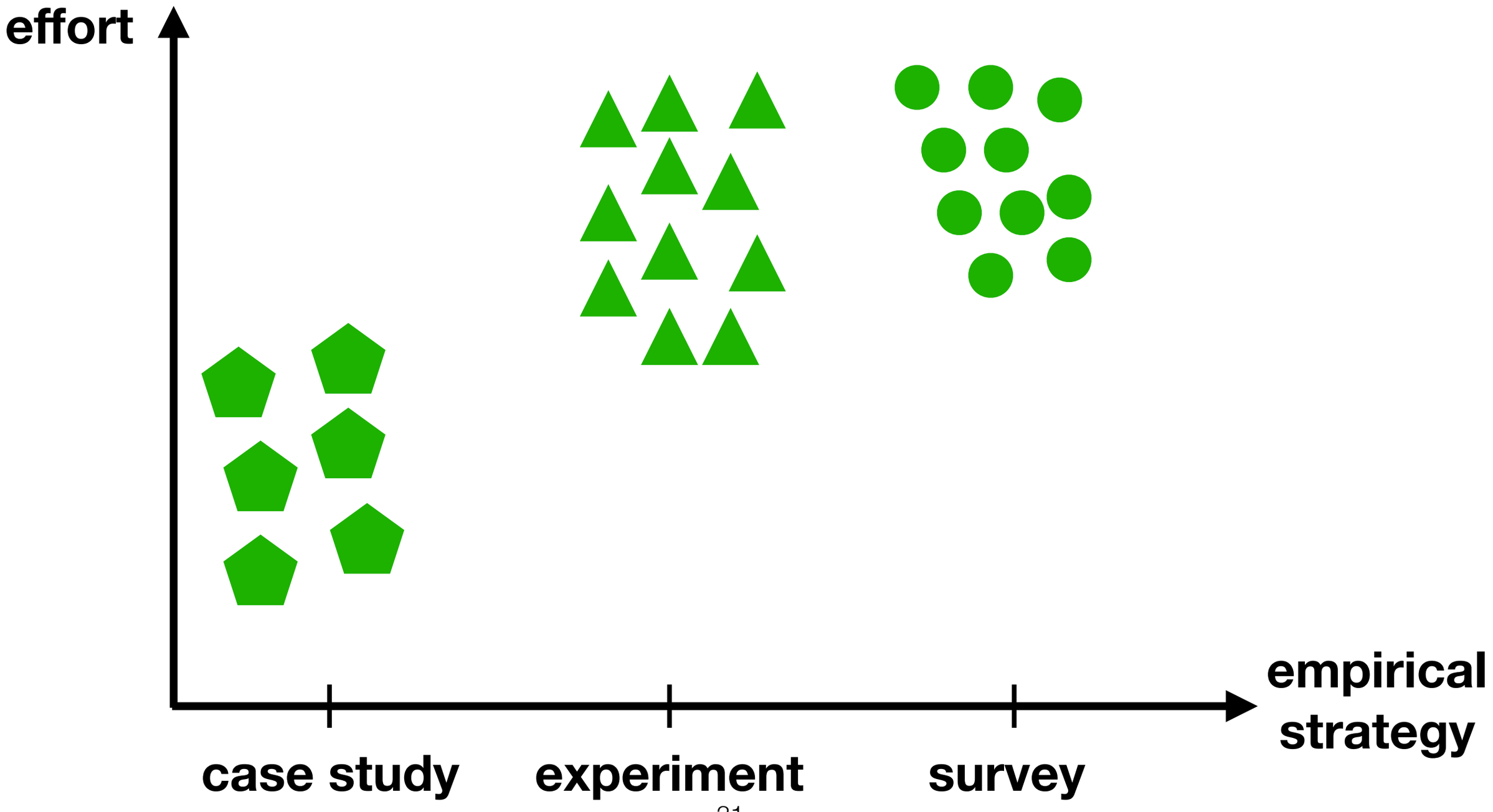
customize software tools and data analysis techniques to problem at hand



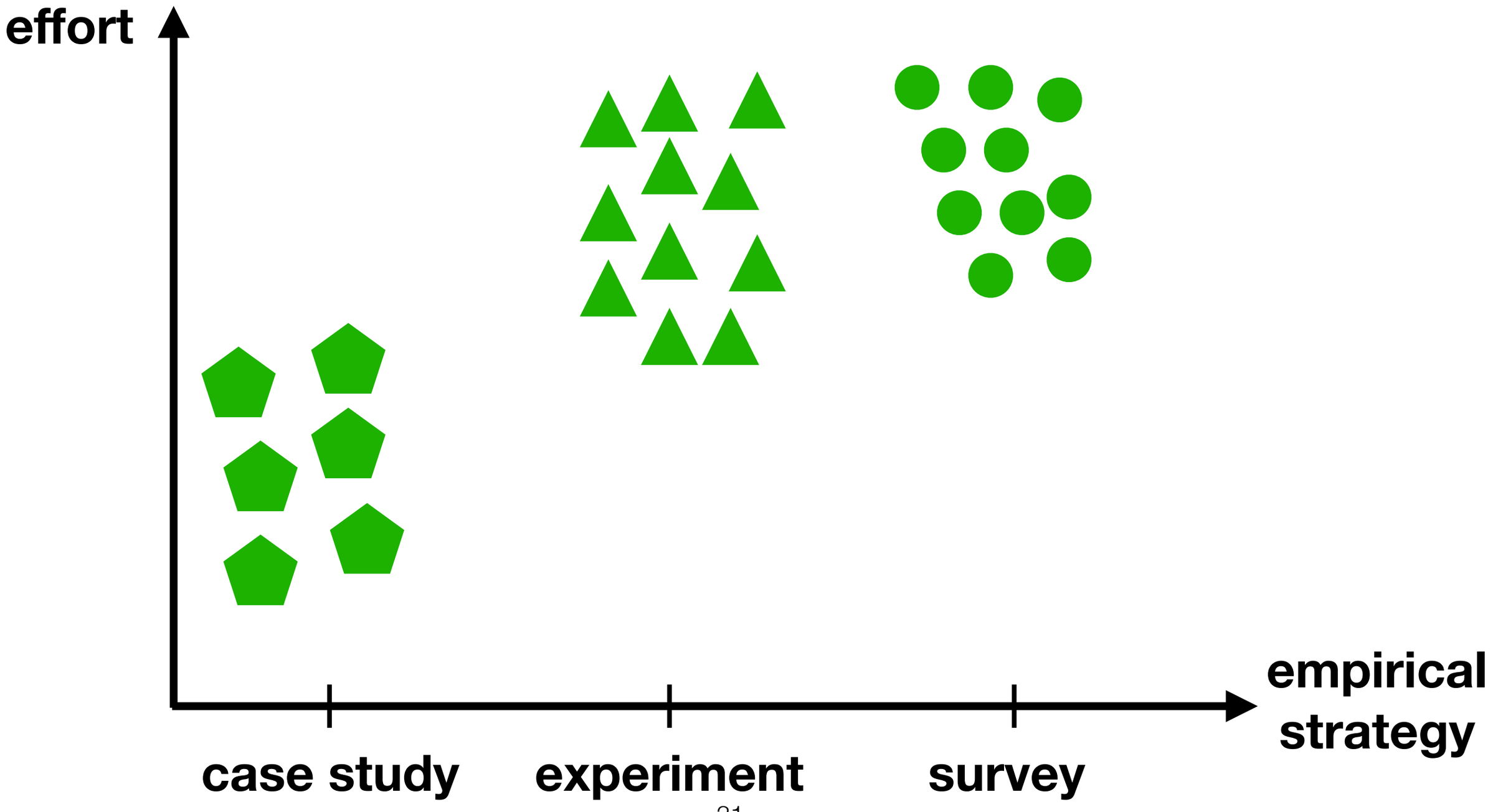

effort
empirical strategy
case study
experiment
survey


effort
empirical strategy
case study
experiment
survey

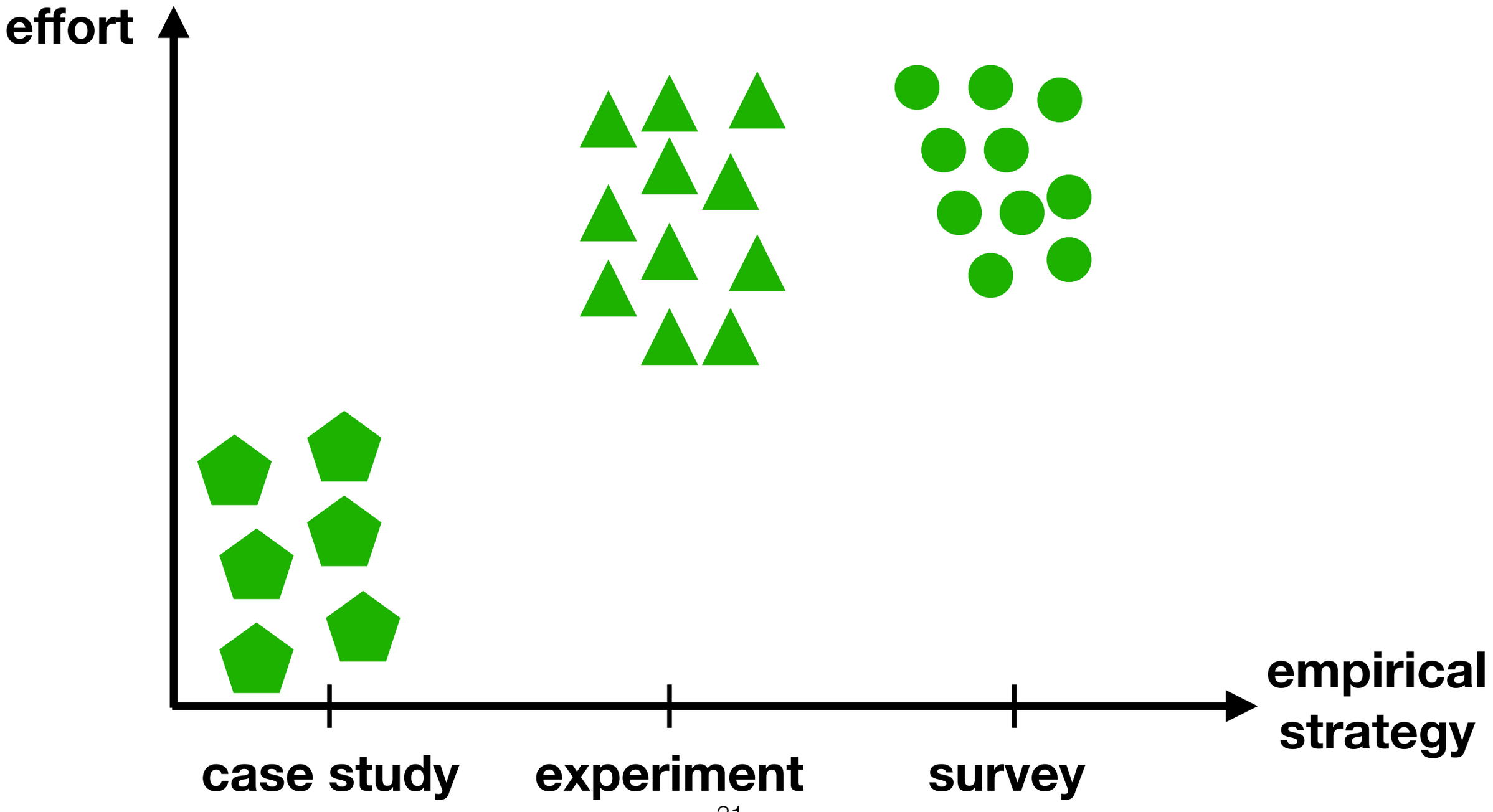

effort
empirical strategy
case study
experiment
survey


effort
empirical strategy
case study
experiment
survey

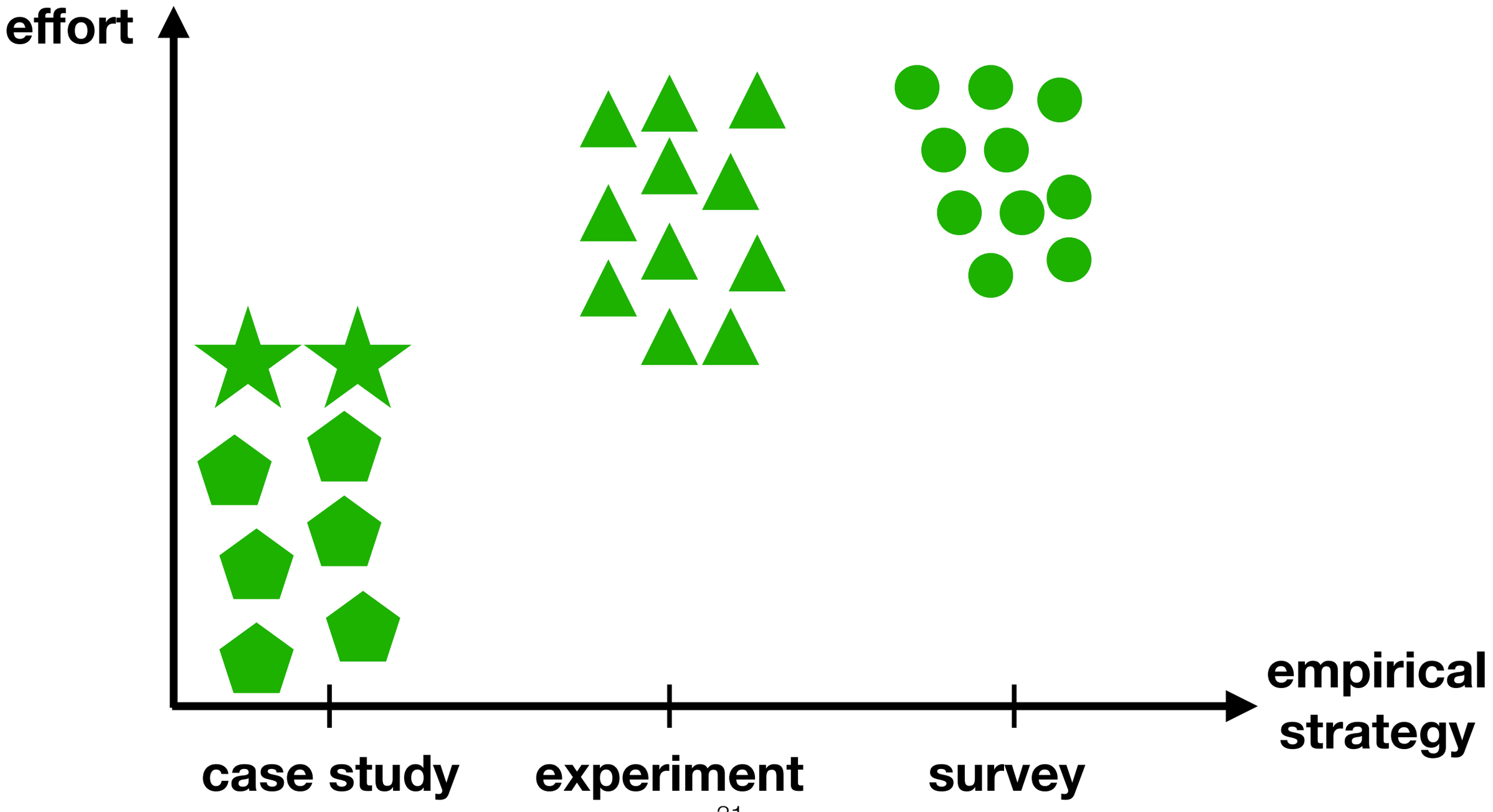
What if Accurate SE Process Data would be Available?



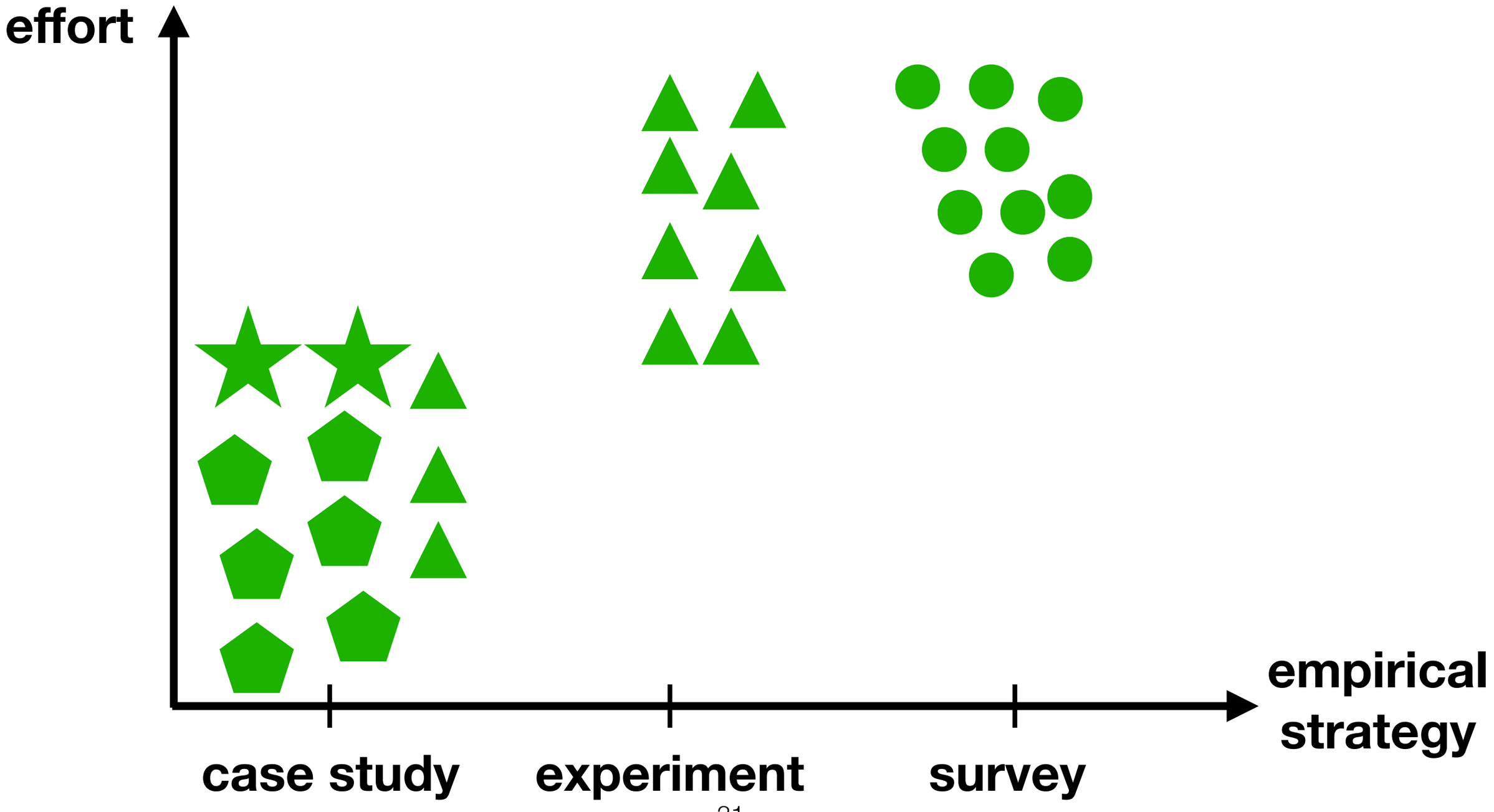
What if Accurate SE Process Data would be Available?



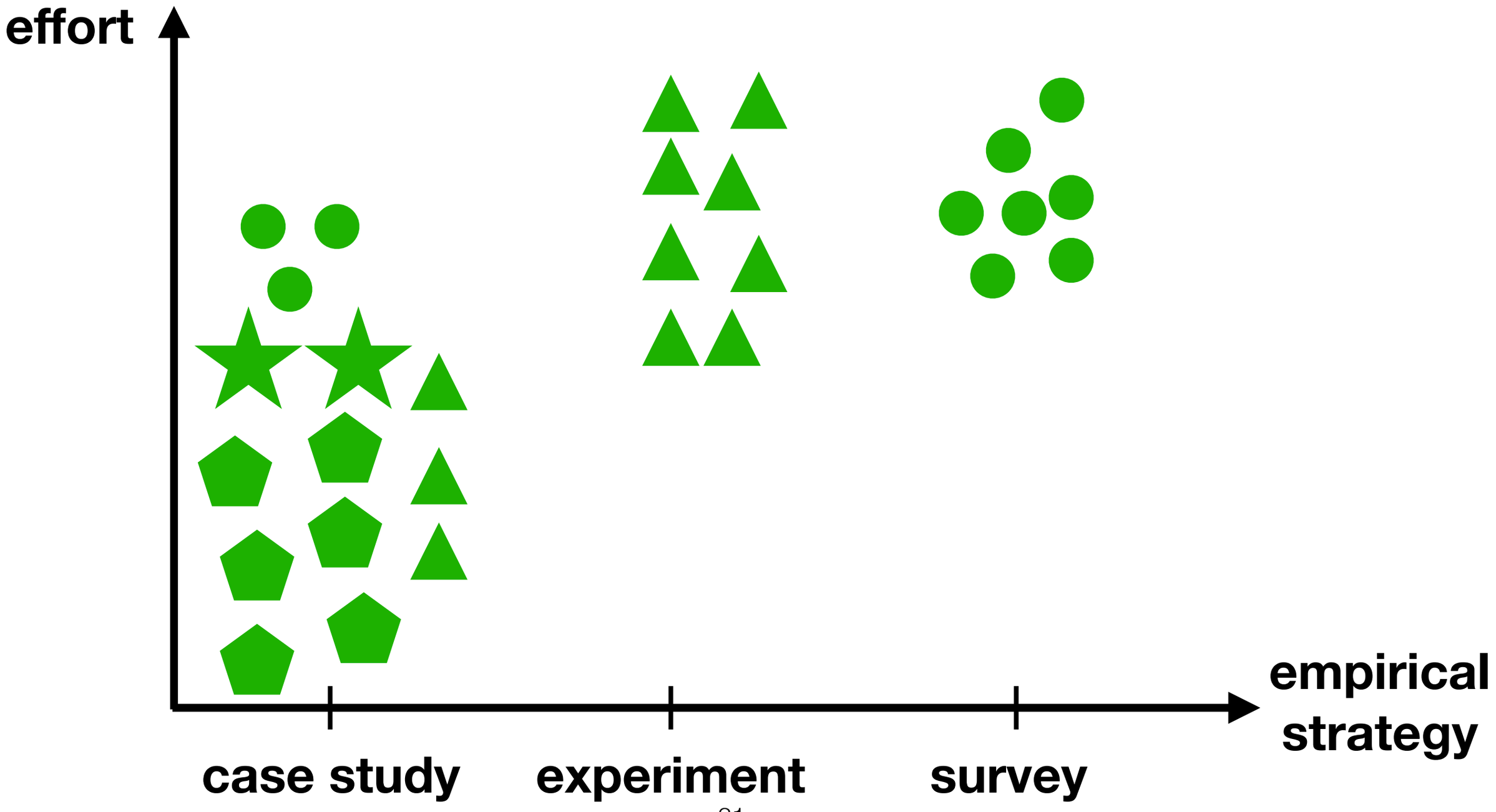
What if Accurate SE Process Data would be Available?



