

Survey Research in Software Engineering

Authors

Marcos Kalinowski (Pontifical Catholic University of Rio de Janeiro)

Allysson Alex Araújo (Federal University of Cariri)

Daniel Mendez (Blekinge Institute of Technology, fortiss)

This tutorial is based on...

- Previous editions given together with



Marco Torchiano
Politecnico di Torino



Guilherme H. Travassos
COPPE,
University of Rio de Janeiro



Rafael M. de Mello
PUC Rio de
Janeiro

- Experiences made in large survey research projects, e.g.



Introduction - Who are you?

Quick round...

- Who are you?
- What are your experiences in conducting survey research?
- Are you currently facing any particular challenges?
- What are your expectations?



What do you think?

**Why do we need survey research
in software engineering?**

This session will be about

Scope

- Brief introduction into survey research and epistemological setting
- Experiences and lessons learnt (Best Practices)
- Discussion / Hands-on session

Out of scope:

- Statistics
- (In-depth) Fundamentals
- Statistics (this is definitively not about statistics)

Ground rule

Whenever you have questions / remarks,
don't ask Google , but
share them with the whole group.

Outline



- A brief introduction into survey research

40'-60'

- (Selected) Best practices

40'-60'

- Lean coffee

0' - 40'



Outline

- A brief introduction into survey research
- (Selected) Best practices
- Lean coffee

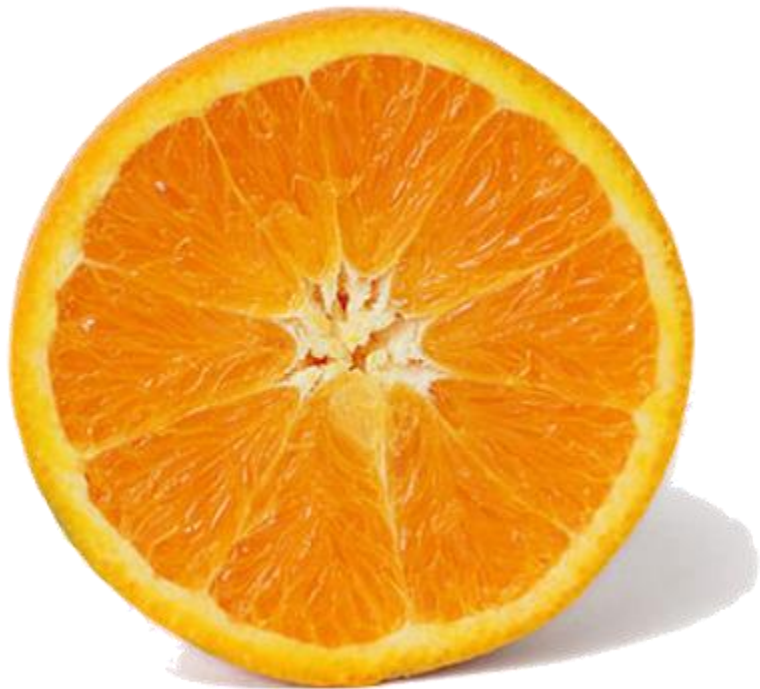
What is survey research?



Systematic observational method to gather qualitative and / or quantitative data from (a sample of) entities to characterise information, attitudes and / or behaviours from different groups of subjects regarding an object of study

- » Observational data with limited control
- » Descriptive and analytic statistics

Observational studies



Surveys
(Cross-sectional)



Case
Studies



Experiments
(Case-control)

Epistemological setting (very briefly)