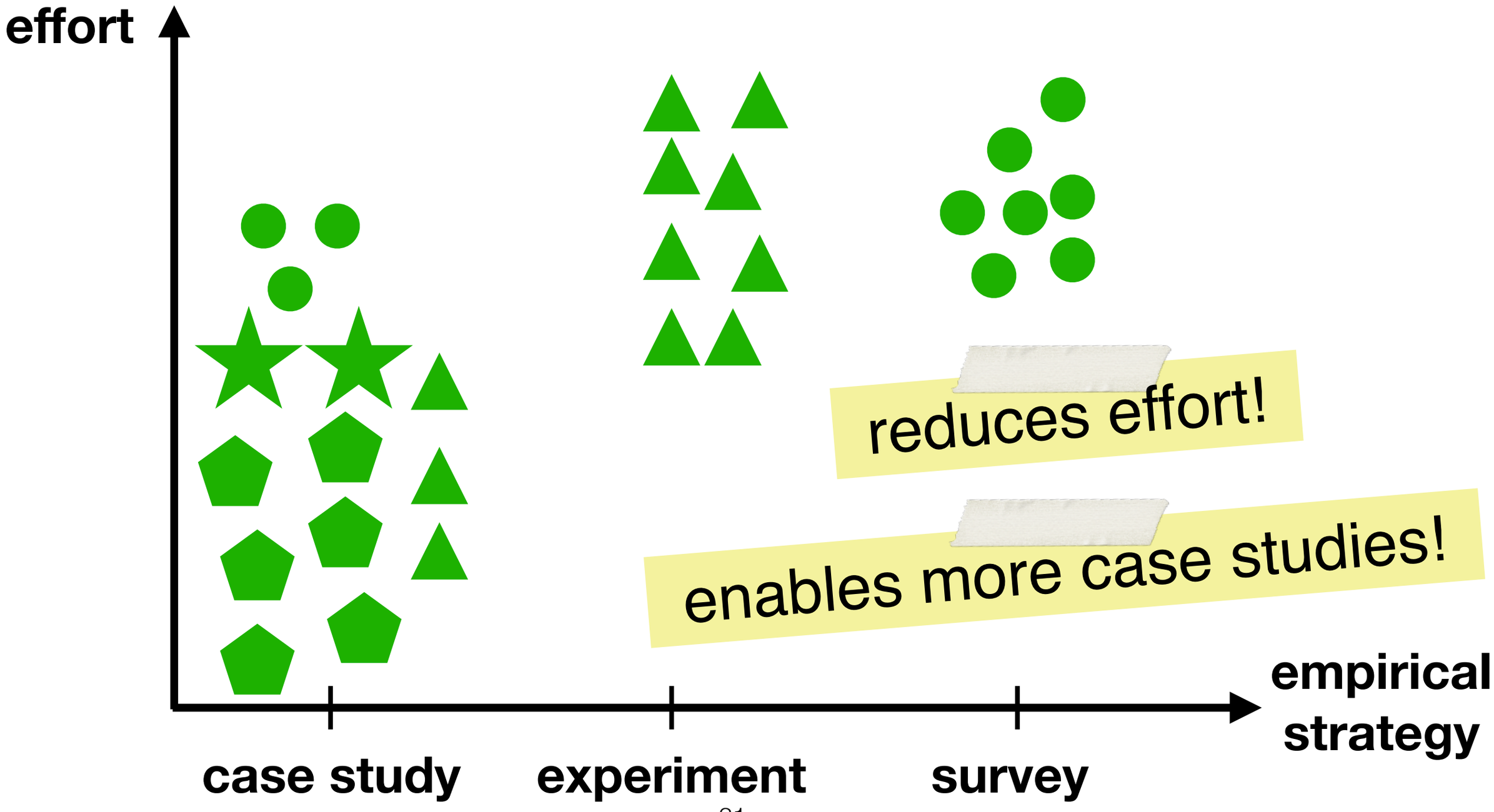
What if Accurate SE Process Data would be Available?



What if Accurate SE Process Data would be Available?



What if Accurate SE Process Data would be Available?



Part III: Software Analytics (aka MSR)

The Complete Open-Source Software Platform

Create, collaborate & distribute to over 33 million users worldwide

This Week: 21,642,929 Downloads 98,742 Code Commits

Search for Software, Business Solutions, or Resources

Make Your Projects Come To Life

With the tools we provide, developers on

For Developers By Developers

SourceForge is an Open Source community resource dedicated to helping open source projects be as successful as possible. We

Zimmermann T, Weißgerber P, Diehl S, Zeller A., "Mining version histories to guide software changes", 26th International Conference on Software Engineering (ICSE), pp. 563–572, 2004.

Mining Version Histories to Guide Software Changes

Thomas Zimmermann tz@acm.org
Peter Weißgerber weissger@st.cs.uni-sb.de
Stephan Diehl diehl@acm.org
Andreas Zeller zeller@acm.org

Saarland University, Saarbrücken, Germany

Abstract

We apply data mining to version histories in order to guide programmers along related changes: "Programmers who changed these functions also changed…". Given a set of existing changes, such rules (a) suggest and predict likely further changes, (b) show up item coupling that is undetectable by program analysis, and (c) prevent errors due to incomplete changes. After an initial change, our ROSE prototype can correctly predict 26% of further files to be changed—and 15% of the precise functions or variables. The topmost three suggestions contain a correct location with a likelihood of 64%.

1. Introduction

Shopping for a book at Amazon.com, you may have come across a section that reads "Customers who bought this book also bought…", listing other books that were typically included in the same purchase. Such information is gathered by *data mining*— the automated extraction of hidden predictive information from large data sets. In this paper, we apply data mining to *version histories:* "Programmers who changed these functions also changed…". Just like the Amazon.com feature helps the customer browsing along related items, our ROSE tool guides the programmer along related changes, with the following aims:

each time some programmer extended the fKeys[] array, she also extended the function that sets the preference default values. If the programmer now wanted to commit her changes *without* altering the suggested location, ROSE would issue a warning.

Detect coupling indetectable by program analysis. As ROSE operates uniquely on the version history, it is able to detect coupling between items that cannot be detected by program analysis—including coupling between items that are not even programs. In Figure 1, position 3 on the list is an ECLIPSE HTML documentation file with a confidence of 0.75—suggesting that after adding the new preference, the documentation should be updated, too.

ROSE is not the first tool to leverage version histories. In earlier work (Section 7), researchers have used history data to understand programs and their evolution [3], to detect evolutionary coupling between files [8] or classes [4], or to support navigation in the source code [6]. In contrast to this state of the art, the present work

- uses full-fledged *data mining techniques* to obtain association rules from version histories,
- detects coupling between fine-grained *program entities* such as functions or variables (rather than, say, classes), thus increasing precision and integrating with program analysis,

Mining Version Histories to Guide Software Changes

Thomas Zimmermann
tz@acm.org

Peter Weißgerber
weissger@st.cs.uni-sb.de

Stephan Diehl
diehl@acm.org

Andreas Zeller
zeller@acm.org

Saarland University, Saarbrücken, Germany

Abstract

We apply data mining to version histories in order to guide pro... ammers ... Given a ... predict ... at is in- ... ors due ... ROSE ... to be ... ables. ... ggestions contain a correct location with a likelihood of 64%.

what other functions should I change for this task?

1. Introduction

Shopping for a book at Amazon.com, you may have come across a section that reads "Customers who bought this book also bought...", listing other books that were typically included in the same purchase. Such information is gathered by *data mining*— the automated extraction of hidden predictive information from large data sets. In this paper, we apply data mining to *version histories:* "Programmers who changed these functions also changed...". Just like the Amazon.com feature helps the customer browsing along related items, our ROSE tool guides the programmer along related changes, with the following aims:

each time some programmer extended the fKeys[] array, she also extended the function that sets the preference default valu... the programmer now wanted ... the suggested

De...

similar to: "what did other people buy (in the past)?"

posi... tation file with a confidence of 0.75—suggesting that after adding the new preference, the documentation should be updated, too.

ROSE is not the first tool to leverage version histories. In earlier work (Section 7), researchers have used history data to understand programs and their evolution [3], to detect evolutionary coupling between files [8] or classes [4], or to support navigation in the source code [6]. In contrast to this state of the art, the present work

- uses full-fledged *data mining techniques* to obtain association rules from version histories,
- detects coupling between fine-grained *program entities* such as functions or variables (rather than, say, classes), thus increasing precision and integrating with program analysis,

Zimmermann T, Weißgerber P, Diehl S, Zeller A., "Mining version histories to guide software changes", 26th International Conference on Software Engineering (ICSE), pp. 563–572, 2004.

Mining Version Histories to Guide Software Changes

Thomas Zimmermann tz@acm.org
Peter Weißgerber weissger@st.cs.uni-sb.de
Stephan Diehl diehl@acm.org
Andreas Zeller zeller@acm.org
Saarland University, Saarbrücken, Germany

Abstract

We apply data mining to version histories in order to guide pro… …ammers … Given a … predict … at is in- … ors due … ROSE … to be … ables. … ggestions contain a correct location with a likelihood of 64%.

what other functions should I change for this task?

similar to: "what did other people buy (in the past)?"

so: mine version control system to **understand** which functions were changed together in the **past**, in order to **predict** the **future**

each time some programmer extended the fKeys[] array, she also extended the function that sets the preference default valu… …programmer now wanted … the suggested …

De…

posi… tation file with a confidence of 0.75—suggesting that after adding the new preference, the documentation … be updated, too.

1. Intro…

Shopping … across a … book als… cally inc… gathered… den pre… per, we … mers wh… like the Amazon.com feature helps the custom… along related items, our ROSE tool guides the programmer along related changes, with the following aims:

… In … data … etect … or to … this … n as- … *enti-* *ties* such as functions or varia… …, say, classes), thus increasing precision and integrating with program analysis.

Zimmermann T, Weißgerber P, Diehl S, Zeller A., "Mining version histories to guide software changes", 26th International Conference on Software Engineering (ICSE), pp. 563–572, 2004.

MSR 2005: International Workshop on Mining Software Repositories
2005.msrconf.org

Call for Papers (PDF)

17th May 2005
Saint Louis,
Missouri, USA

Co-located with ICSE 2005,
IEEE International Conference on Software Engineering
http://www.cs.wustl.edu/icse05/

Quick Links:

- Program
- Registration W10 and Accommodations
- new Proceedings new
- Important Dates
- MSR 2004 Website (MSR 2004 Summary)

Organizers

Ahmed E. Hassan
(aeehassa at plg dot uwaterloo.ca)
Richard C. Holt
(holt at plg dot uwaterloo.ca)
School of Computer Science
University of Waterloo
Ontario, Canada

Stephan Diehl
(diehl at cs dot uni-sb.de)
Dept. of Computer Science
Catholic University Eichstätt
Germany

Overview

Software repositories such as *source control systems*, *archived communications between project personnel*, and *defect tracking systems* are used to help manage the progress of software projects. Software practitioners and researchers are beginning to recognize the potential benefit of mining this information to *support the maintenance of software systems*, *improve software design/reuse*, and *empirically validate novel ideas and techniques*. Research is now proceeding to uncover the ways in which mining these repositories can help to understand software development, to support predictions about software development, and to plan various aspects of software projects.

The goal of this one-day workshop is to establish a community of researchers and practitioners who are working to recover and use the data stored in software repositories to further understanding of software development practices. We expect the presentations and discussions in this workshop to continue on a number of general themes and challenges, from the previous

The Mining Software Repositories (MSR) field **analyzes and cross-links** the rich data available in **software repositories** to uncover interesting and **actionable** information about software systems and projects

[Ahmed E. Hassan, 2008]

Ahmed E. Hassan, “The road ahead for mining software repositories”, Frontiers of Software Maintenance, pp. 48-57, 2008.

What Repositories are we Talking About?

What Repositories are we Talking About?

git

What Repositories are we Talking About?

What Repositories are we Talking About?

git

JIRA

source forge

What Repositories are we Talking About?














What Repositories are we Talking About?

What Repositories are we Talking About?









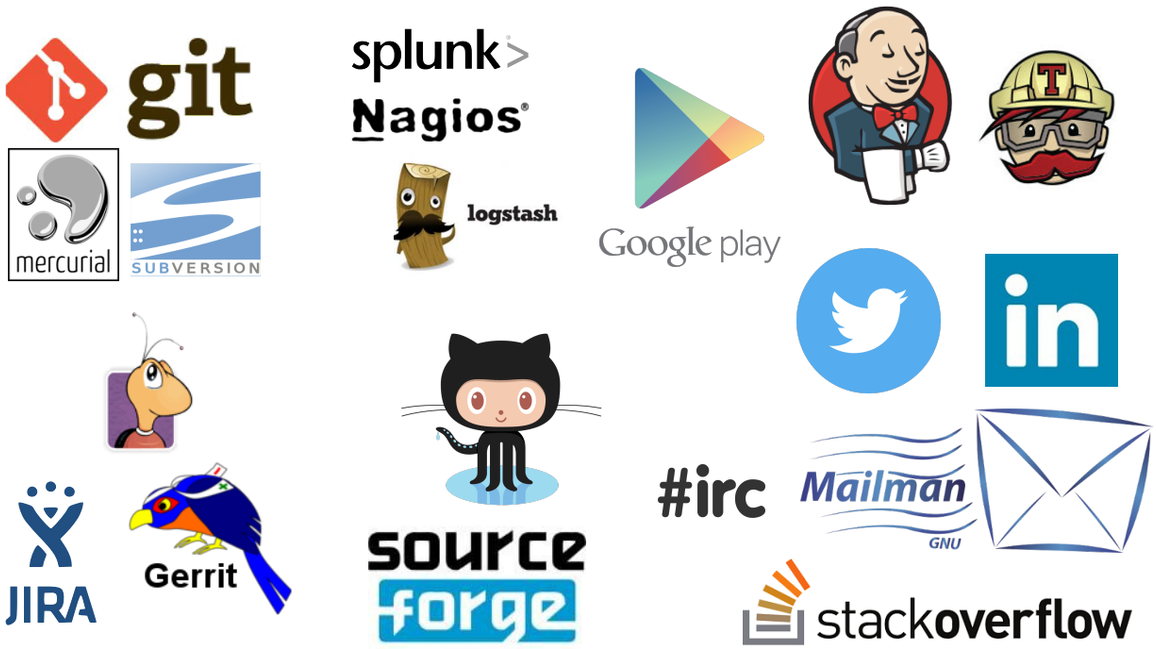


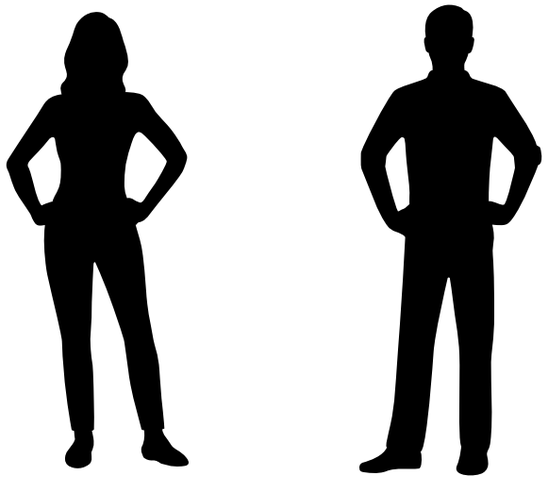


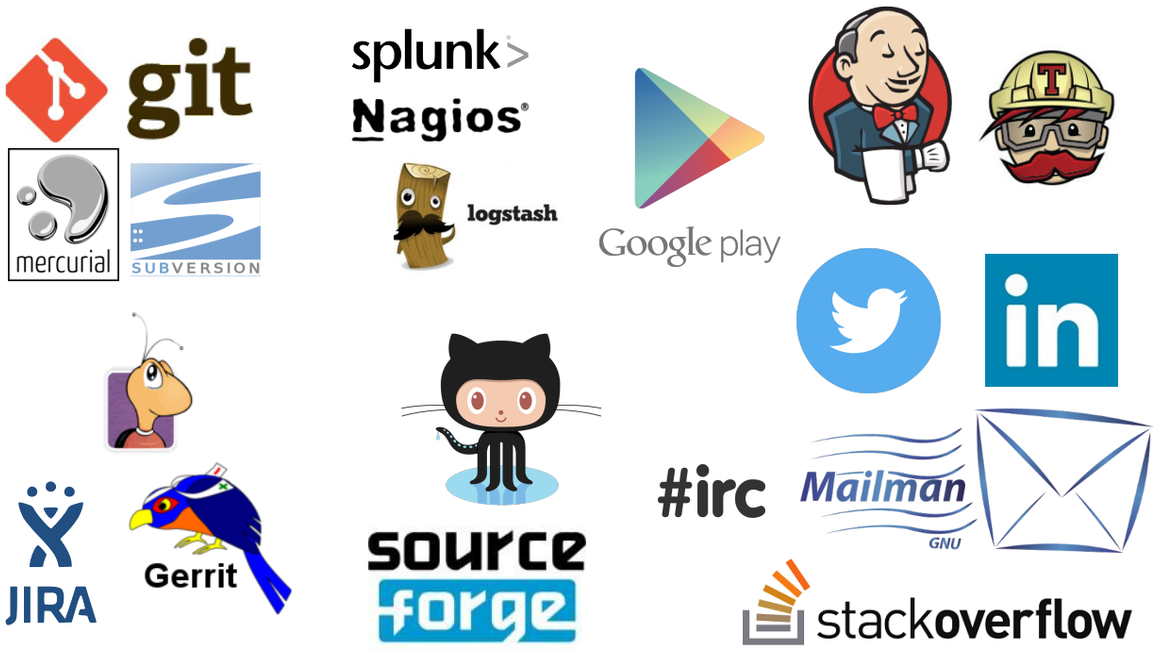


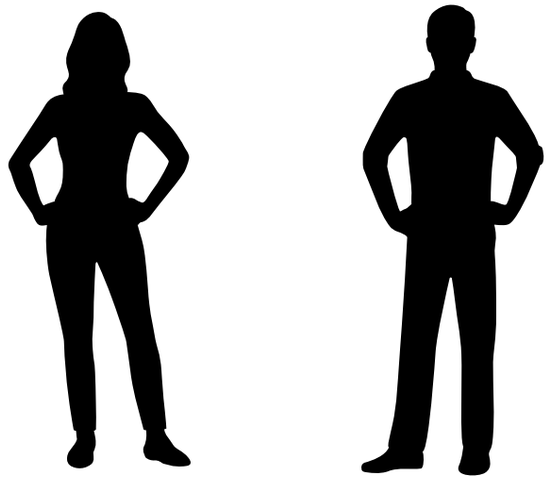


What Repositories are we Talking About?

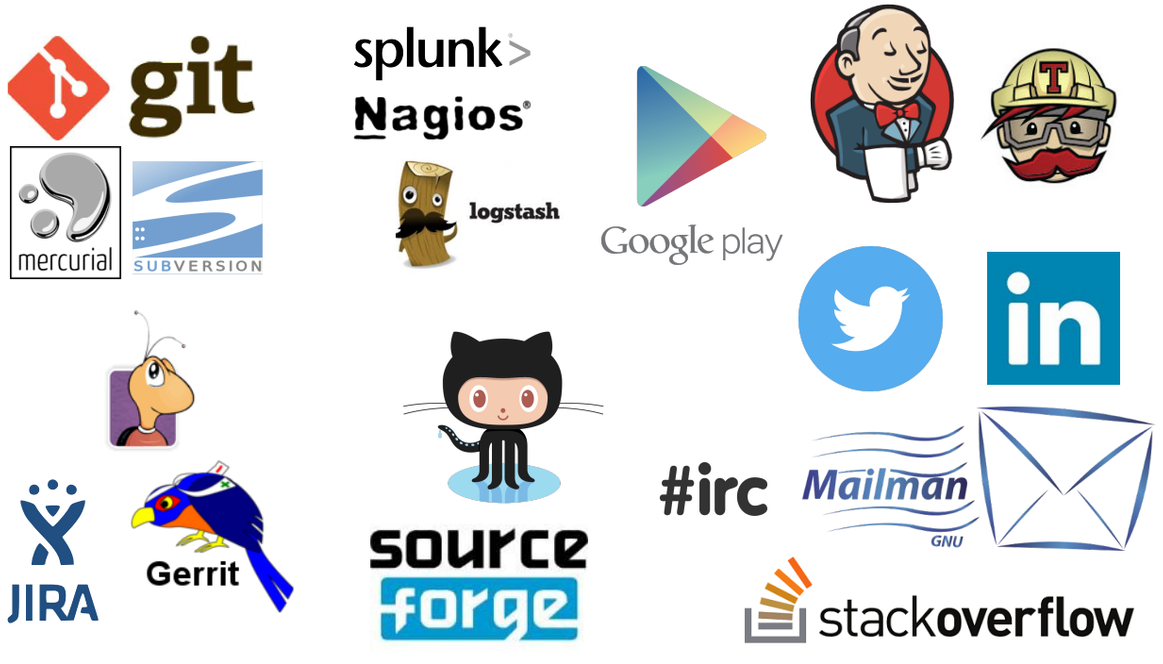

git
mercurial
SUBVERSION
splunk>
Nagios
logstash
Google play
JIRA
Gerrit
source
forge
#irc
Mailman
GNU
stackoverflow



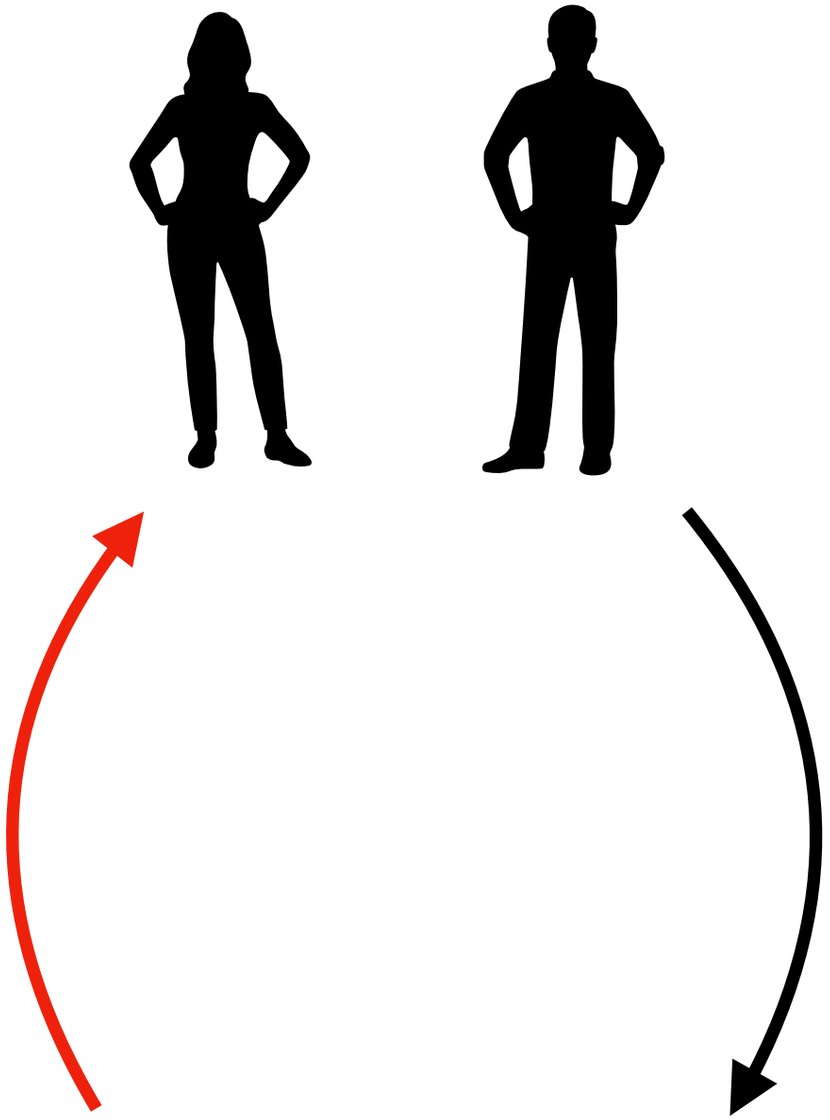

git
splunk>
Nagios
logstash
mercurial
SUBVERSION
Google play
#irc
Mailman
GNU
JIRA
Gerrit
source
forge
stackoverflow



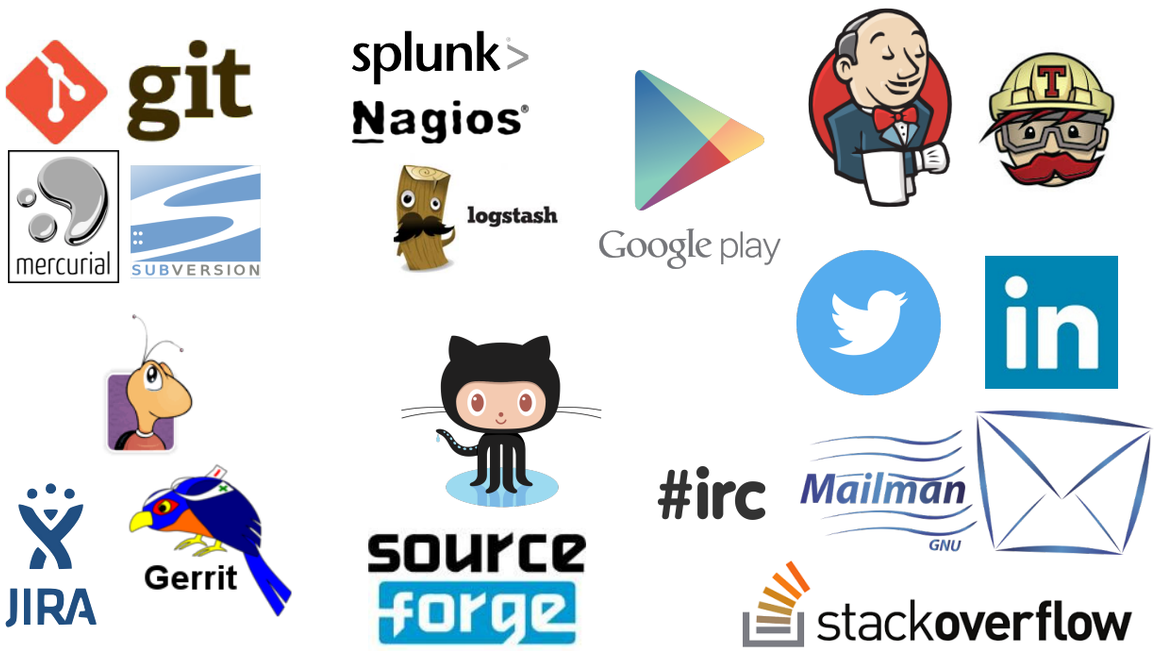
recording traces of activities



actionable insights

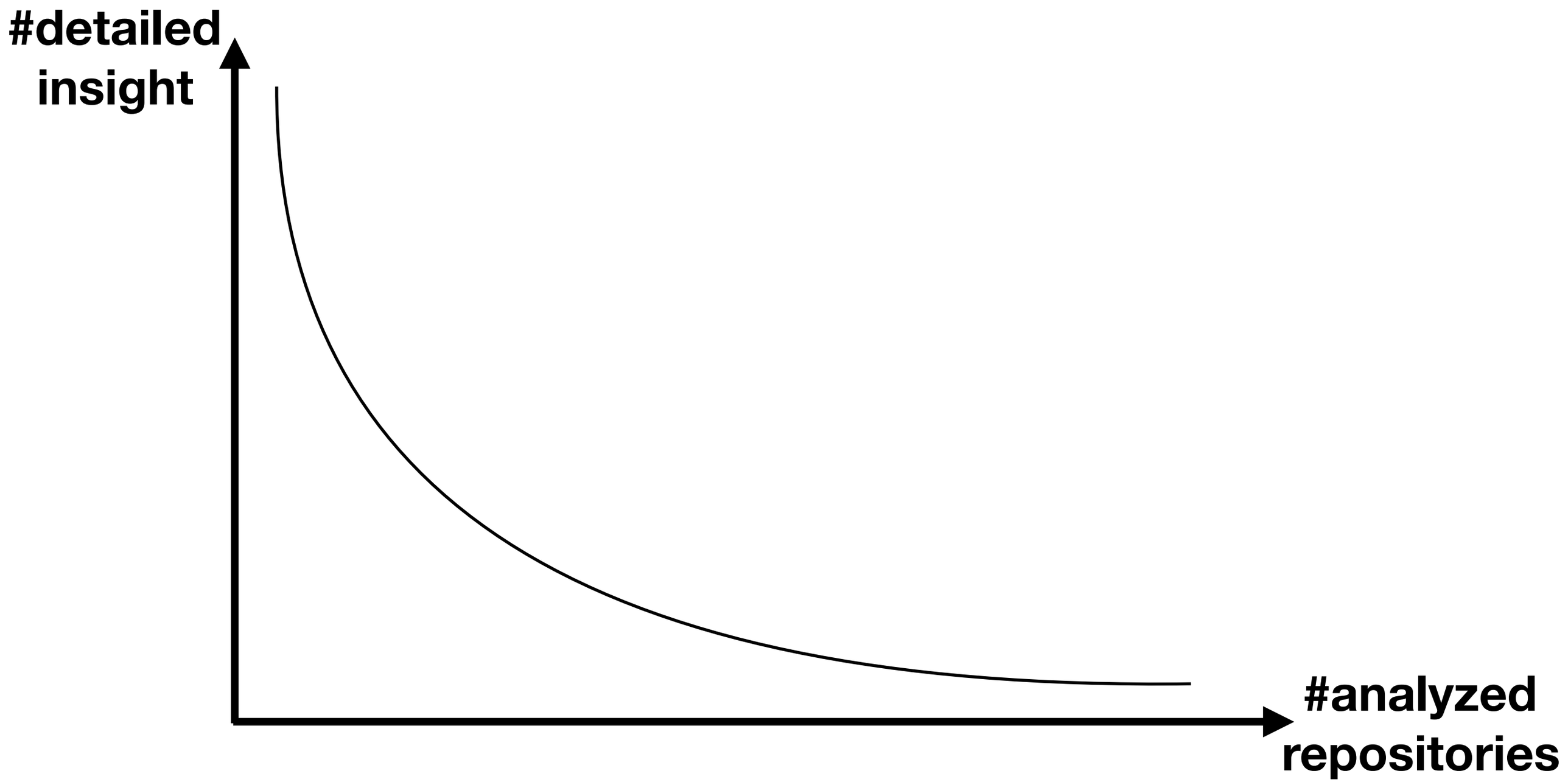


recording traces of activities

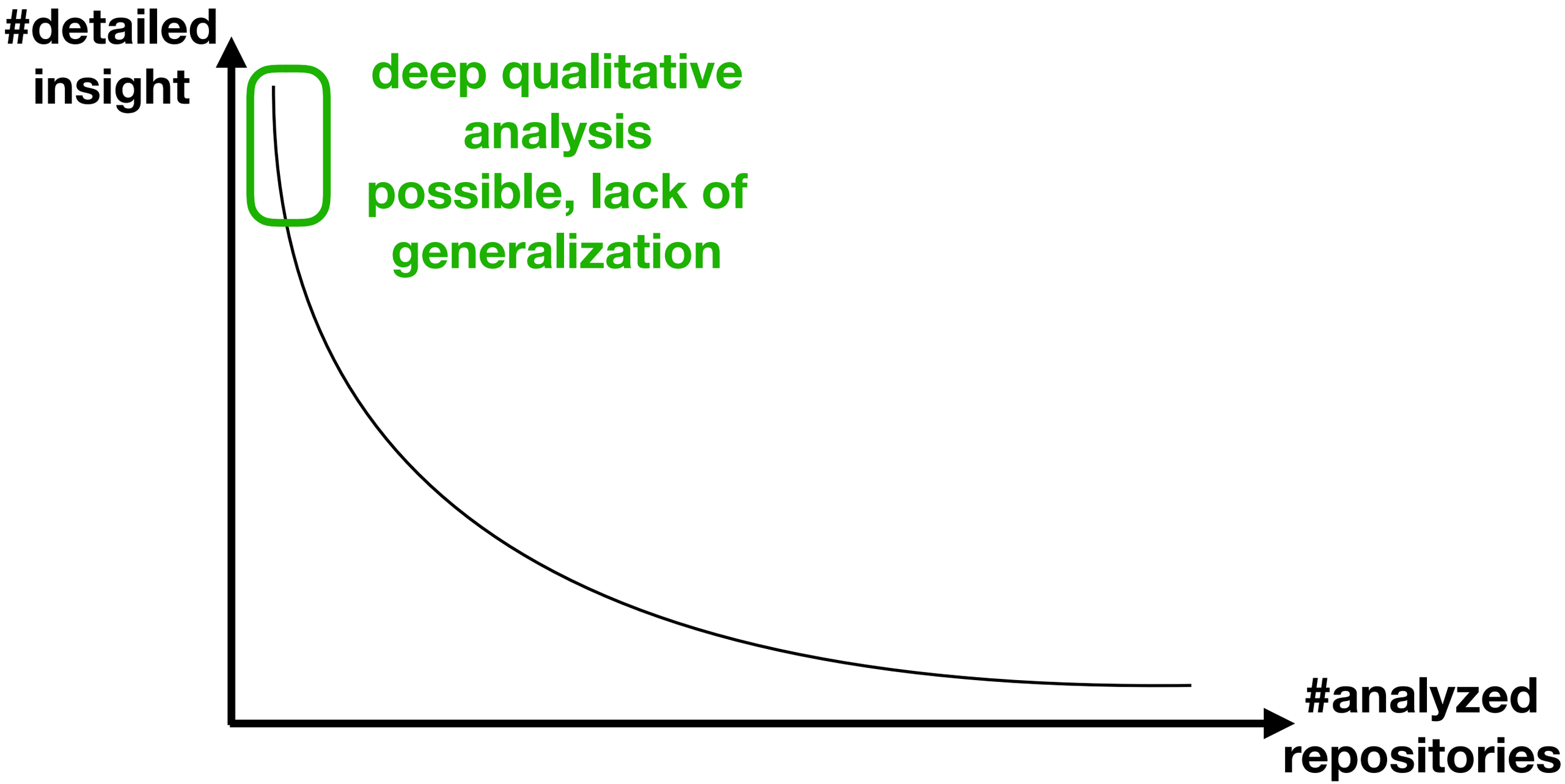


How Many Repositories and Projects are Required?

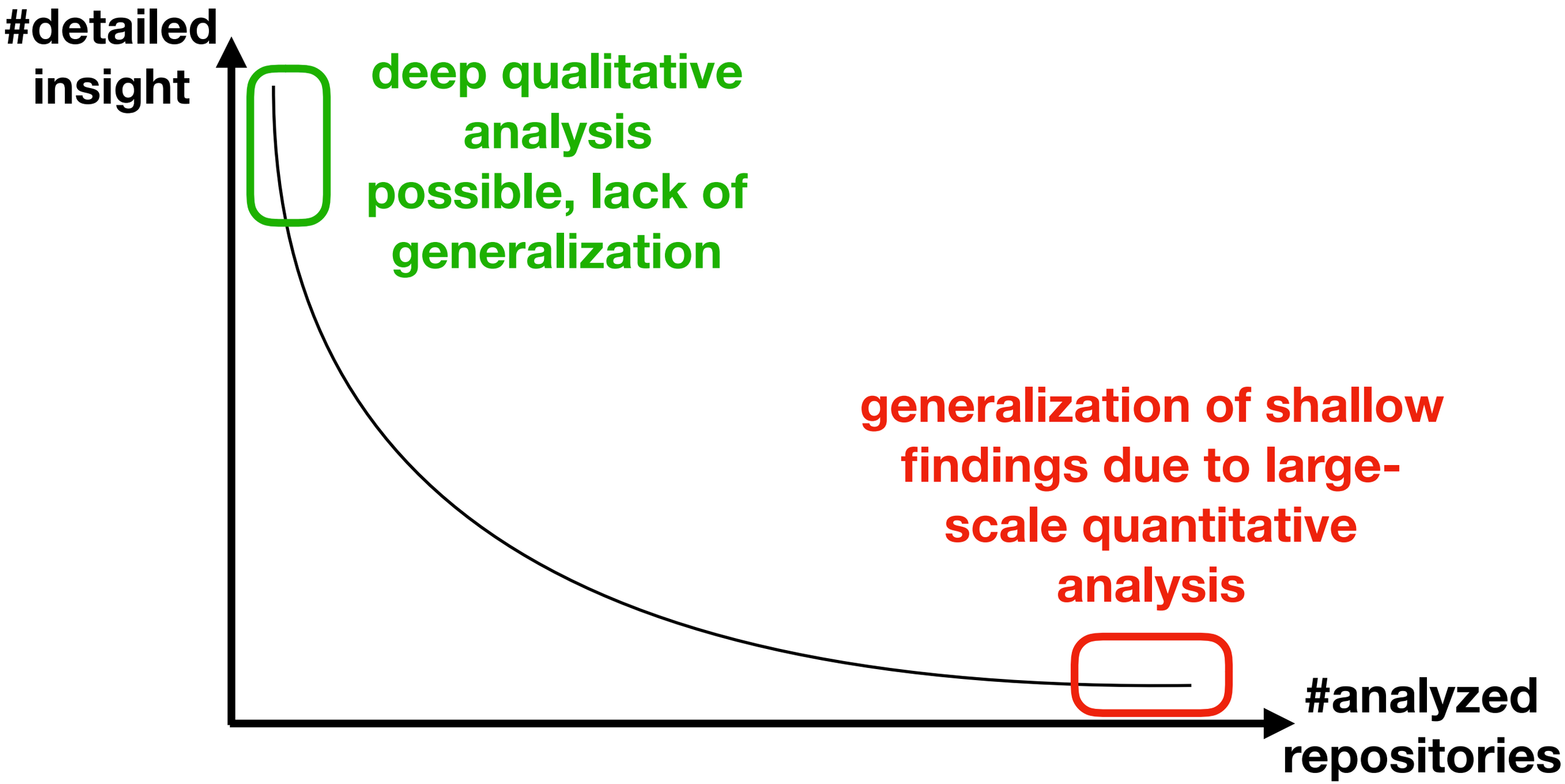
How Many Repositories and Projects are Required?



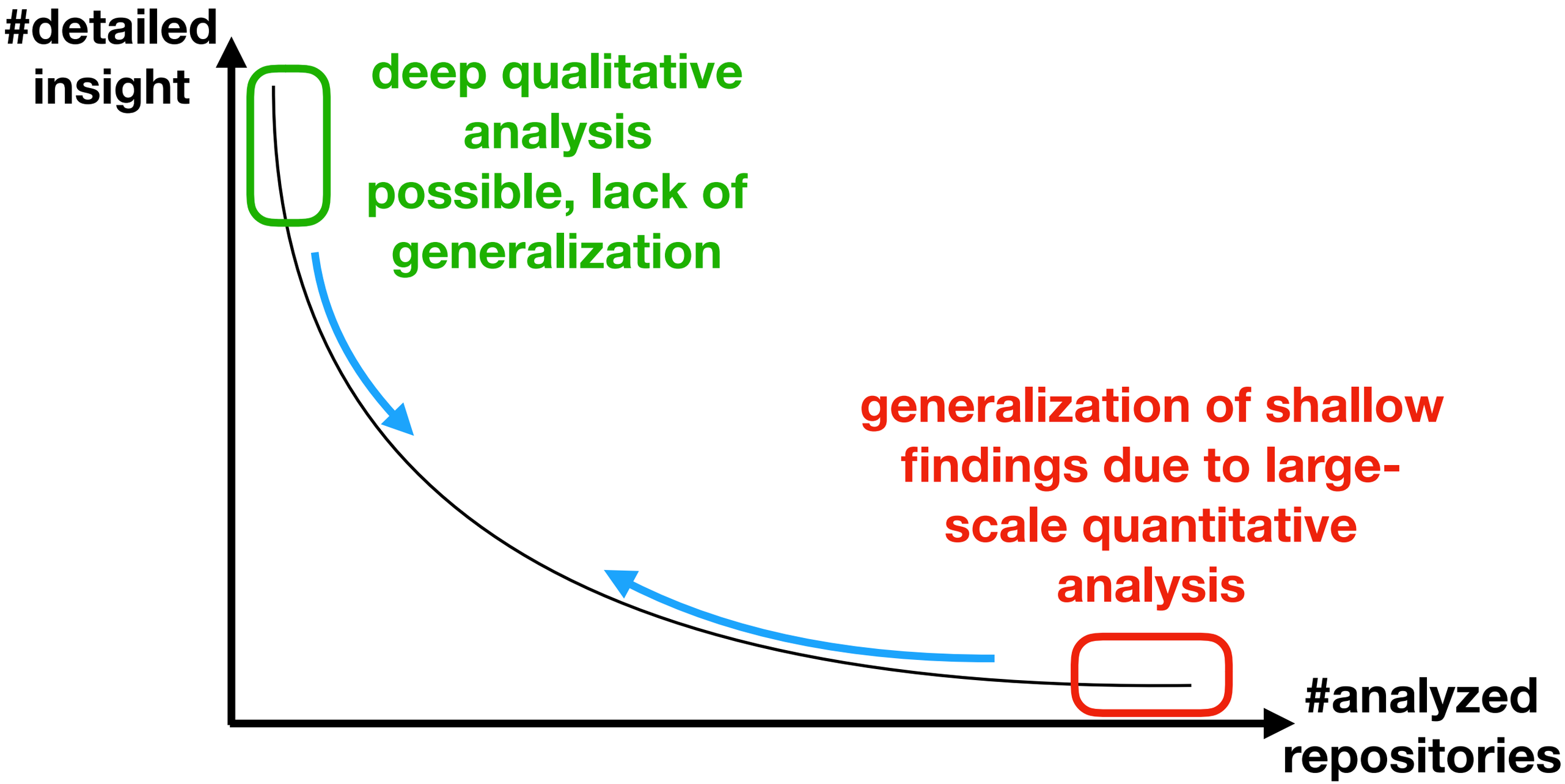
How Many Repositories and Projects are Required?



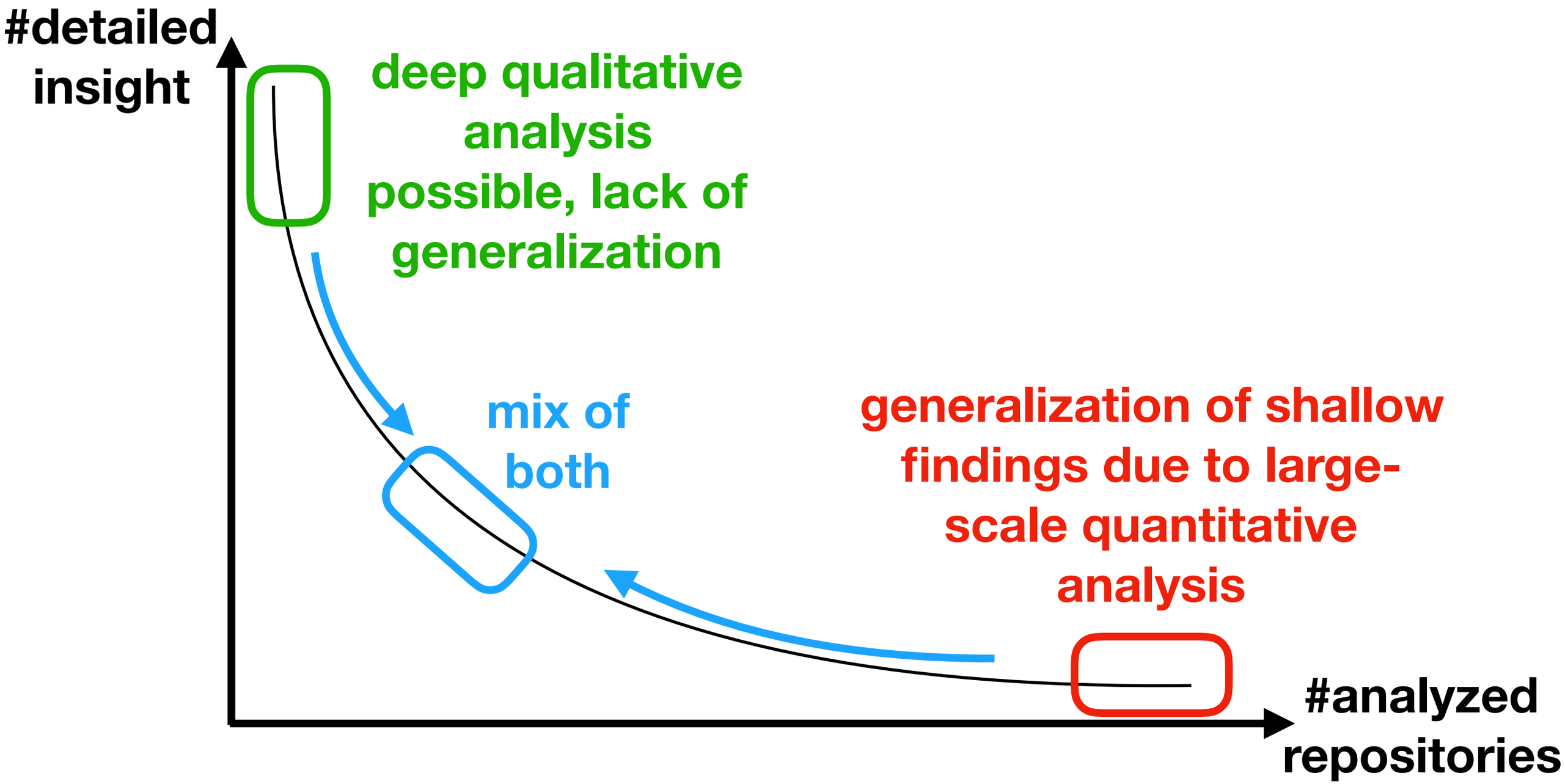
How Many Repositories and Projects are Required?