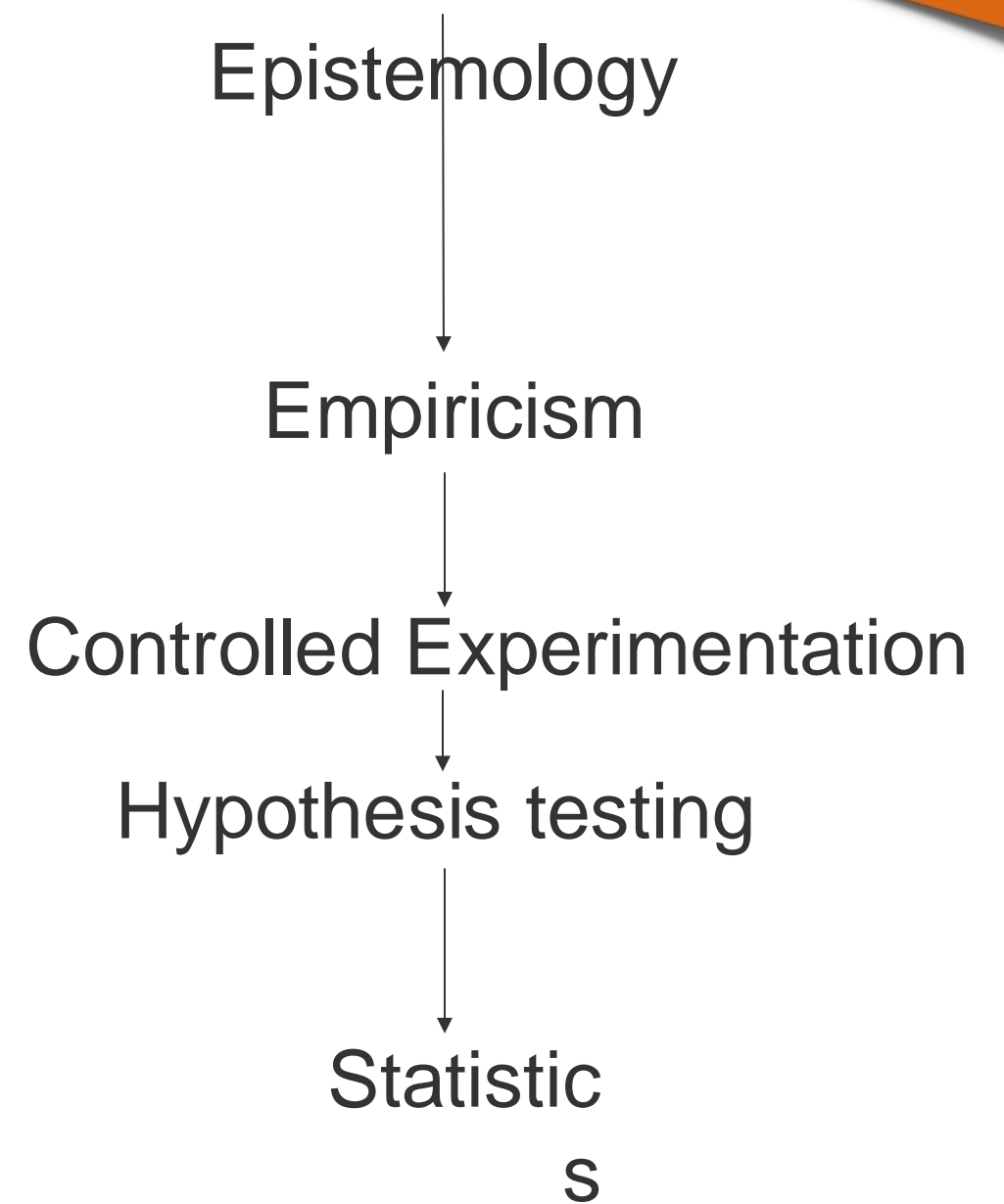


Philosophy of science

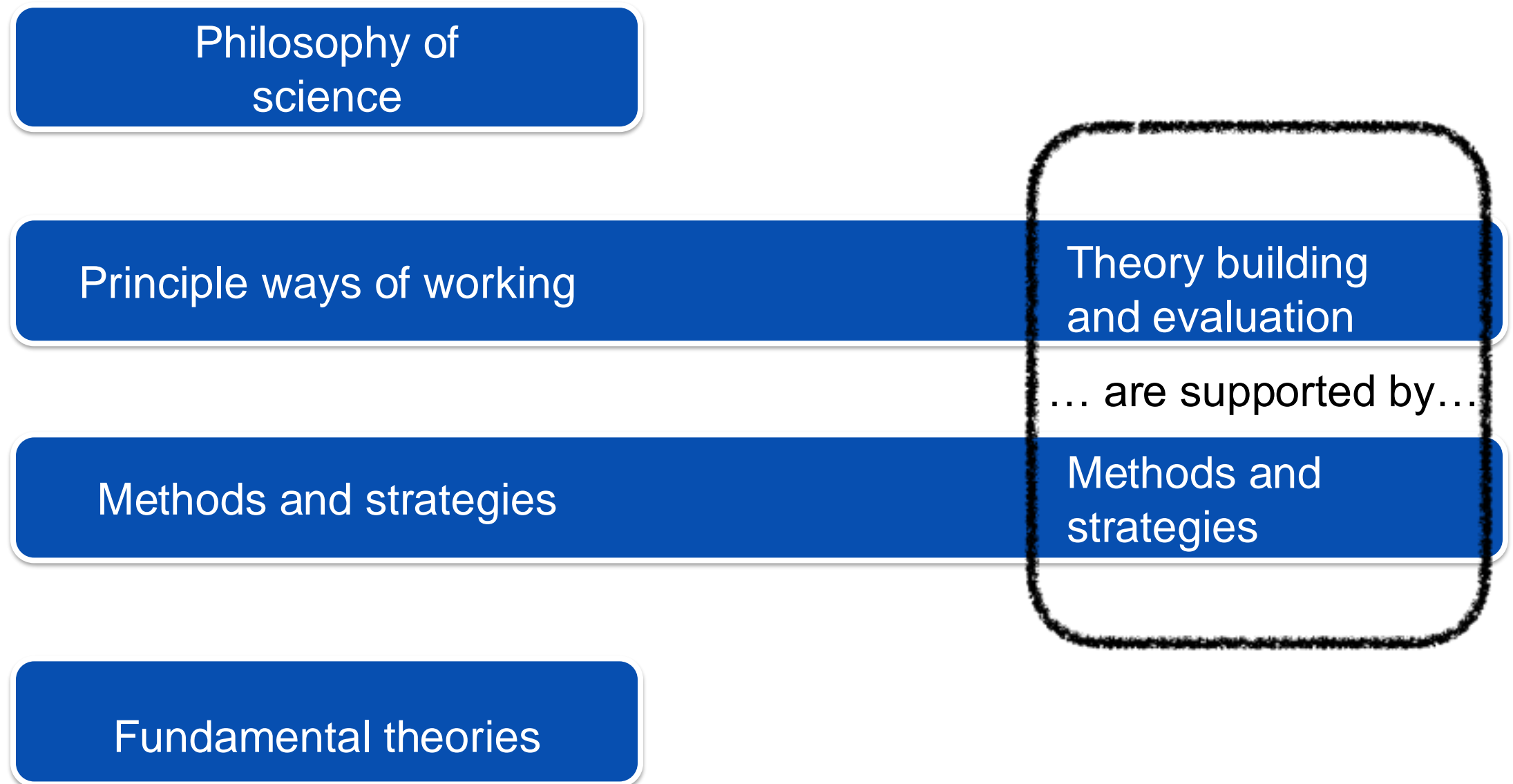
Principle ways of working

Methods and strategies

Fundamental theories

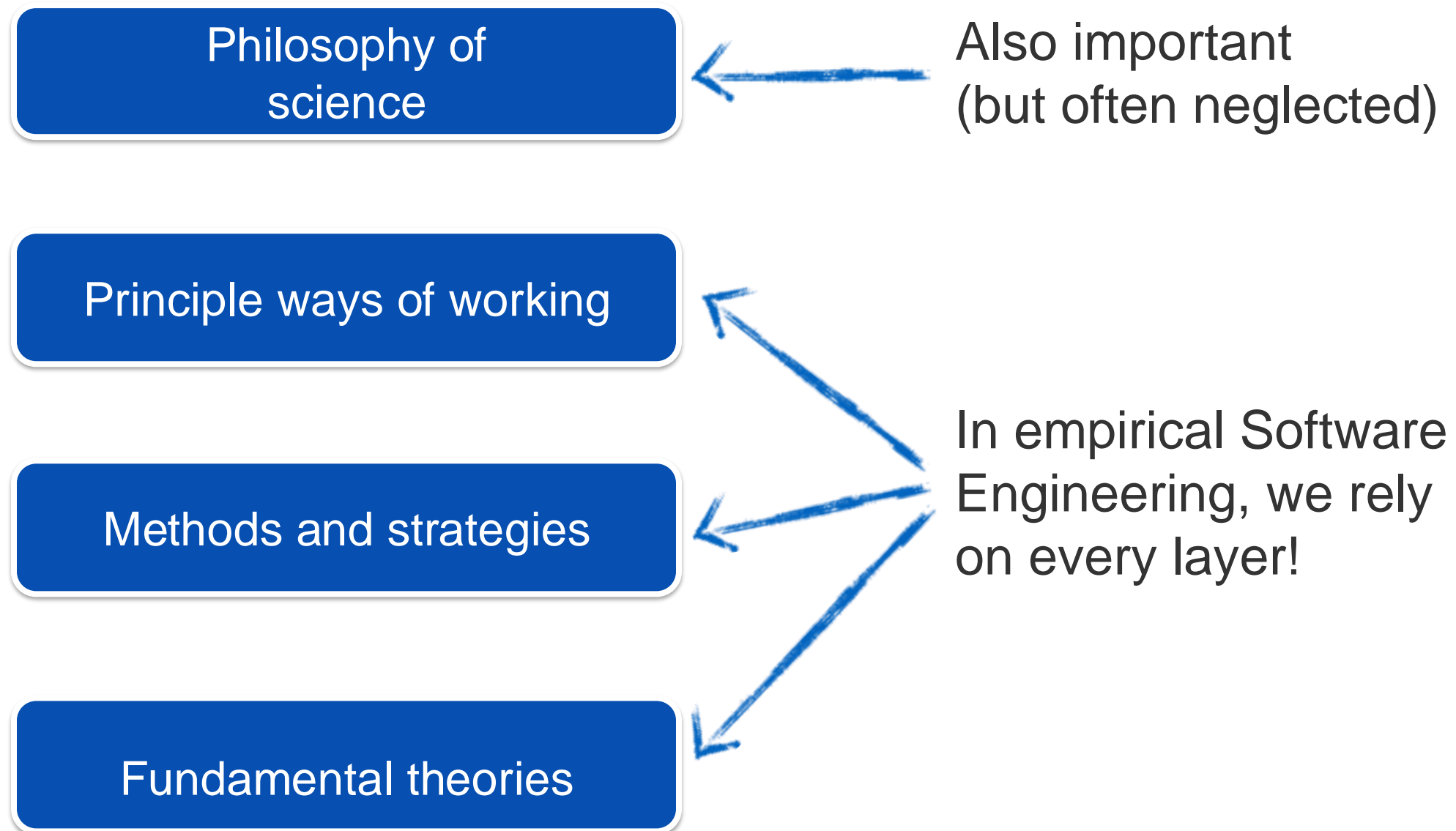