How Many Repositories and Projects are Required?



How Many Repositories and Projects are Required?




Hold On, What about that “Cross-Linking”?

Search or jump to... / Pull requests Issues Marketplace Explore

wireshark / wireshark

Watch 187

Code Pull requests 0 Insights

*shark: Update help and manpage for name resolving

```
Add 'v' option for VLAN ID resolving and get rid of
deprecated 'C' option.

Bug: 14826
Change-Id: I63104f4a465d251048693ad02882ea7eb2c4d926
Reviewed-on: https://code.wireshark.org/review/30029
Petri-Dish: Anders Broman <a.broman58@gmail.com>
Tested-by: Petri Dish Buildbot
Reviewed-by: Anders Broman <a.broman58@gmail.com>
Reviewed-on: https://code.wireshark.org/review/30039
```

master-2.4

uhei authored and **Anders Broman** committed 4 days ago 1 parent 0a74a3b commit ceed

Showing 10 changed files with **14 additions** and **10 deletions.**

Search or jump to... / Pull requests Issues Marketplace Explore

wireshark / wireshark

Watch 187

Code Pull requests 0 Insights

*shark: Update help and manpage for name resolving

```
Add 'v' option for VLAN ID resolving and get rid of
deprecated 'C' option.


Bug: 14826
Change-Id: I63104f4a465d251048693ad02882ea7eb2c4d926
Reviewed-on: https://code.wireshark.org/review/30029
Petri-Dish: Anders Broman <a.broman58@gmail.com>
Tested-by: Petri Dish Buildbot
Reviewed-by: Anders Broman <a.broman58@gmail.com>
Reviewed-on: https://code.wireshark.org/review/30039
```

how did reviewers react to this commit?

master-2.4

uhei authored and **Anders Broman** committed 4 days ago 1 parent 0a74a3b commit ceed

Showing 10 changed files with **14 additions** and **10 deletions**.

All Projects Documentation

Search term

Open Merged Abandoned

Change 30029 - **Merged** Included in ▼ Patch Sets (3/3) ▼ Download ▼

```
*shark: Update help and manpage for name resolving

Add 'v' option for VLAN ID resolving and get rid of
deprecated 'C' option.

Bug: 14826
Change-Id: I63104f4a465d251048693ad02882ea7eb2c4d926
Reviewed-on: https://code.wireshark.org/review/30029
Petri-Dish: Anders Broman <a.broman58@gmail.com>
Tested-by: Petri Dish Buildbot
Reviewed-by: Anders Broman <a.broman58@gmail.com>
```

Owner	Uli Heilmeier
Uploader	Anders Broma
Assignee	
Reviewers	Anders Brom Petri Dish Buildb
Project	wireshark
Branch	master
Topic	bug/14826
Updated	4 days ago

Author	Uli Heilmeier <uh@heilmeier.eu>	Oct 5, 2018 9:54 AM
Committer	Anders Broman <a.broman58@gmail.com>	Oct 5, 2018 4:19 PM
Commit	8dfaa8fa7c97cd1372a0a233b83fbc7945447b75	
Parent(s)	75c46e80bf2e6db3d59be99e63ac0d4243c5878e	
Change-Id	I63104f4a465d251048693ad02882ea7eb2c4d926	

Code-Review	+2 Anders Bro
Petri-Dish	+1 Anders Bro
Verified	+1 Petri Dish Buil

Files Open All Diff against: Base

File Path	Comments	Size
Commit Message		
doc/rawshark.pod		
doc/tshark.pod		

Search or jump to... / Pull requests Issues Marketplace Explore

wireshark / wireshark

Watch 187

Code Pull requests 0 Insights

*shark: Update help and manpage for name resolving

```
Add 'v' option for VLAN ID resolving and get rid of
deprecated 'C' option.

Bug: 14826
Change-Id: I63104f4a465d251048693ad02882ea7eb2c4d926
Reviewed-on: https://code.wireshark.org/review/30029
Petri-Dish: Anders Broman <a.broman58@gmail.com>
Tested-by: Petri Dish Buildbot
Reviewed-by: Anders Broman <a.broman58@gmail.com>
Reviewed-on: https://code.wireshark.org/review/30039
```

master-2.4

uhei authored and **Anders Broman** committed 4 days ago 1 parent 0a74a3b commit ceed

Showing 10 changed files with **14 additions** and **10 deletions.**

Search or jump to... / Pull requests Issues Marketplace Explore

wireshark / wireshark

Watch 187

<> Code | Pull requests 0 | Insights

*shark: Update help and manpage for name resolving

```
Add 'v' option for VLAN ID resolving and get rid of
deprecated 'C' option.

Bug: 14826
Change-Id: I63104f4a465d251048693ad02882ea7eb2c4d926
Reviewed-on: https://code.wireshark.org/review/30029
Petri-Dish: Anders Broman <a.broman58@gmail.com>
Tested-by: Petri Dish Buildbot
Reviewed-by: Anders Broman <a.broman58@gmail.com>
Reviewed-on: https://code.wireshark.org/review/30039
```

does this commit fix a bug or add a new feature?

master-2.4

uhei authored and Anders Broman committed 4 days ago 1 parent 0a74a3b commit ceed

Showing 10 changed files with 14 additions and 10 deletions.

Bug 14826 - Undocumented sub-option for -N option in man page and tshark -N help

Status: RESOLVED FIXED
Alias: None
Product: Wireshark
Component: TShark (show other bugs)
Version: Git
Hardware: x86 Linux
Importance: Low Normal (vote)
Target Milestone: ---
Assignee: Bugzilla Administrator
URL:
Depends on:
Blocks:

Reported: 2018-06-05 10:00 UTC by Michal Ruprich
Modified: 2018-10-06 16:39 UTC (History)
CC List: 1 user (show)
See Also: *14692*

Quick Search

Attachments

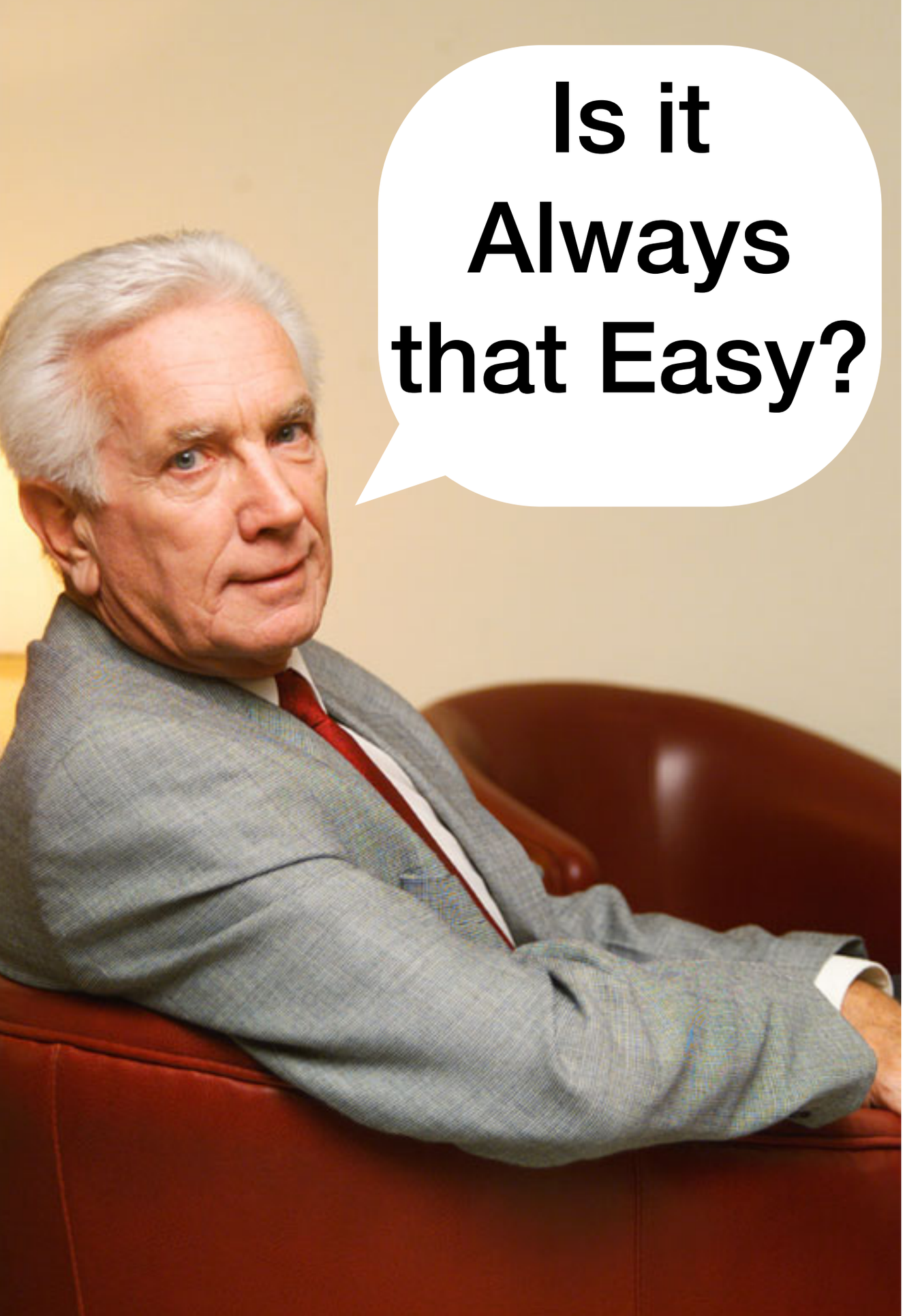
Add an attachment (proposed patch, testcase, etc.)

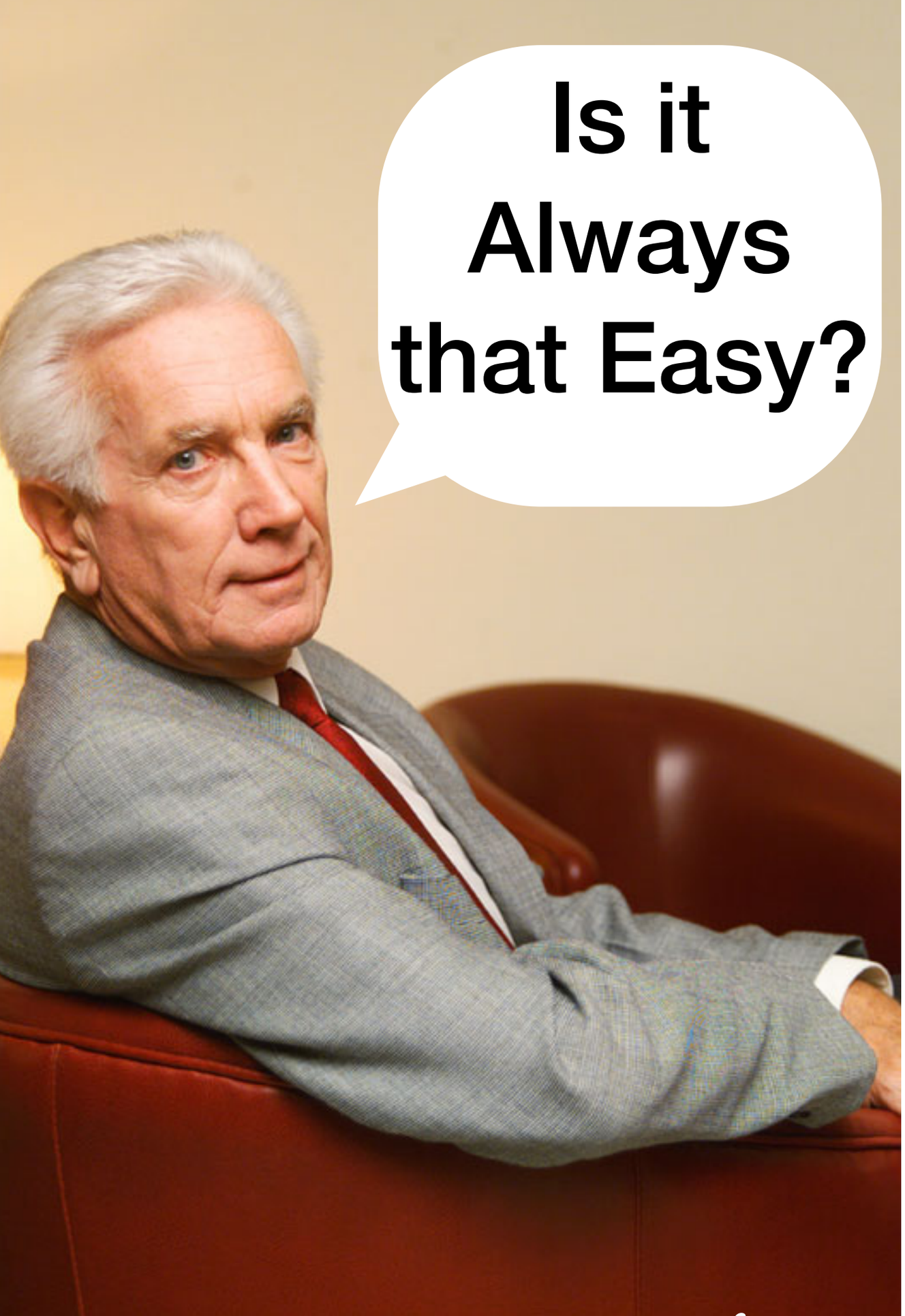
Note

You need to log in before you can comment on or make changes to this bug.

Michal Ruprich **2018-06-05 10:00:28 UTC** **Description**

```
Build Information:
Paste the COMPLETE build information from "Help->About Wireshark", "wireshark -v",
or "tshark -v".
--
Man page on -N option mentions sub-options d, m, n, N and t. There is, however,
```


Is it
Always
that Easy?



NO!

Linking Repositories is Challenging!

Linking Repositories is Challenging!

no recorded links due to lack of **discipline** developers

lack of access to confidential/ closed repositories

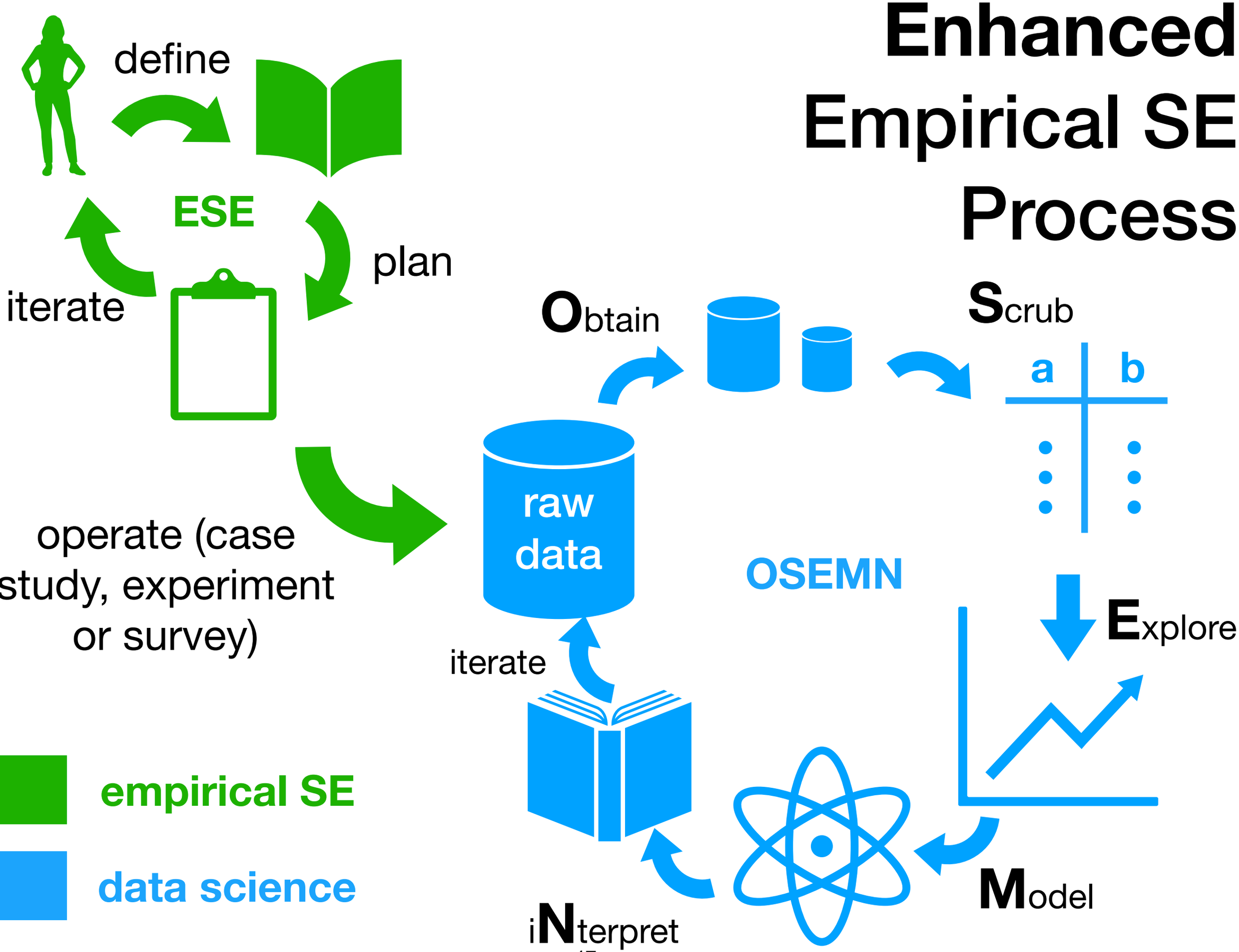
ambiguous or incorrect links

invalid identifiers, e.g., after **migration** to newer repository technology

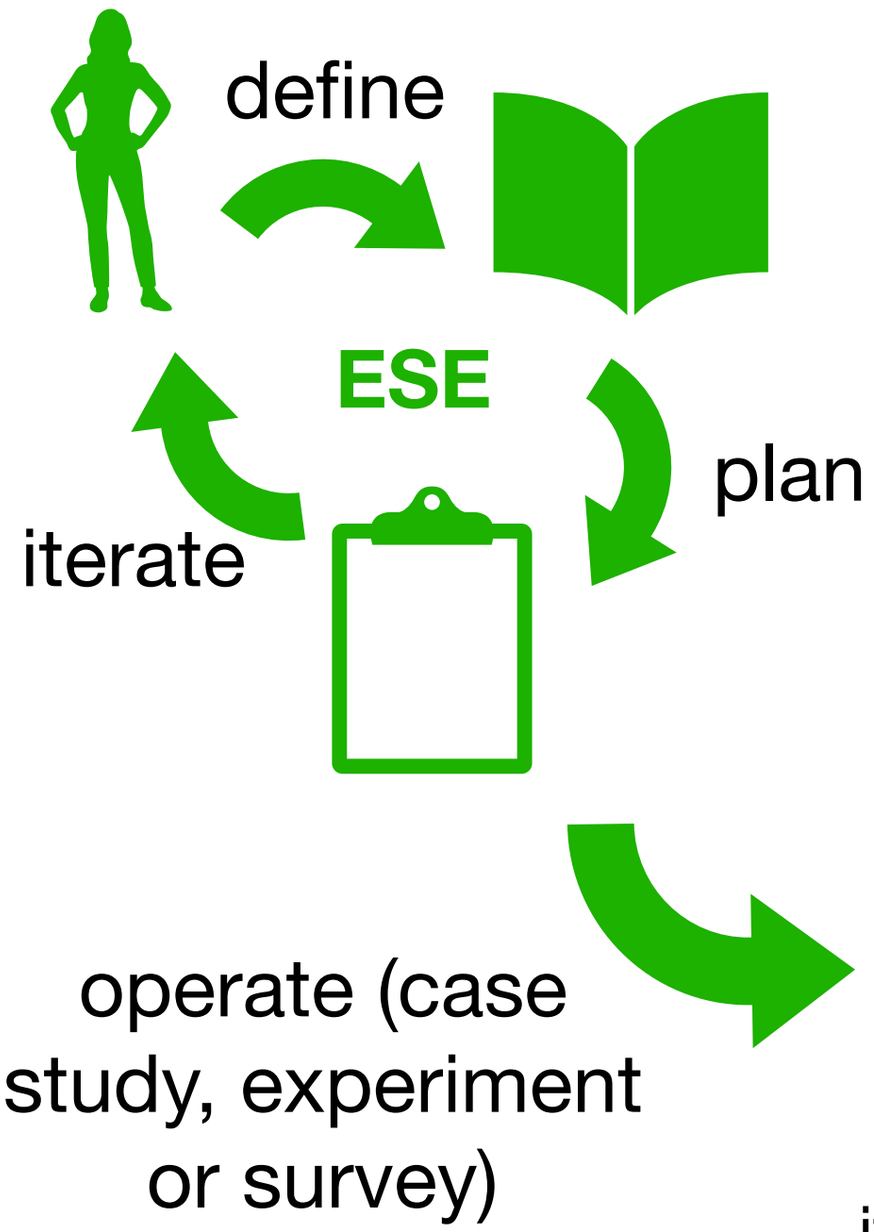
imprecise heuristics based on email addresses, …