Setting: Empirical Software Engineering



Analogy: Theoretical and Experimental Physics

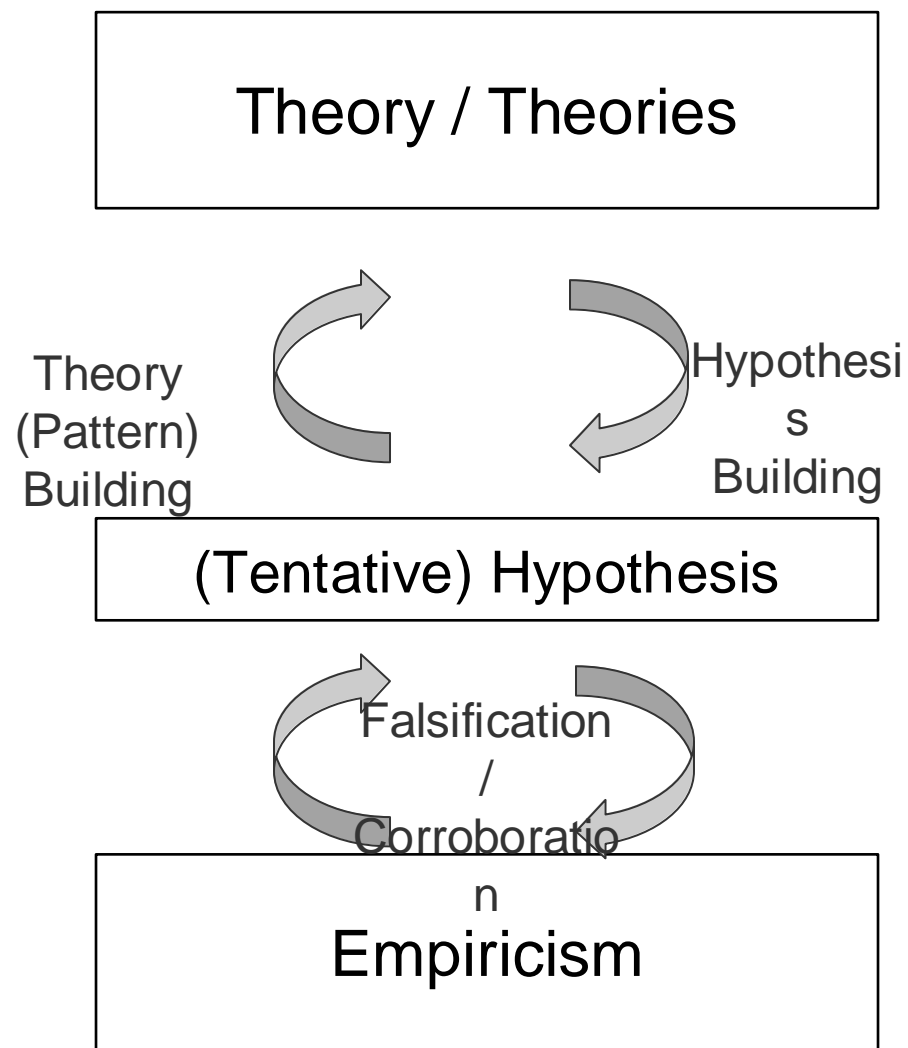
Setting: Empirical Software Engineering



A very quick step into the philosophy of science
... and back

Theories and hypotheses

*By the way
We don't "test theories", but their
consequences (via hypotheses)*



Scientific theory

- “[...] based on hypotheses tested and verified multiple times by detached researchers” (J. Bortz and N. Döring, 2003)