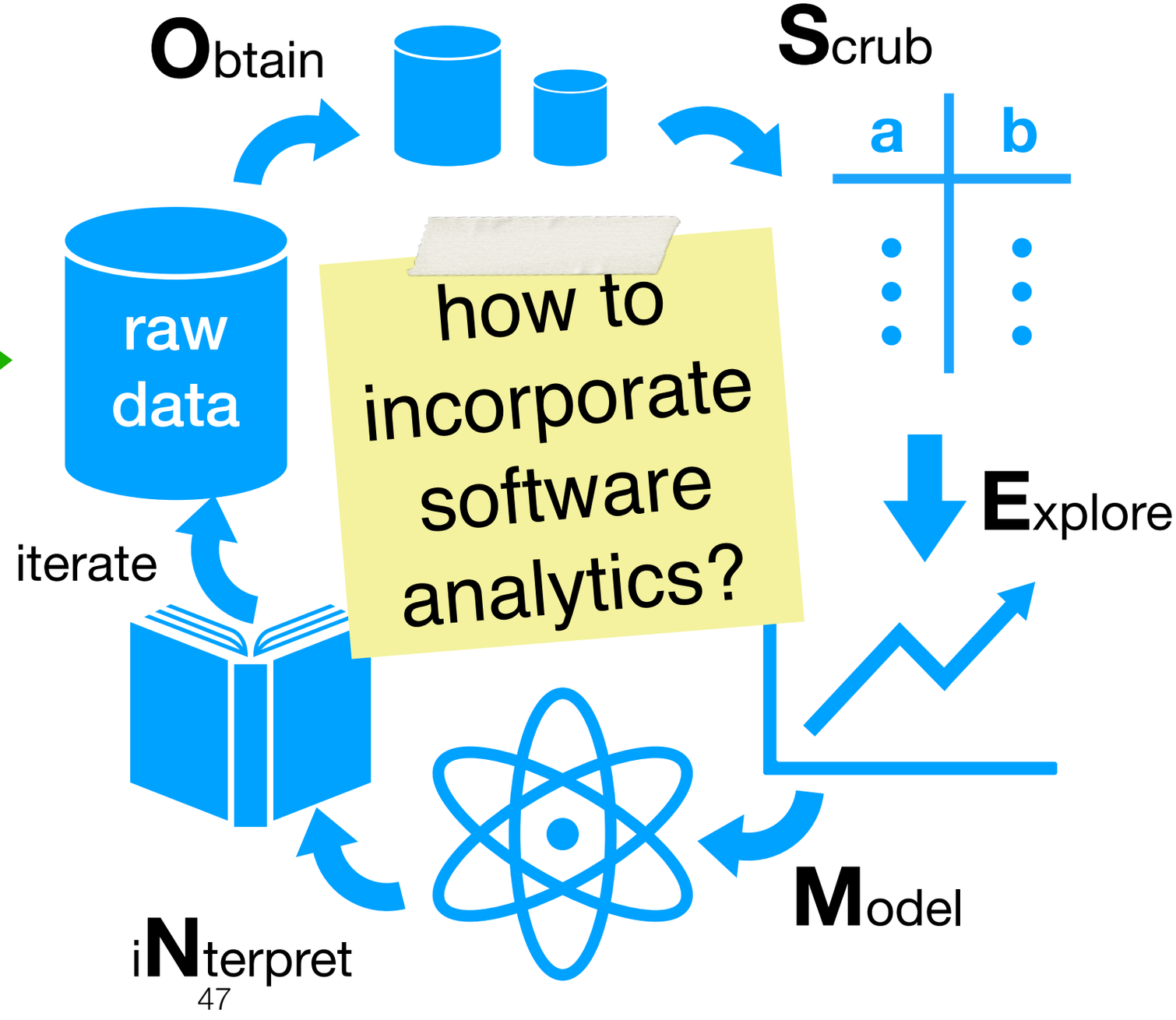
not enough **data left** after linking

Enhanced Empirical SE Process



Enhanced Empirical SE Process

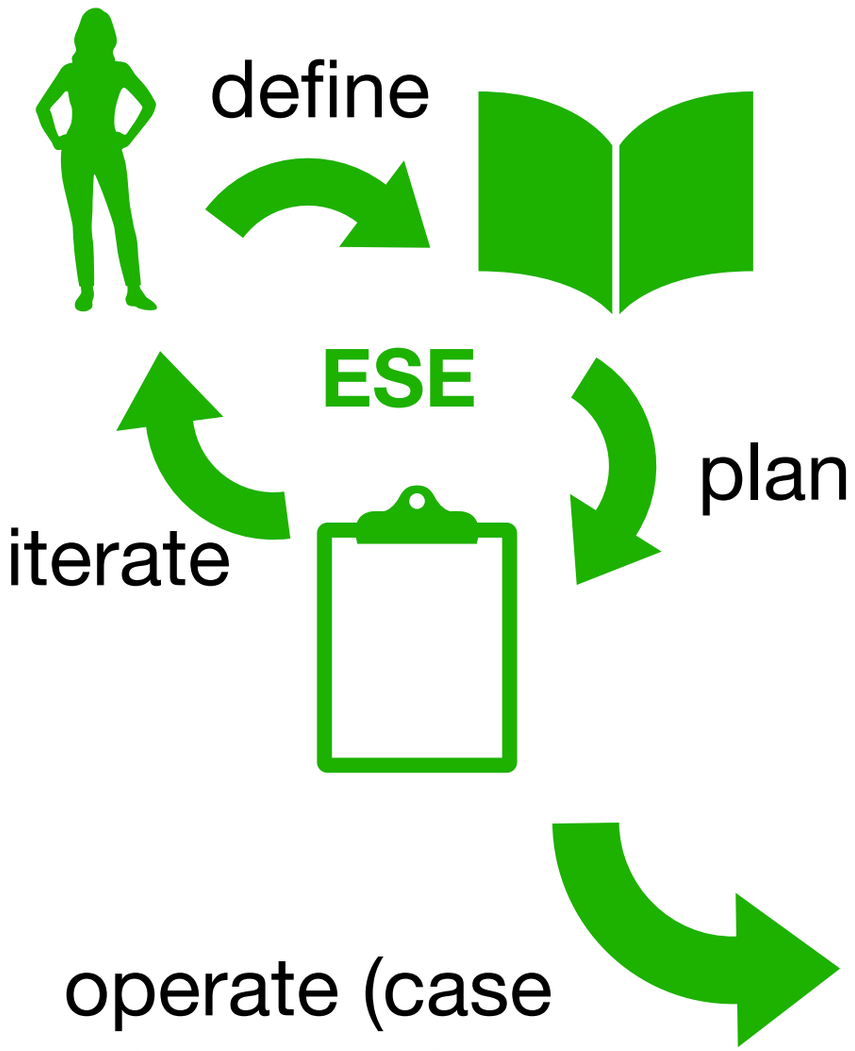




empirical SE

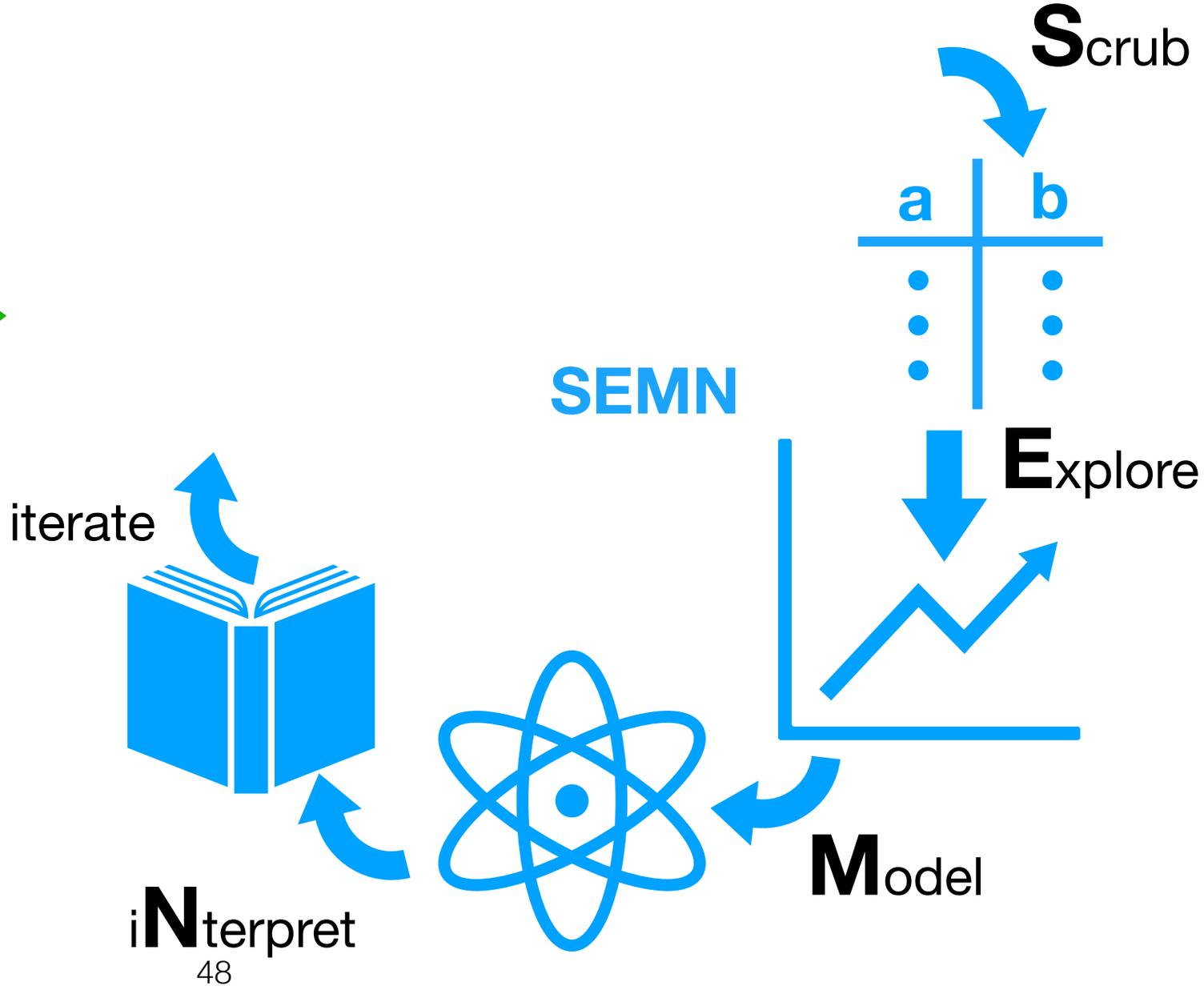
data science

Today's Empirical SE Process

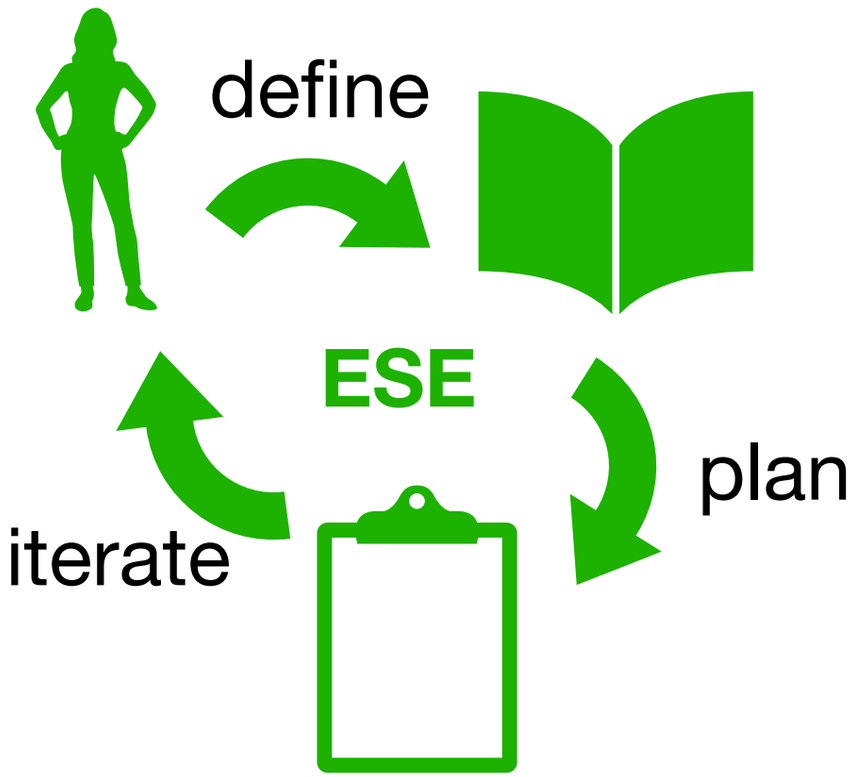


empirical SE

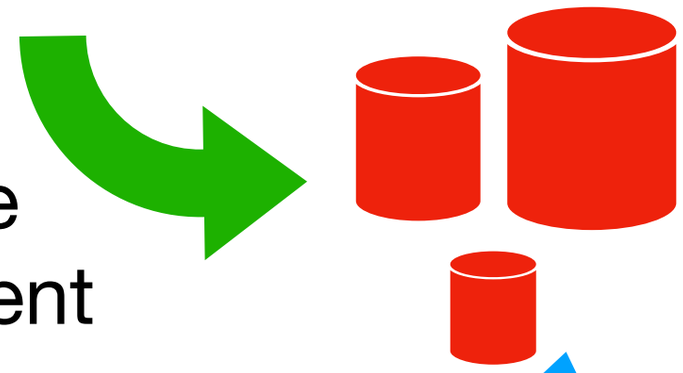
data science



Today's Empirical SE Process

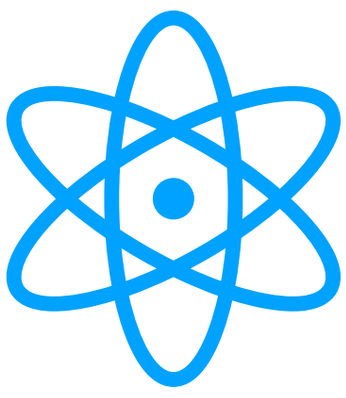


operate (case study, experiment or survey)



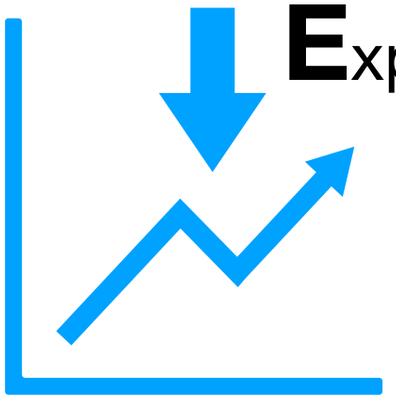
iterate

iNterpret



SEMN

Model



Explore

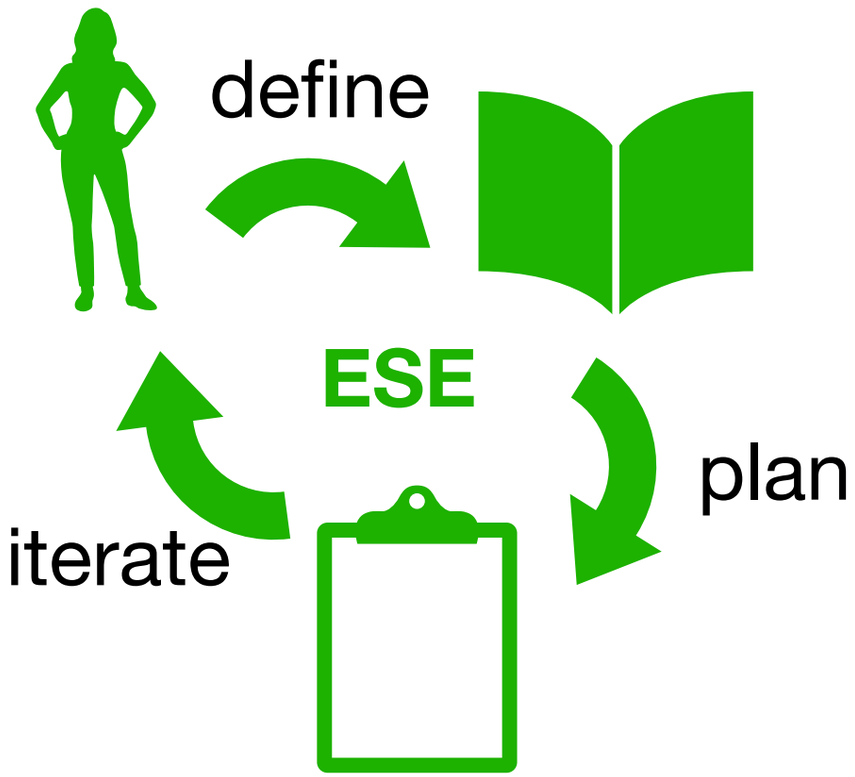
Scrub

empirical SE

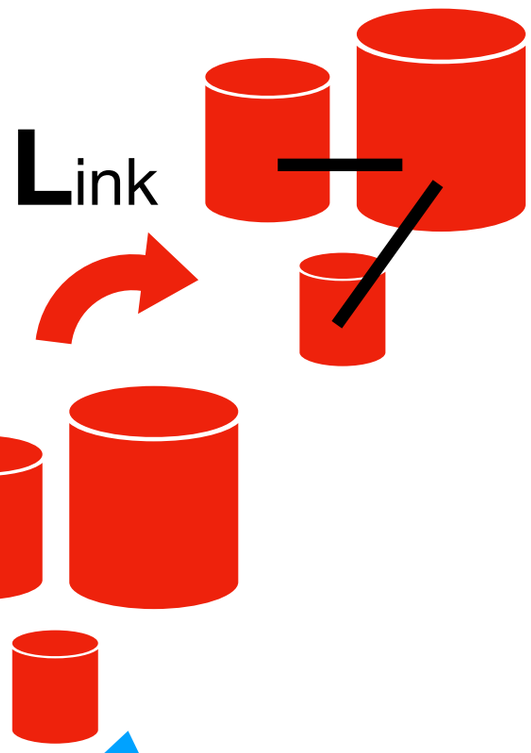
data science

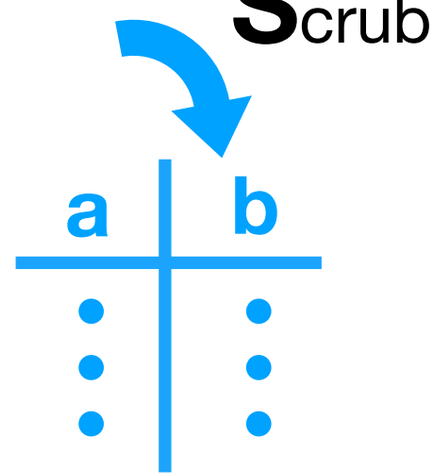
software analytics

Today's Empirical SE Process

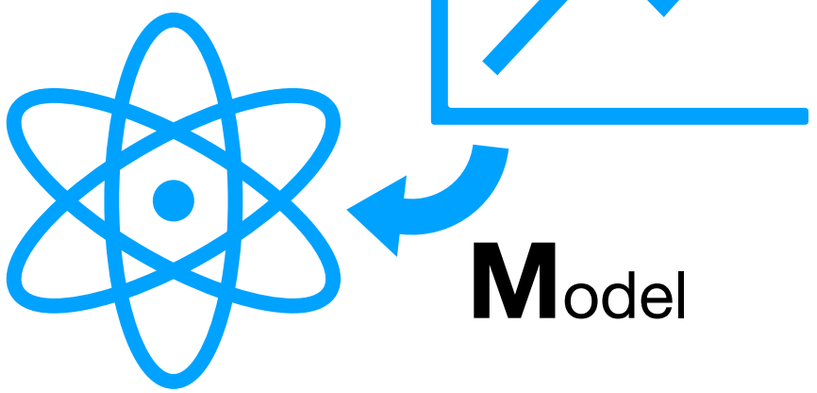


operate (case study, experiment or survey)





SEMN

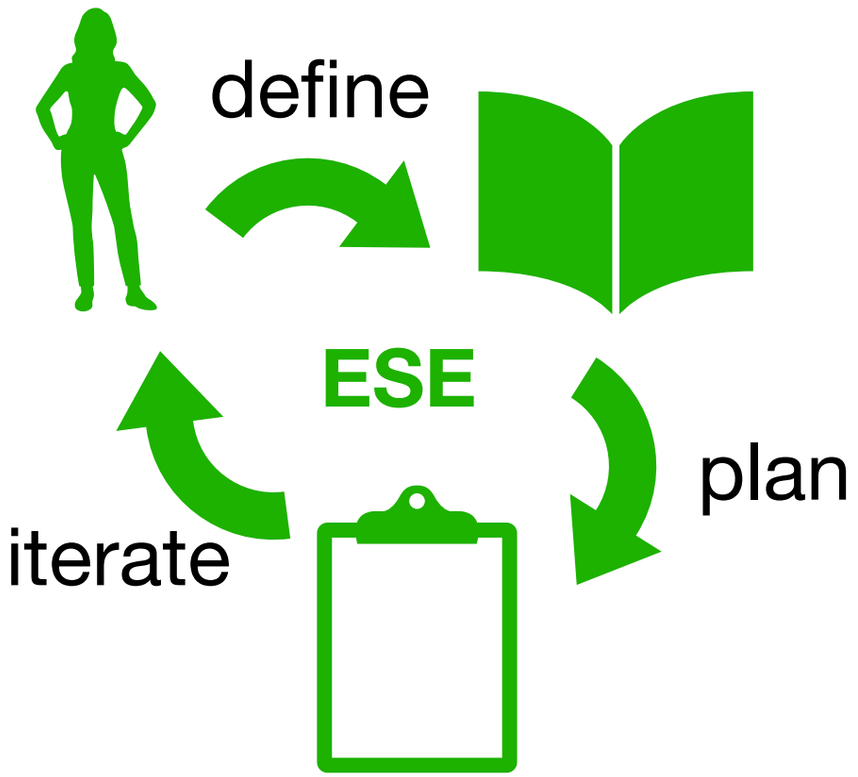


empirical SE

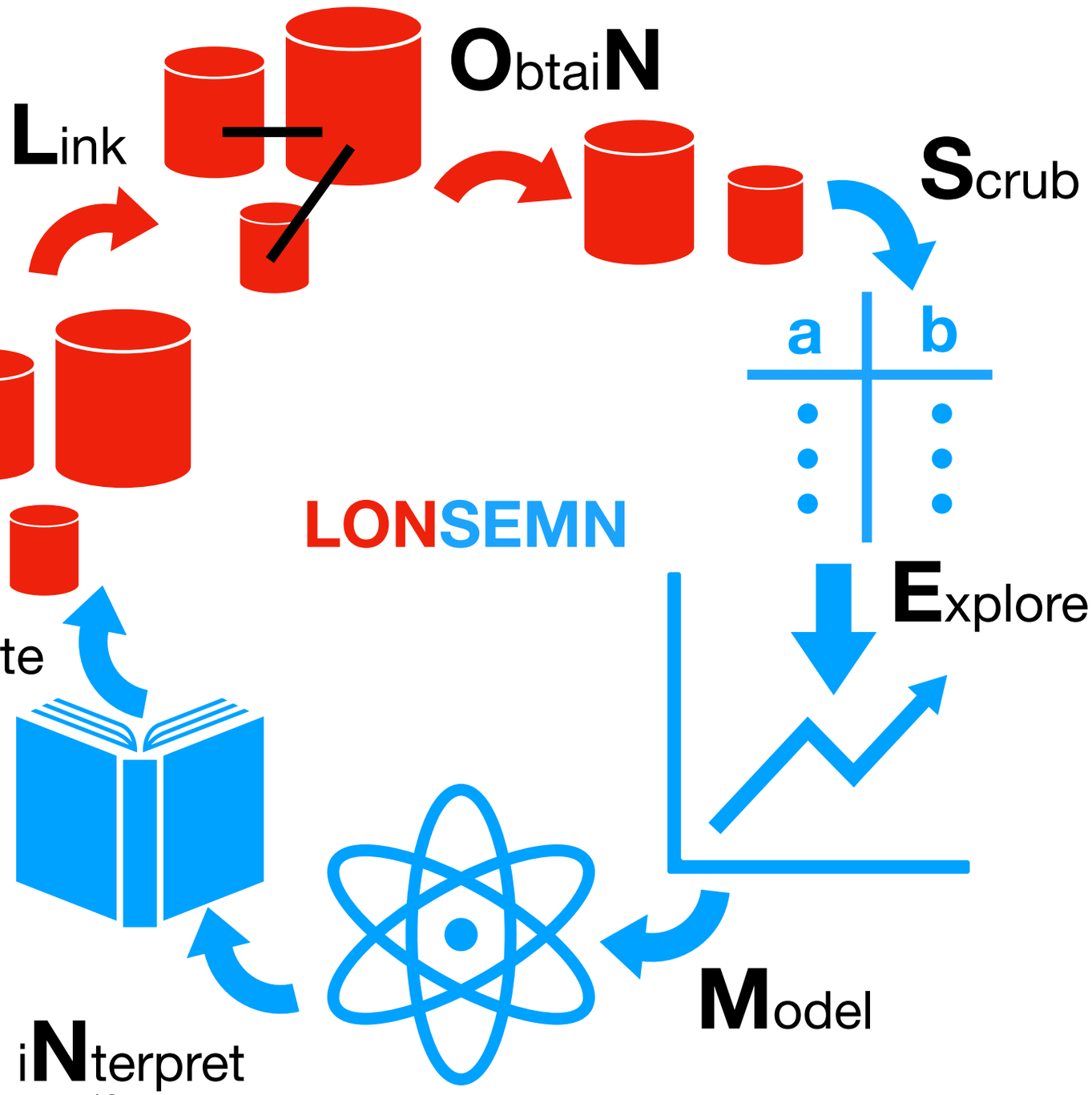
data science

software analytics

Today's Empirical SE Process



operate (case study, experiment or survey)



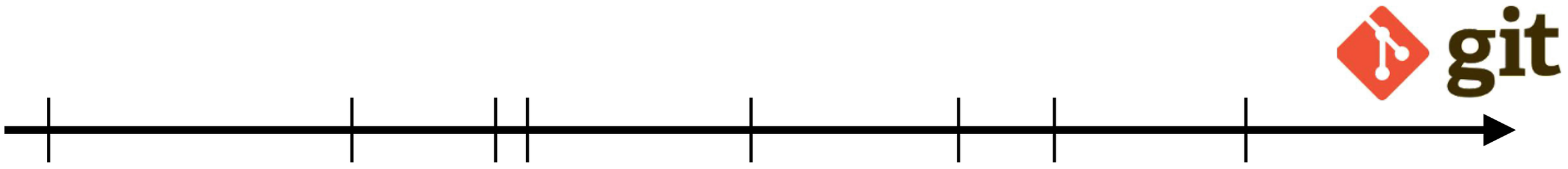
empirical SE

data science

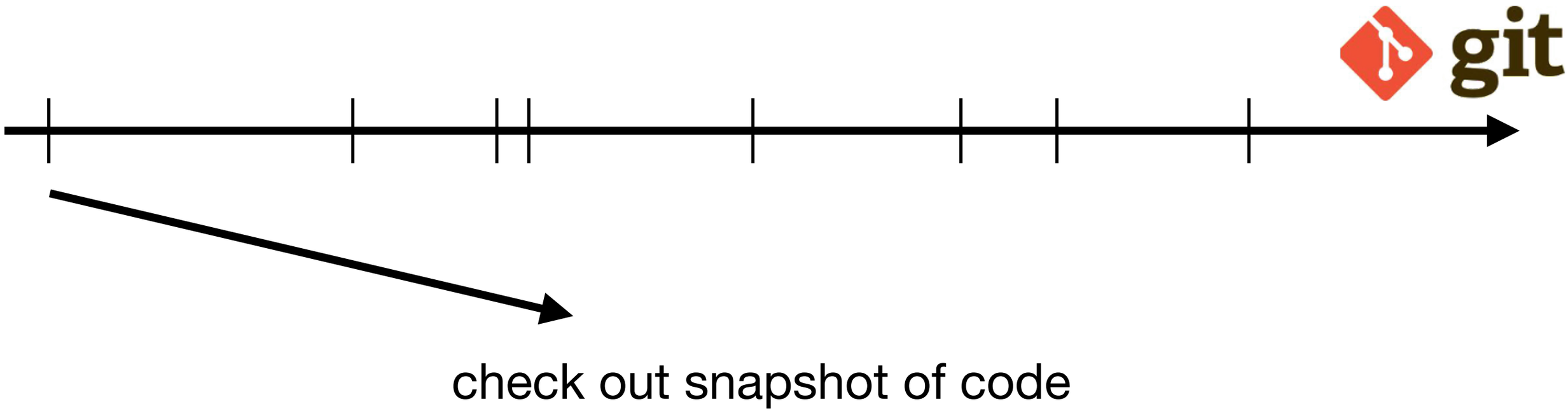
software analytics

ObtaiN Activity should be Time-aware!

ObtaiN Activity should be Time-aware!

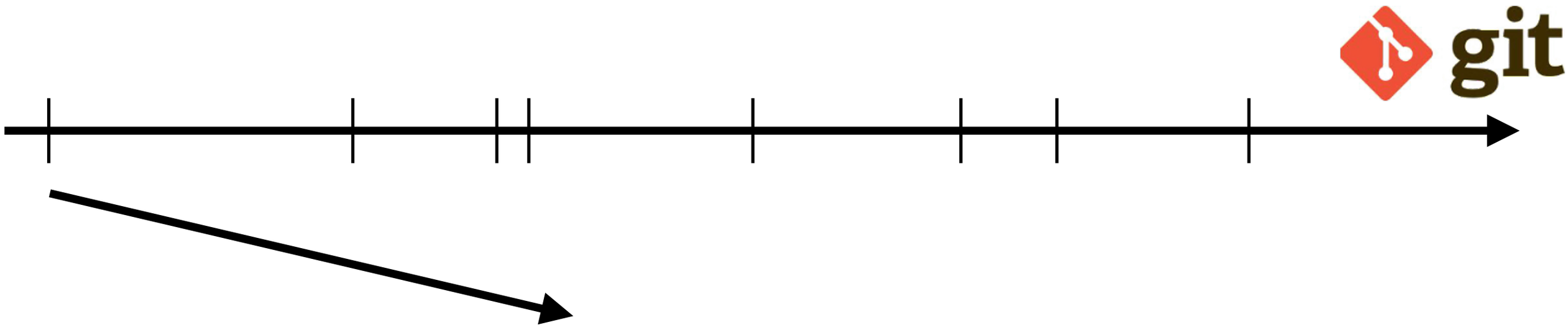


ObtaiN Activity should be Time-aware!



check out snapshot of code

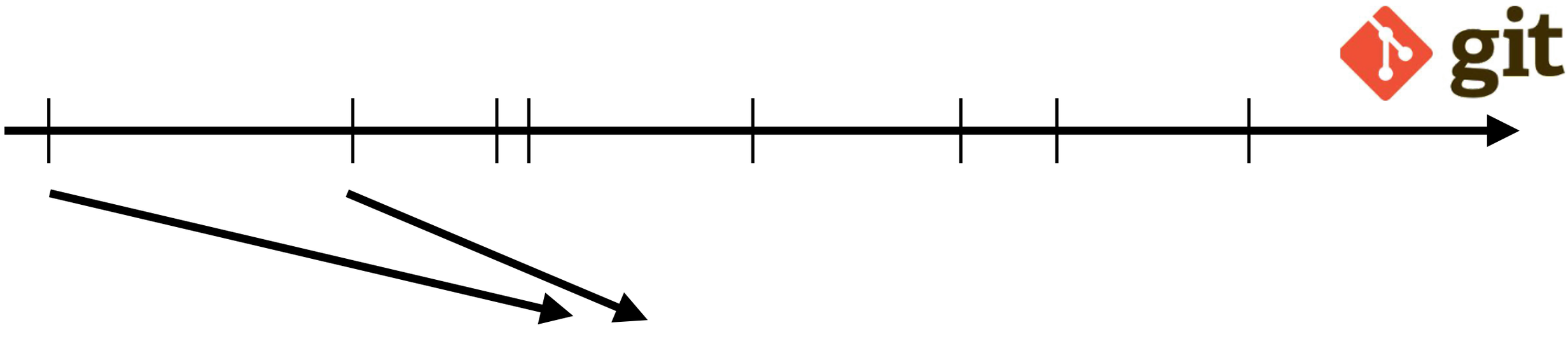
ObtaiN Activity should be Time-aware!



check out snapshot of code

apply time-**un**aware metric extraction

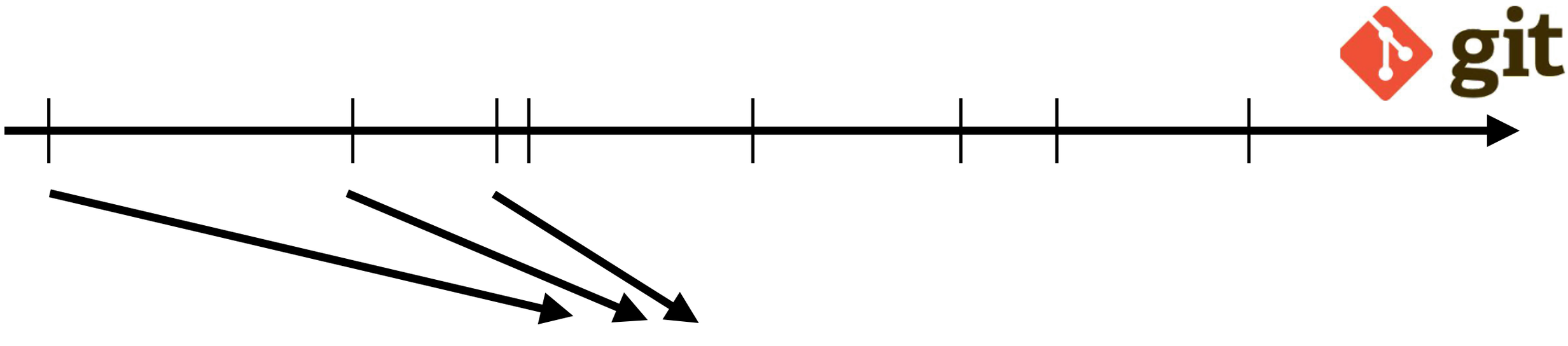
ObtaiN Activity should be Time-aware!



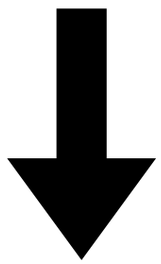
check out snapshot of code

apply time-**un**aware metric extraction

ObtaiN Activity should be Time-aware!

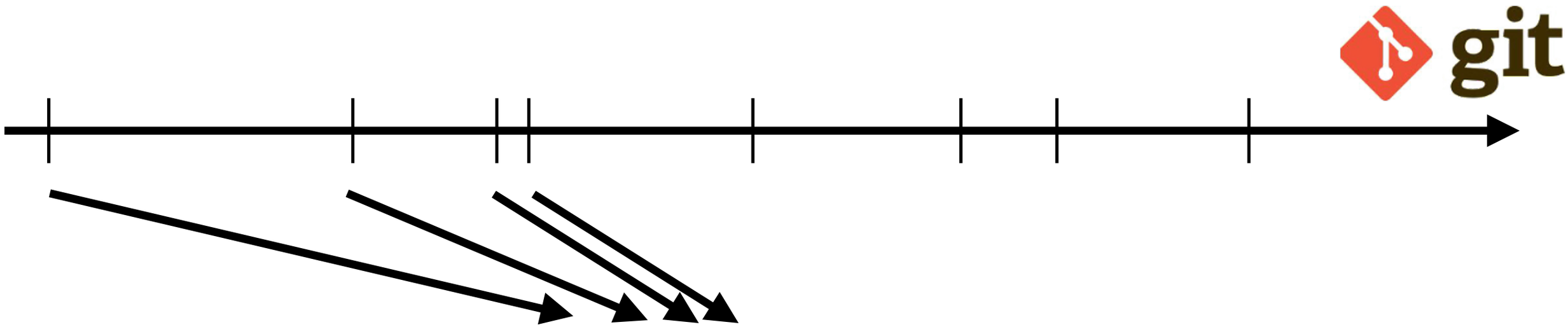


check out snapshot of code



apply time-**un**aware metric extraction

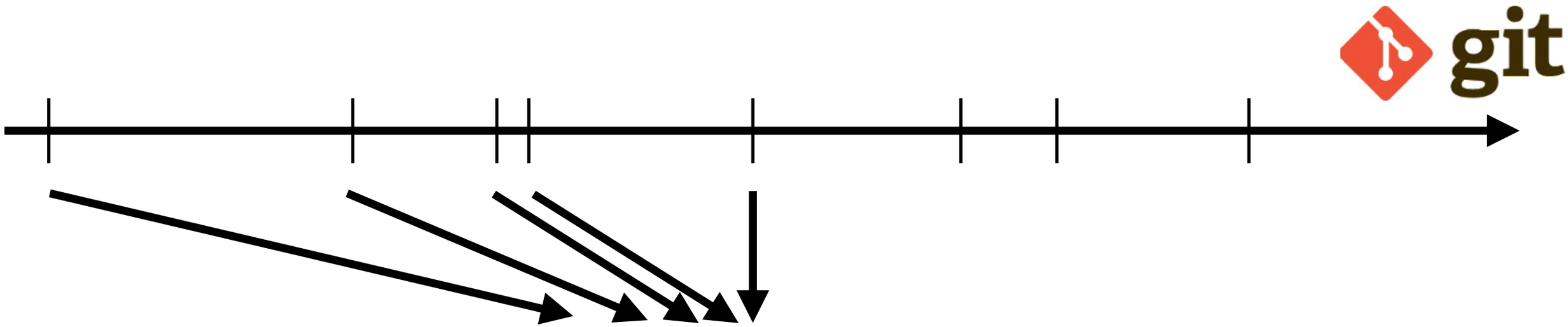
ObtaiN Activity should be Time-aware!



check out snapshot of code

apply time-**un**aware metric extraction

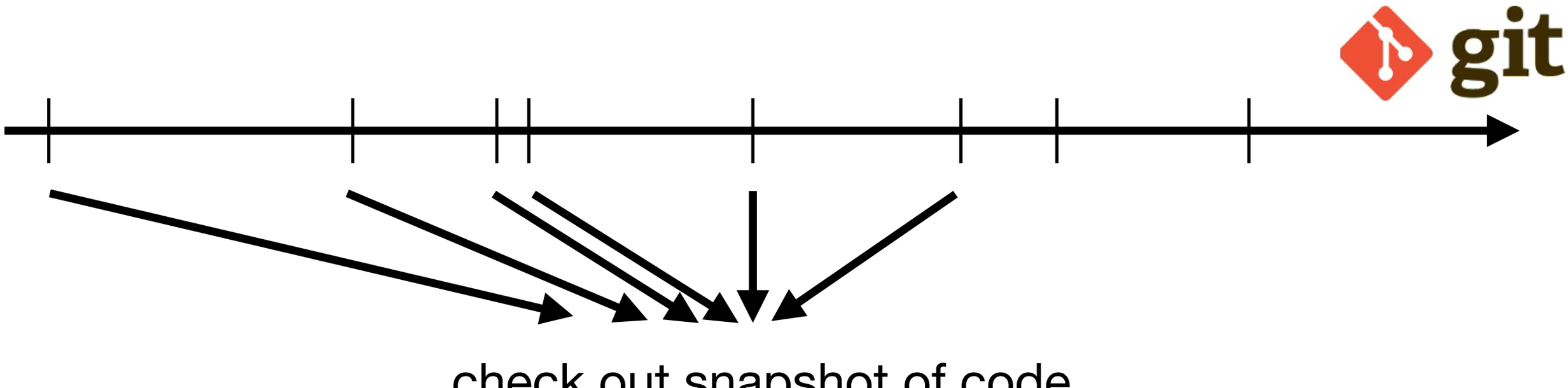
ObtaiN Activity should be Time-aware!



check out snapshot of code

apply time-**un**aware metric extraction

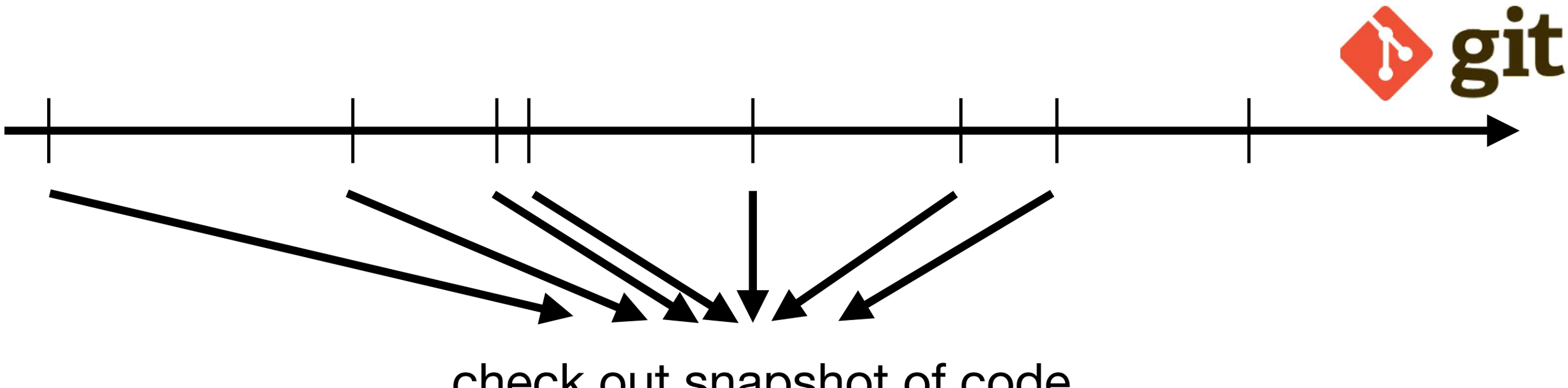
ObtaiN Activity should be Time-aware!



check out snapshot of code

apply time-**un**aware metric extraction

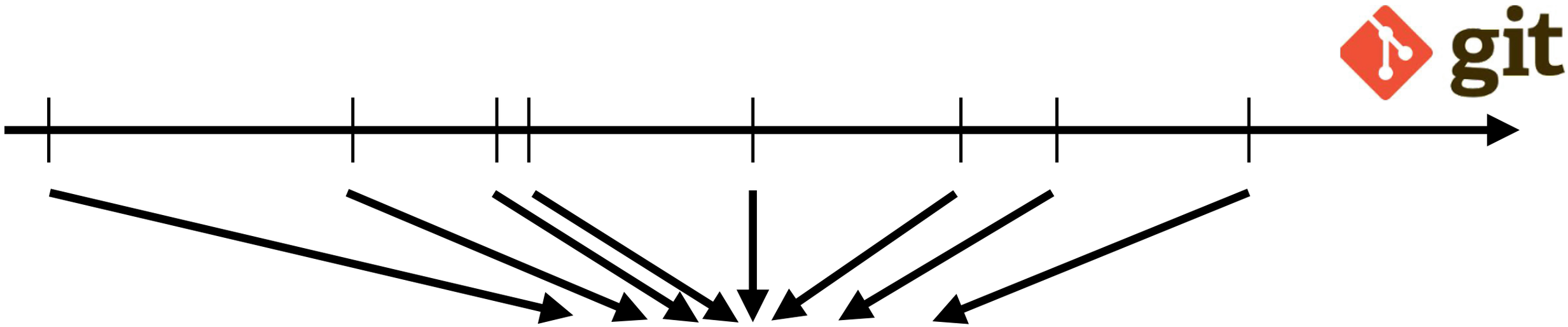
ObtaiN Activity should be Time-aware!



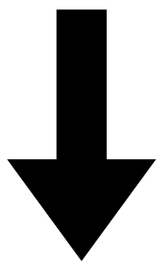
check out snapshot of code

apply time-**un**aware metric extraction

ObtaiN Activity should be Time-aware!

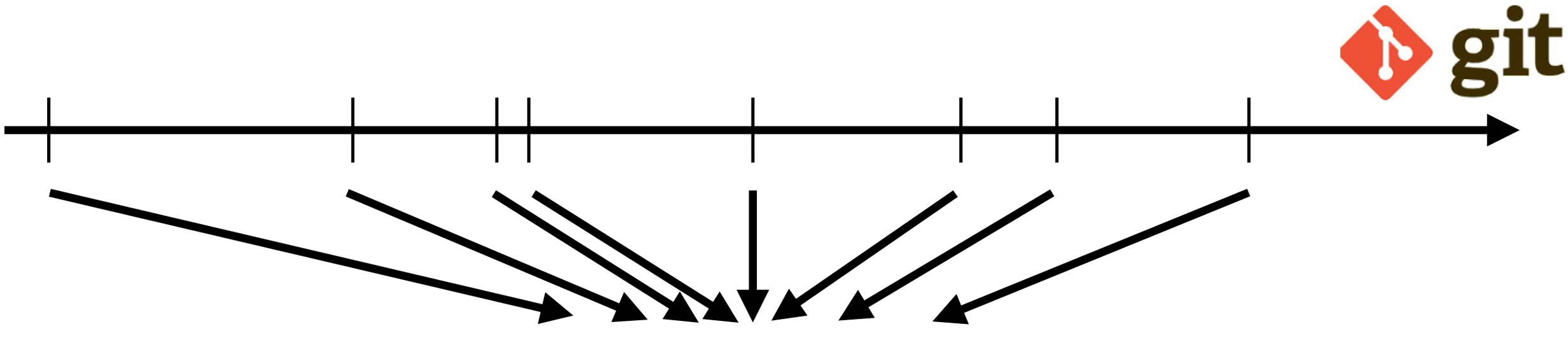


check out snapshot of code



apply time-**un**aware metric extraction

ObtaiN Activity should be Time-aware!

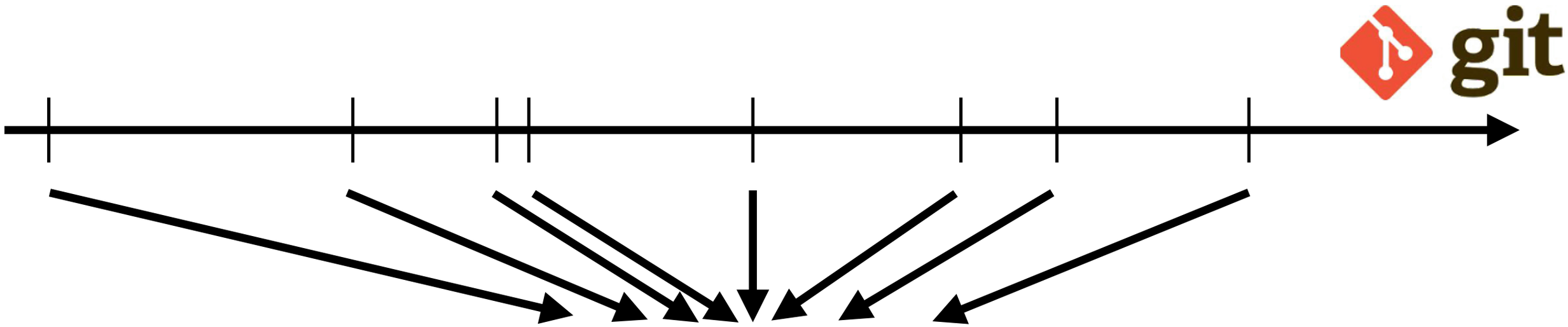


check out snapshot of code

what if snapshot does not compile?

apply time-**un**aware metric extraction

ObtaiN Activity should be Time-aware!



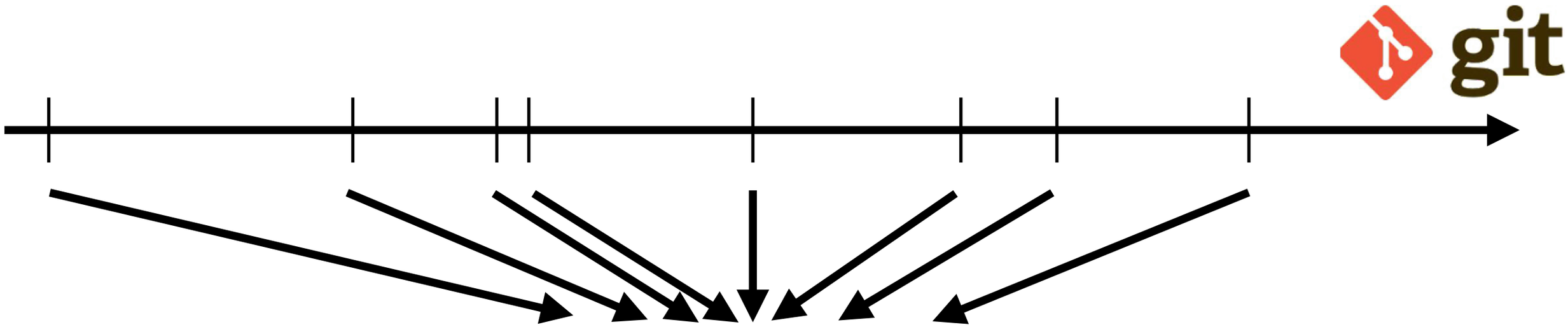
check out snapshot of code

what if snapshot does not compile?

what if analysis is time-intensive?

apply time-**un**aware metric extraction

ObtaiN Activity should be Time-aware!



check out snapshot of code

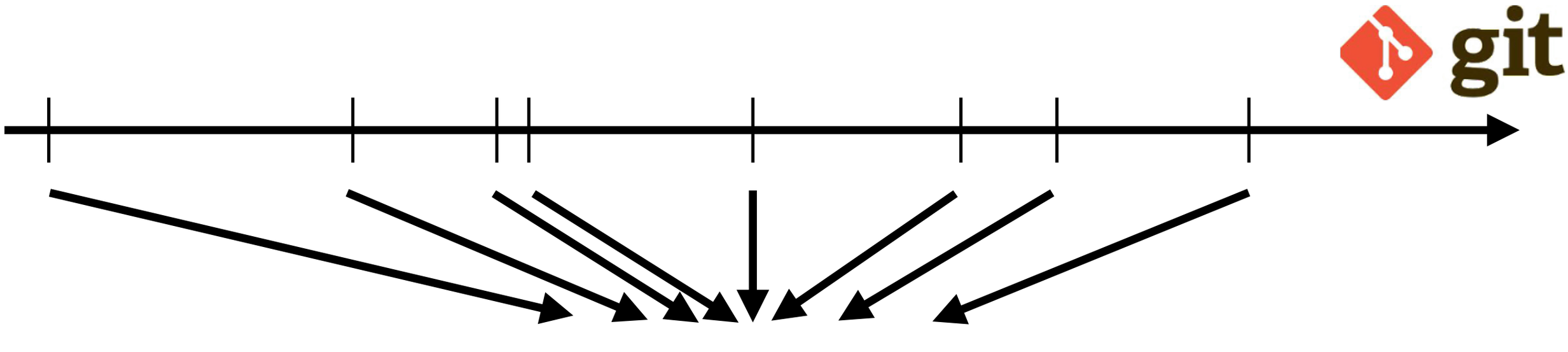
what if snapshot does not compile?

what if more than one branch?

what if analysis is time-intensive?

apply time-**un**aware metric extraction

ObtaiN Activity should be Time-aware!



check out snapshot of code

what if snapshot does not compile?

what if more than one branch?

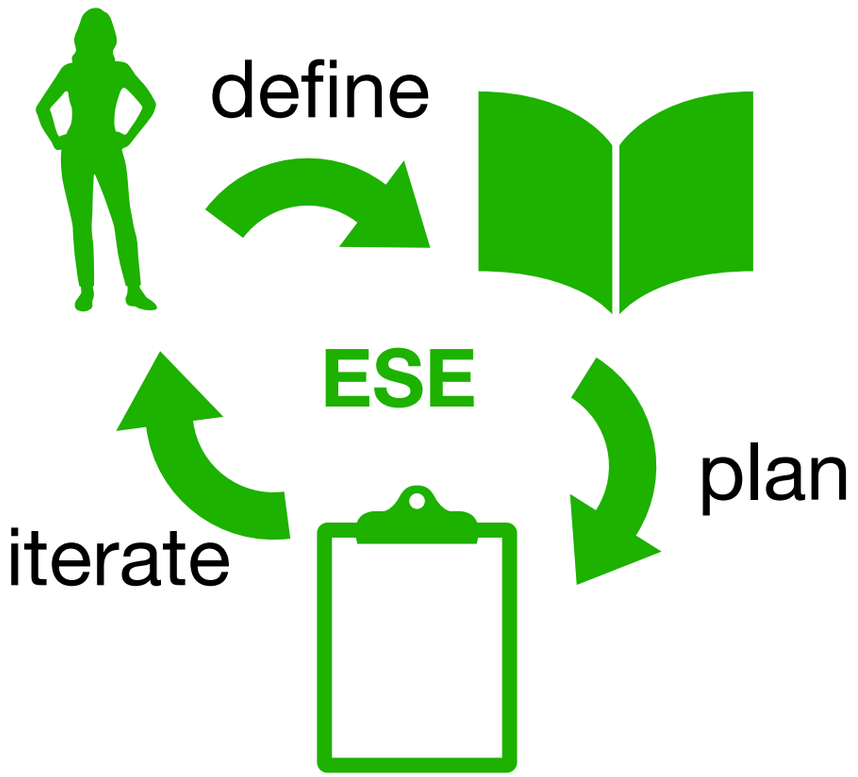
what if analysis is time-intensive?

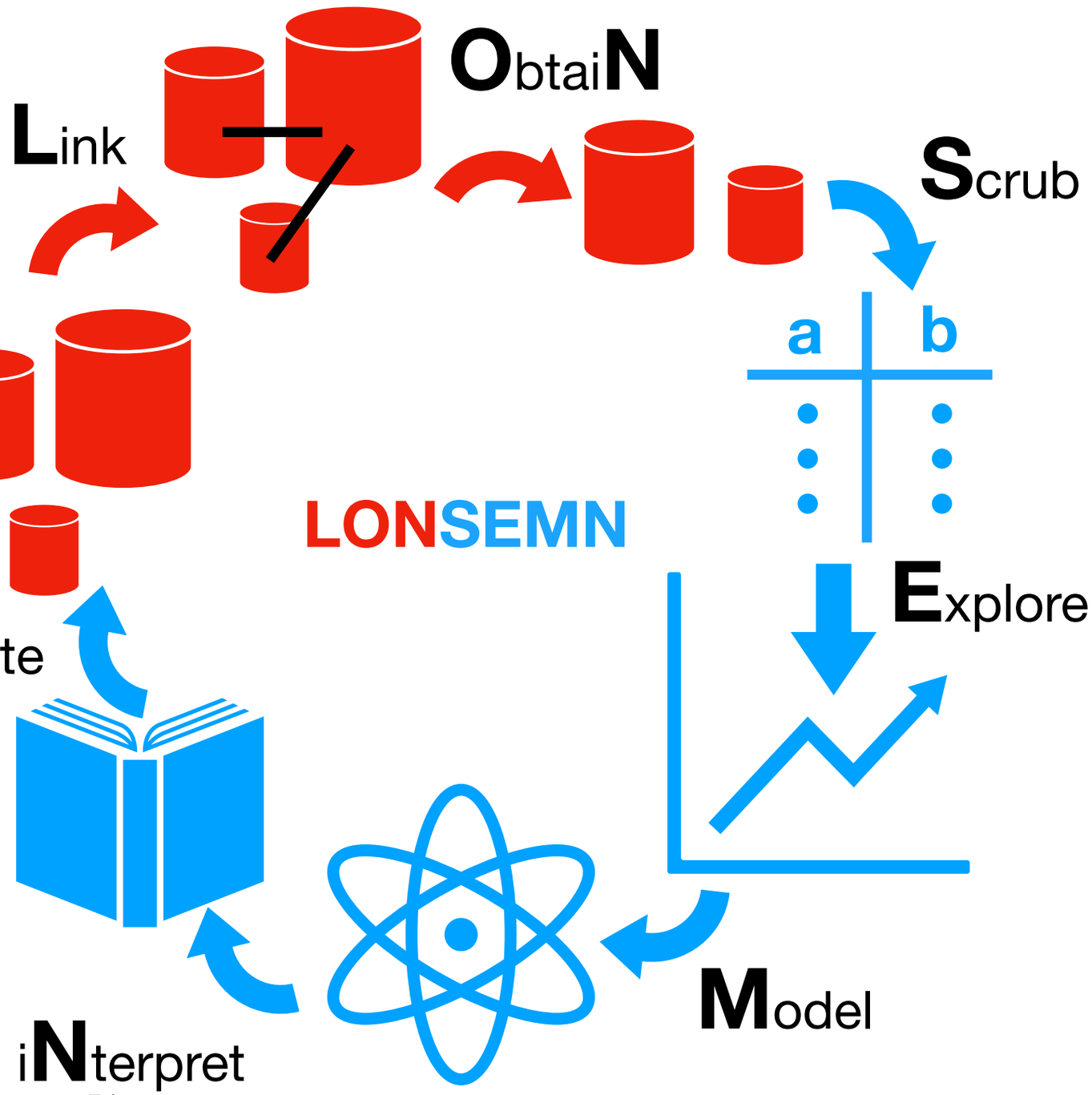
apply time-**un**aware metric extraction

what if results of prior snapshots needed?

Exercise: Predicting Bug-introducing Commits (1)

Today's Empirical SE Process





empirical SE

data science

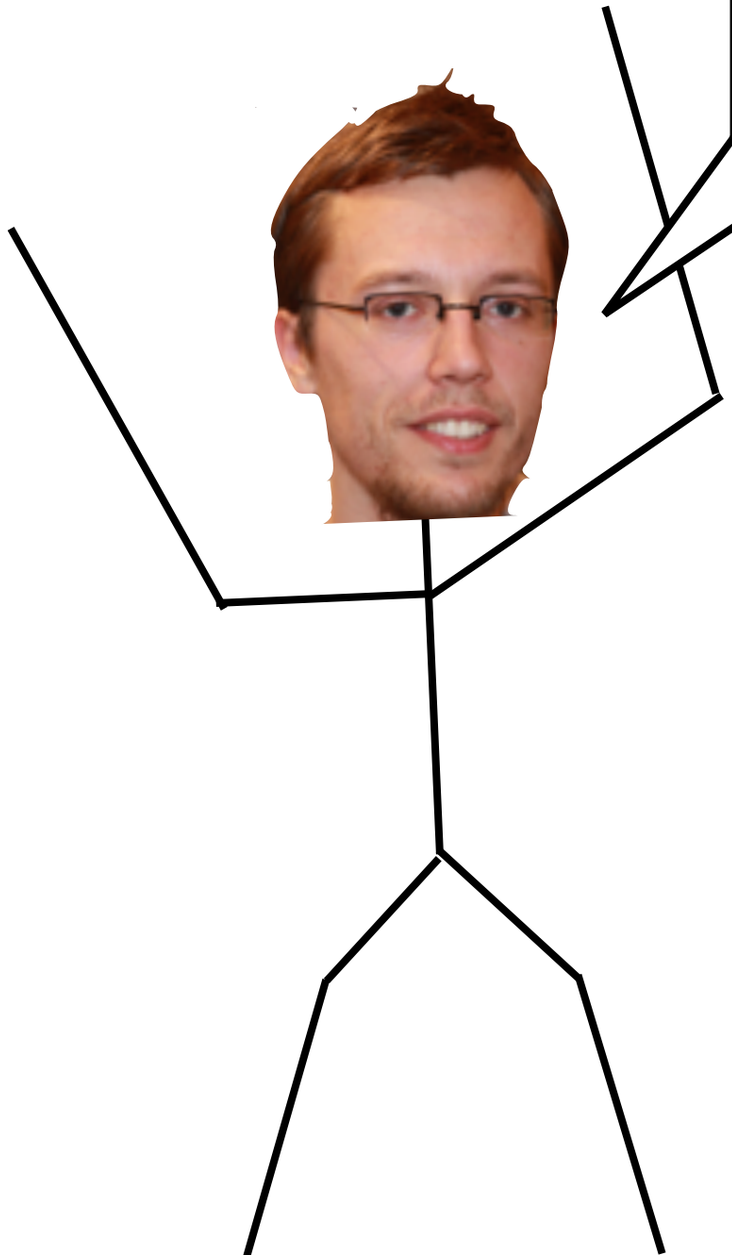
software analytics

Part IV: Now What?

The 3 concepts are in fact complementary!

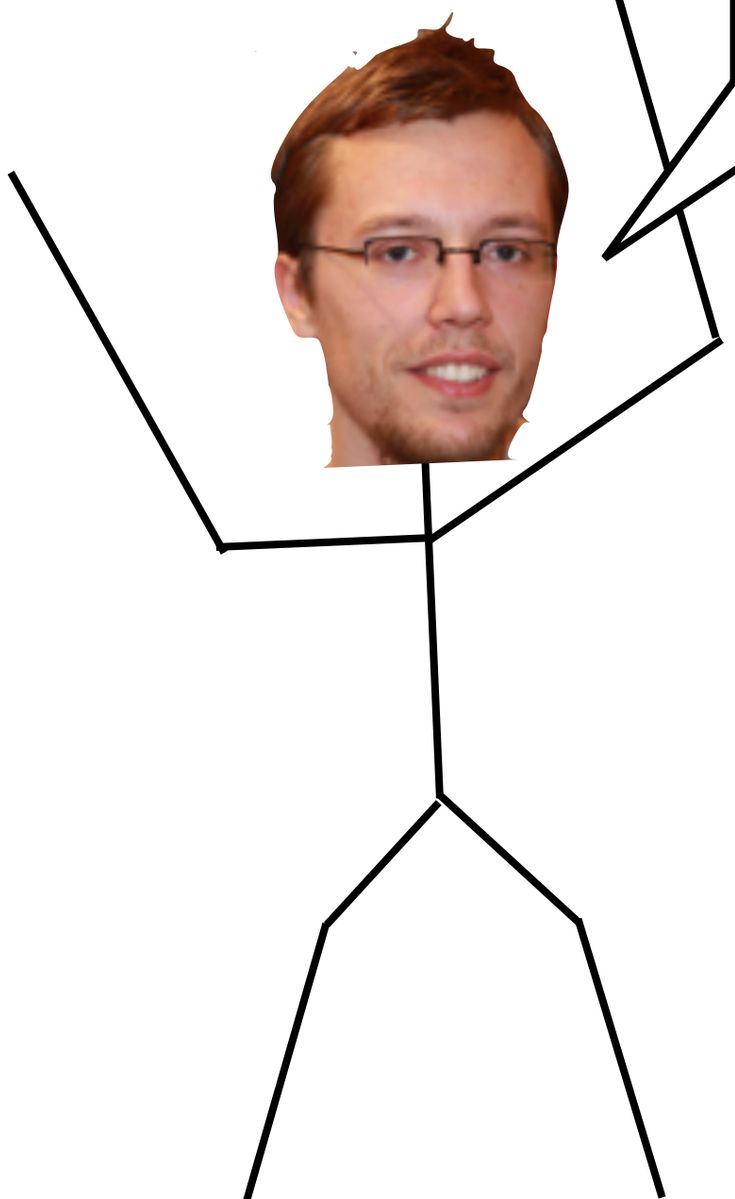
The 3 concepts are in fact complementary!

… so can you just **submit your work to venues of either domain?**



Yes, you can!

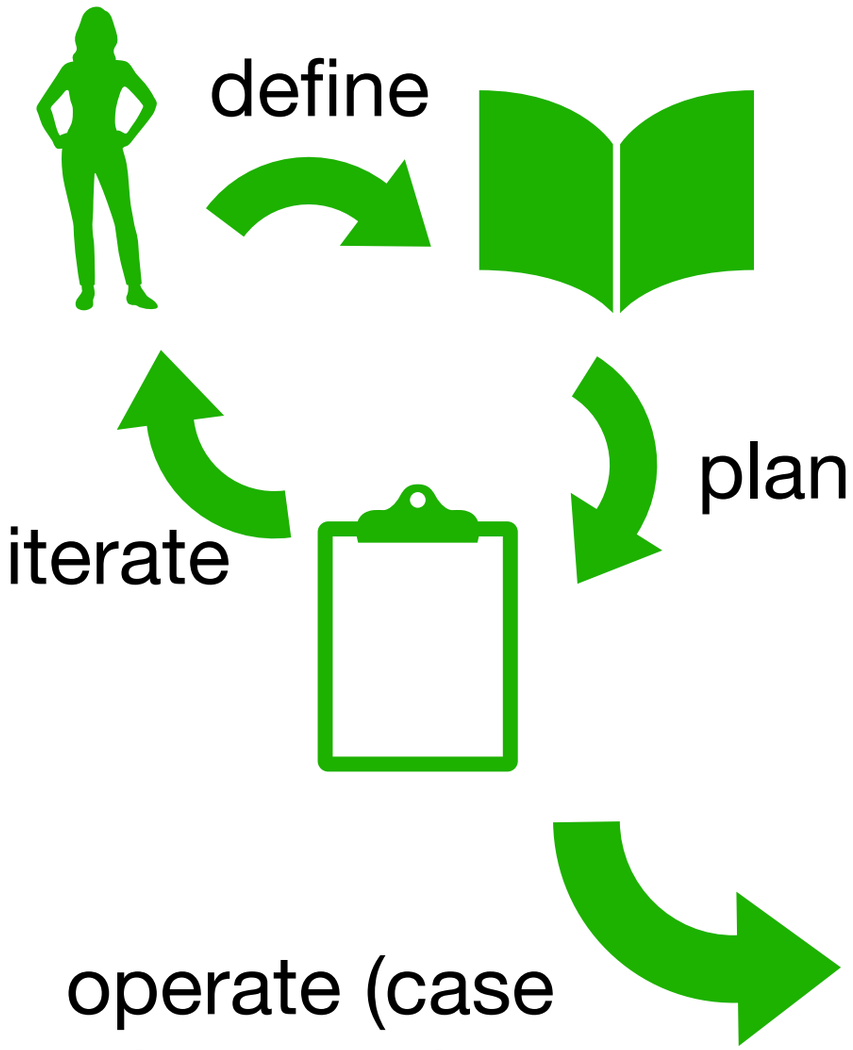
(Depending on the **contributions** you want to stress)



Yes, you can!

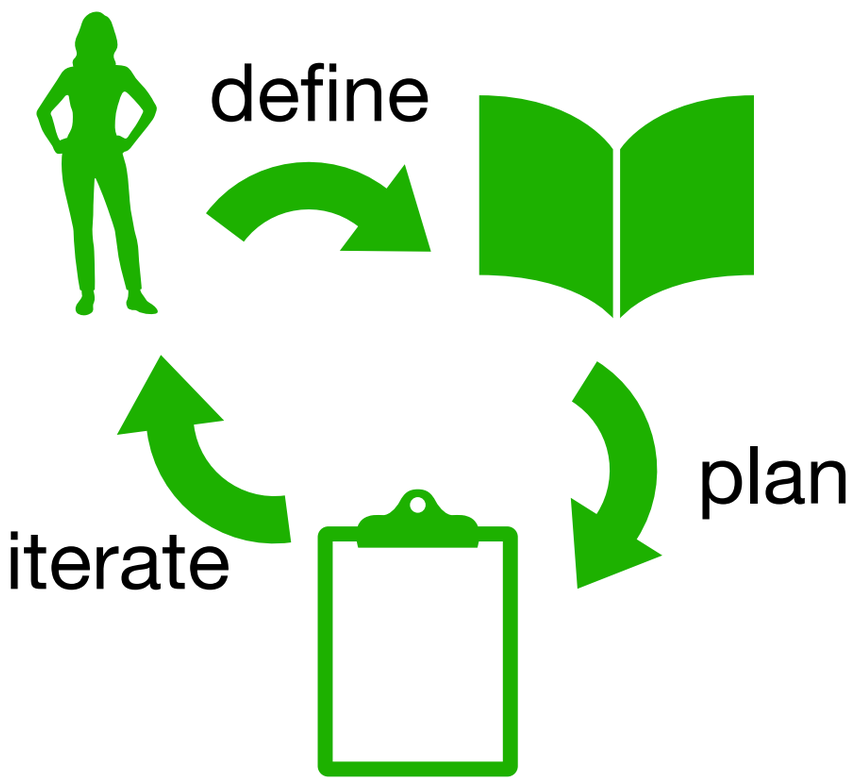
(Depending on the **contributions** you want to stress)

NOTE: “stress” != “include” (i.e., all activities will be included, but with different importance)



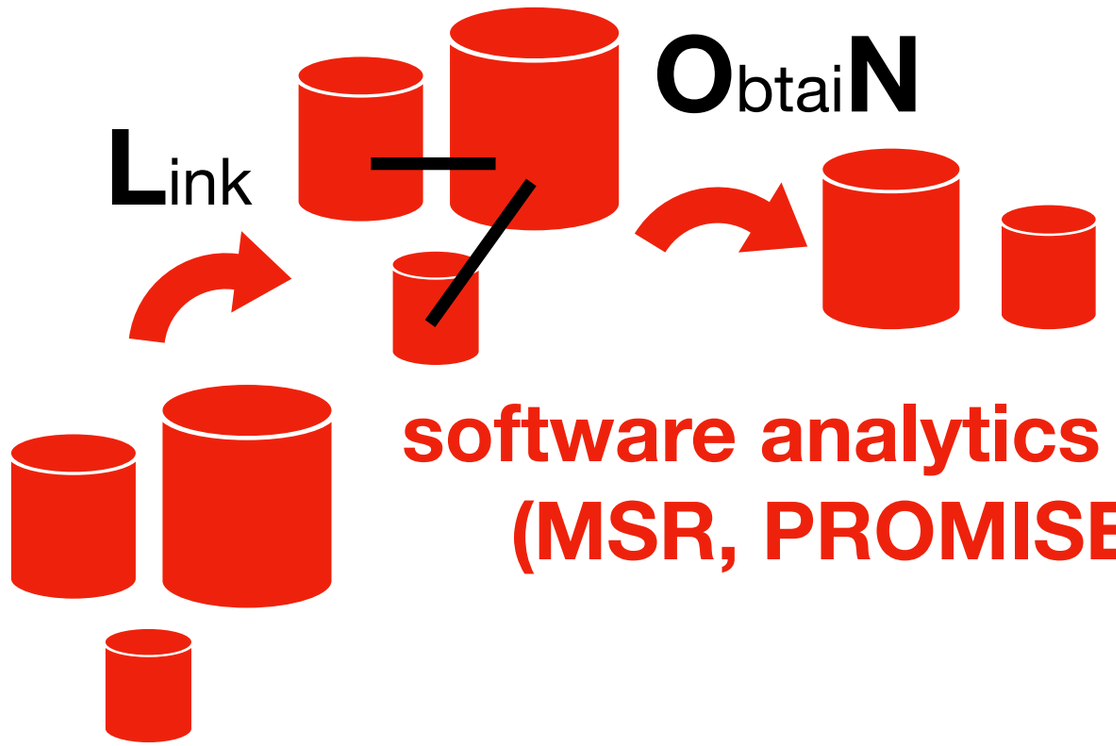
operate (case study, experiment or survey)

empirical venues (ESEM, EMSE, …)



operate (case study, experiment or survey)

empirical venues (ESEM, EMSE, …)



software analytics venues (MSR, PROMISE, …)

define

plan

iterate

operate (case study, experiment or survey)

empirical venues (ESEM, EMSE, …)

Link

Obtai**N**

software analytics venues (MSR, PROMISE, …)

Scrub

a

b

Explore

Model

i**N**terpret

iterate

data science/general SE venues (KDD, ICDM, ICSE, TSE, …)

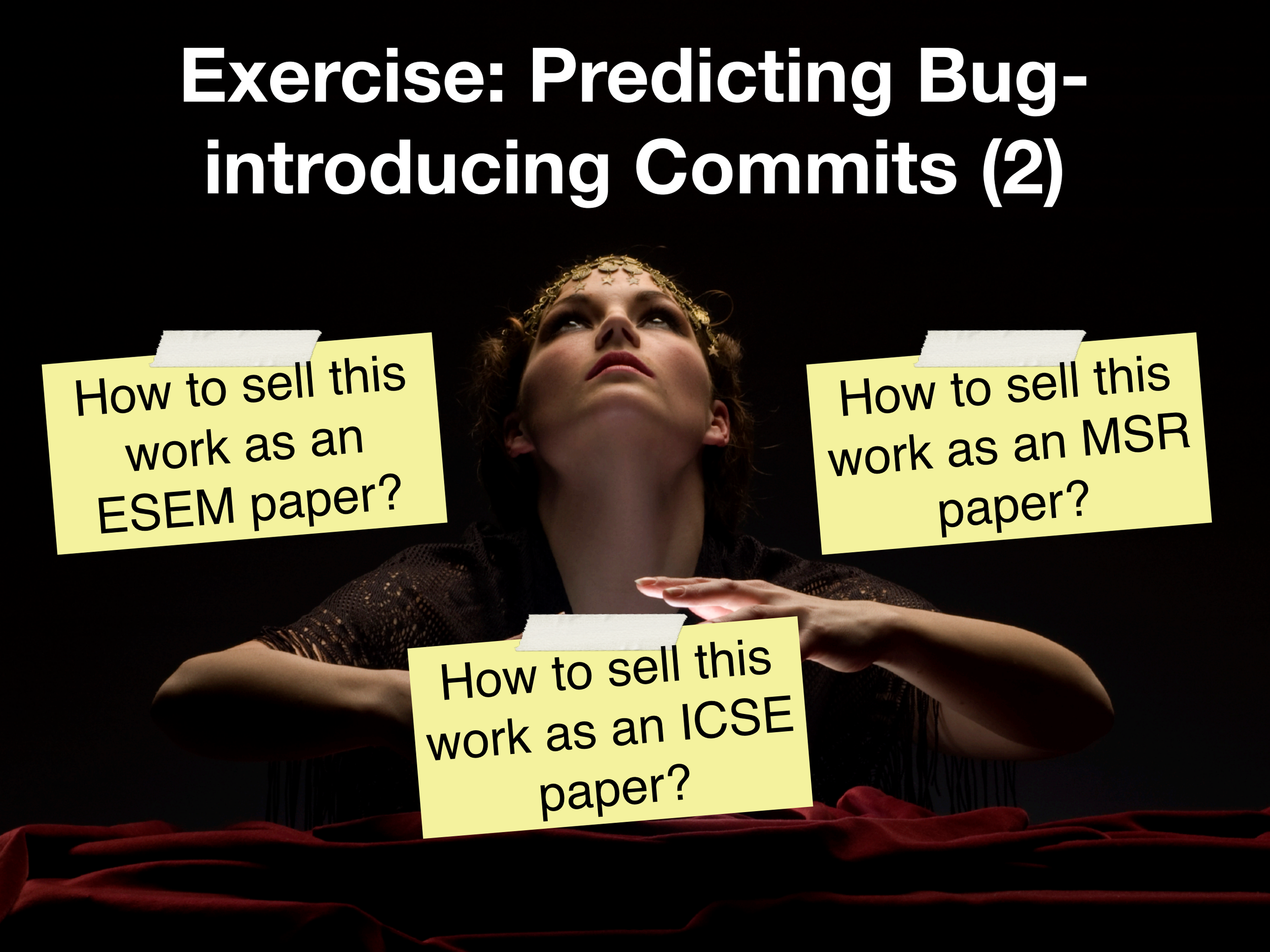
Exercise: Predicting Bug-introducing Commits (2)

Exercise: Predicting Bug-introducing Commits (2)

How to sell this work as an ESEM paper?

Exercise: Predicting Bug-introducing Commits (2)

Exercise: Predicting Bug-introducing Commits (2)



define

plan

iterate

operate (case study, experiment or survey)

empirical venues (ESEM, EMSE, …)

Link

Obtai**N**

software analytics venues (MSR, PROMISE, …)

Scrub

a

b

Explore

Model

i**N**terpret

iterate

data science/general SE venues (KDD, ICDM, ICSE, TSE, …)

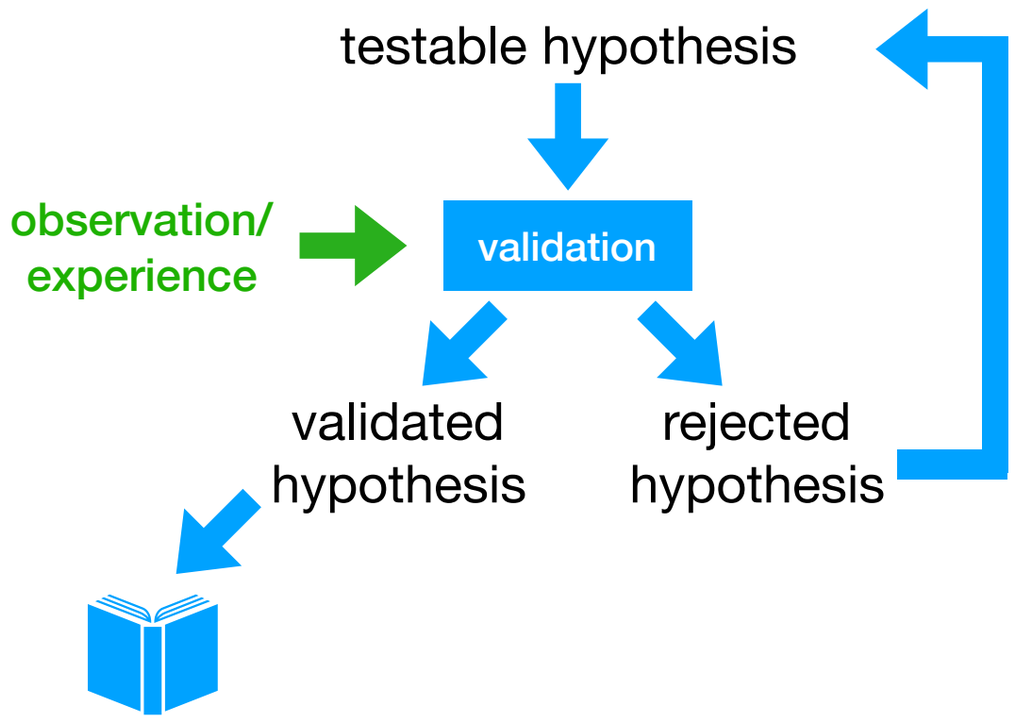
How Will Tomorrow's ESE Process Look Like?

hard to predict …

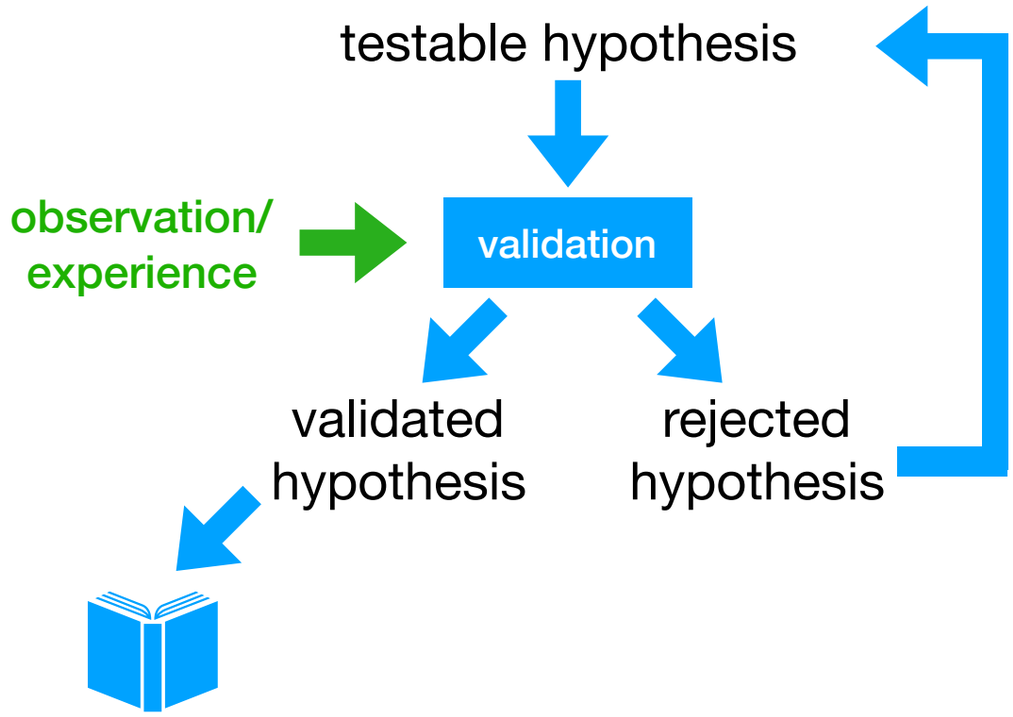
hard to predict …

… but very likely an evolution of today’s process!

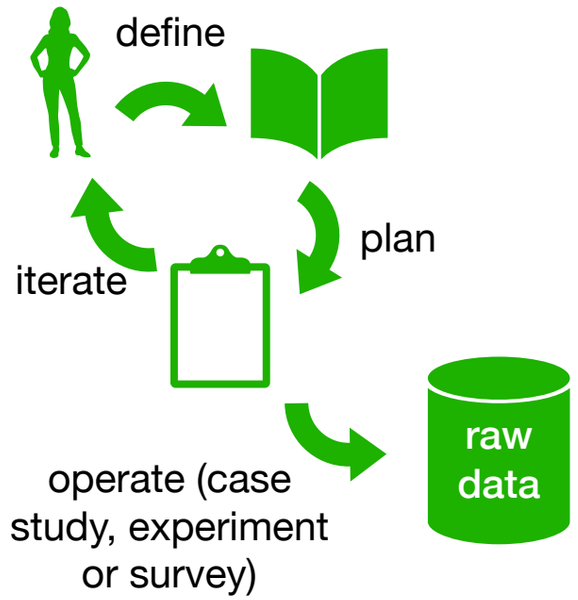
Empirical Science?

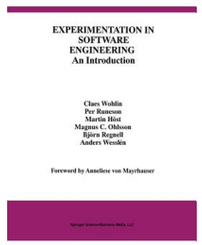


Empirical Science?



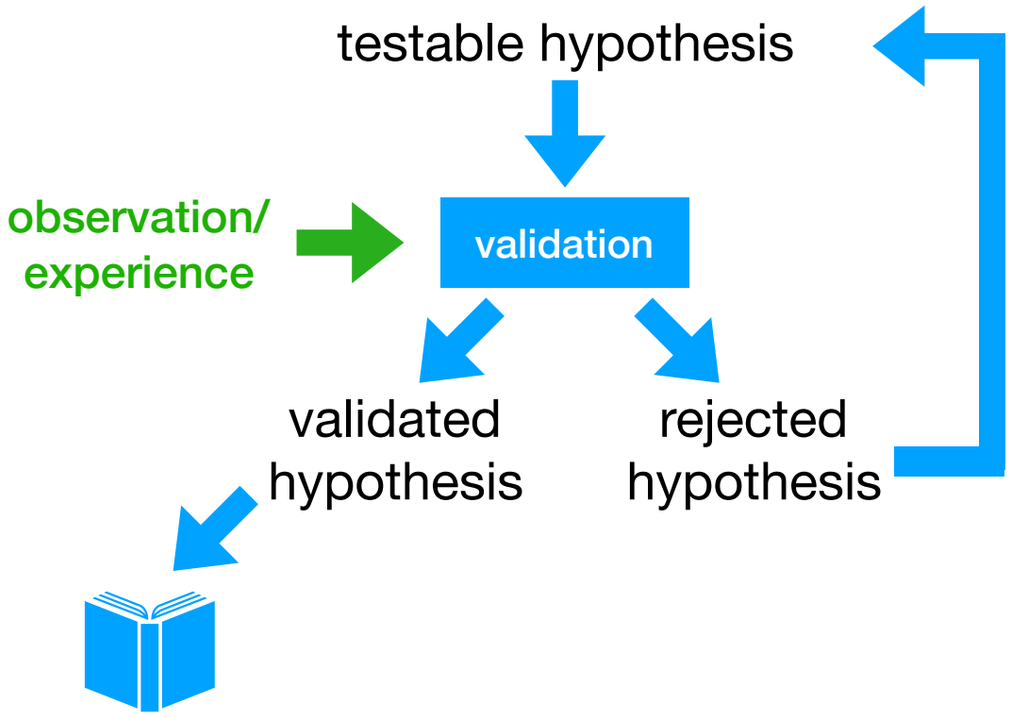
Traditional Empirical SE Process



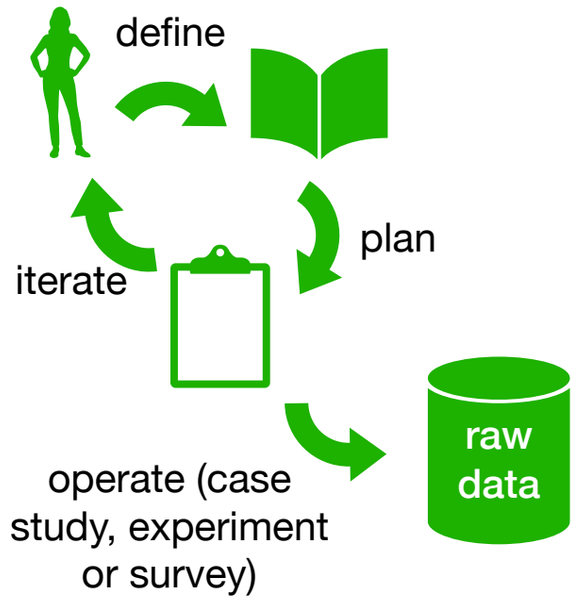


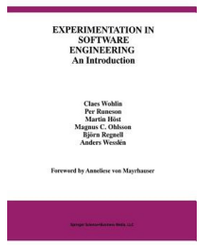
Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

Empirical Science?

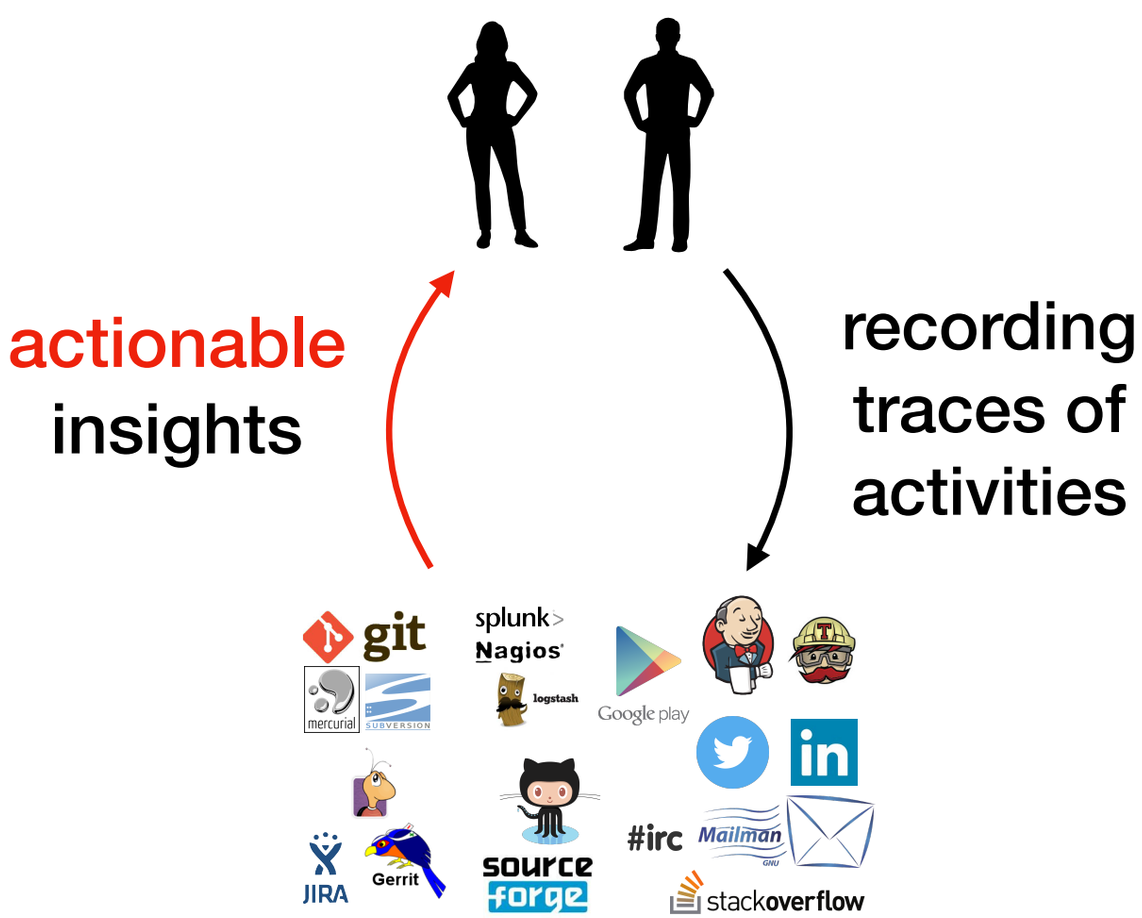


Traditional Empirical SE Process

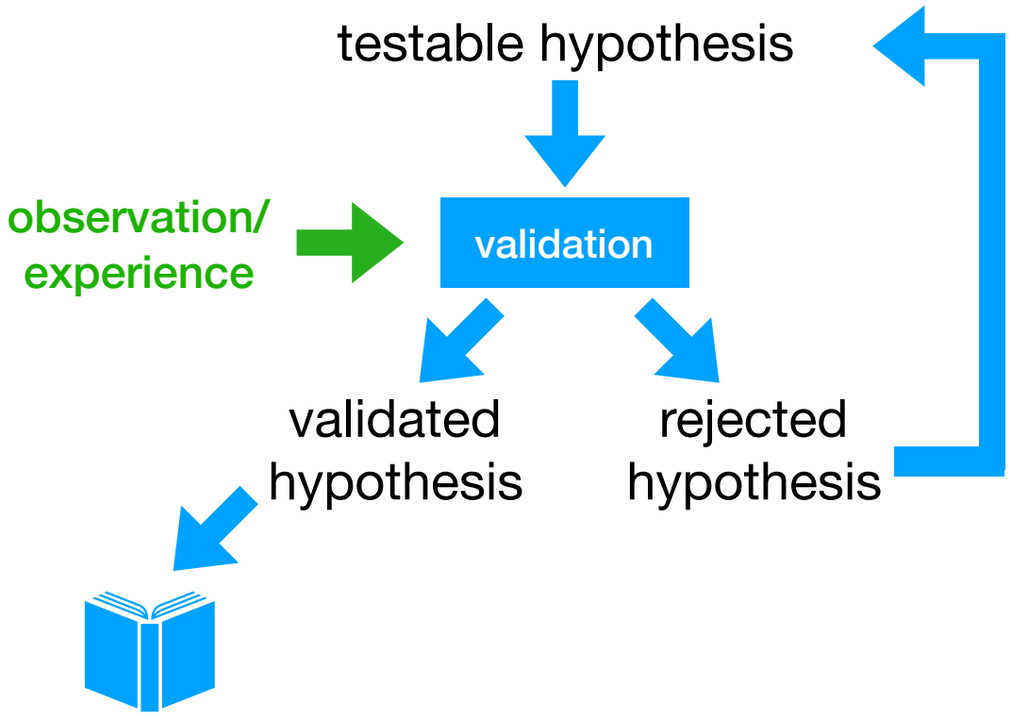




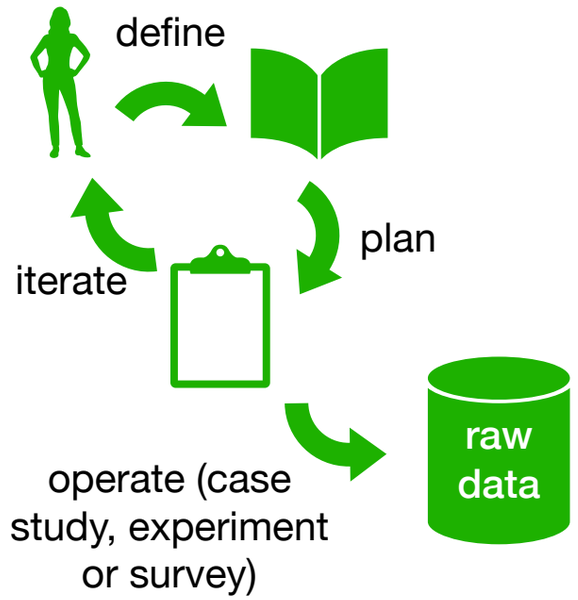
Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

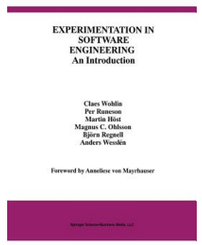


Empirical Science?



Traditional Empirical SE Process





Claes Wohlin et al., “Experimentation in Software Engineering - An Introduction”, 2000.

Today’s Empirical SE Process