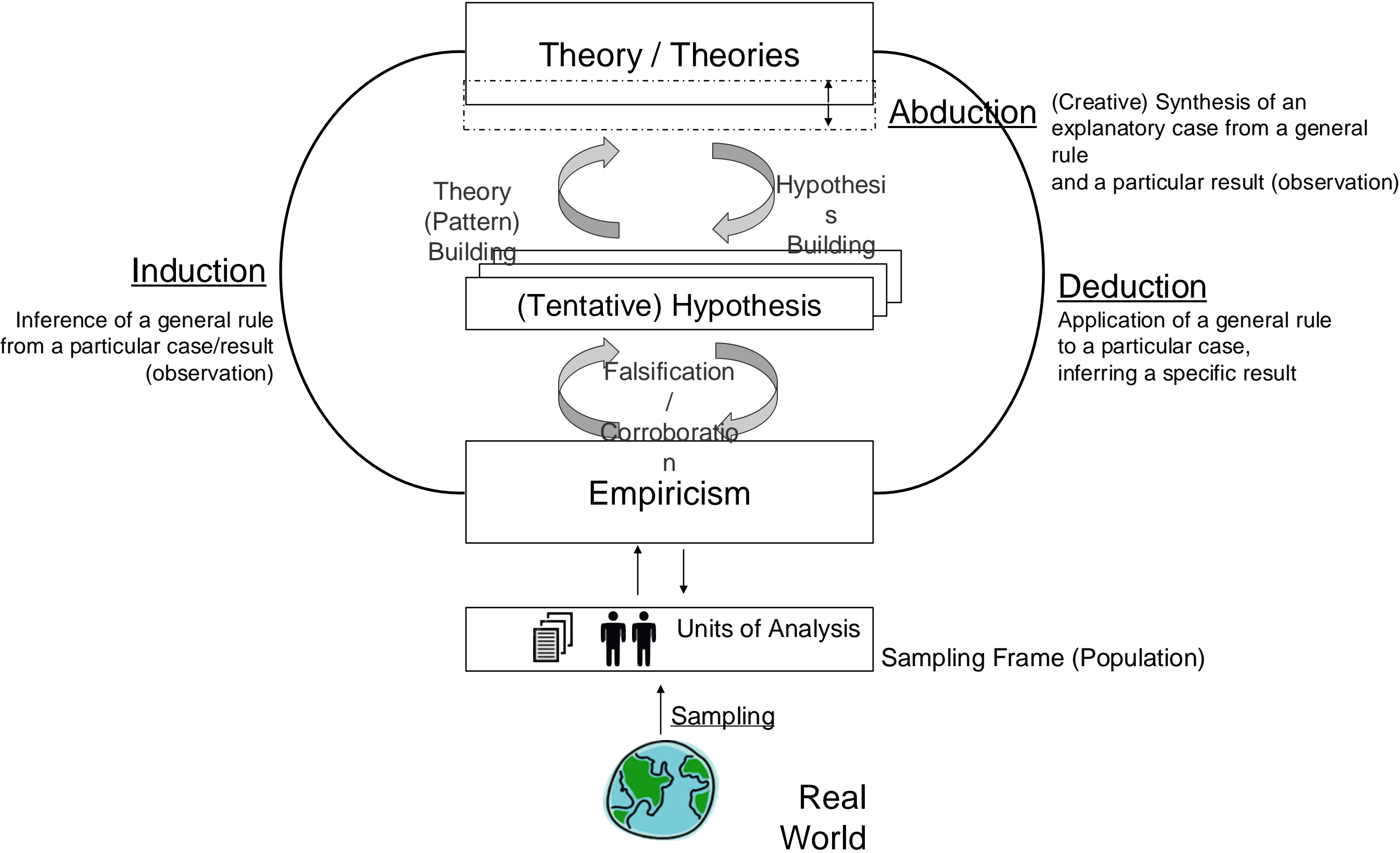
Hypothesis

- “[...] a statement that proposes a possible explanation to some phenomenon or event” (L. Given, 2008)
- Grounded in theory, testable and falsifiable
- Often quantified and written as a conditional statement

If cause/assumption (independent variables)
then (\Rightarrow)
consequence (dependent variables)

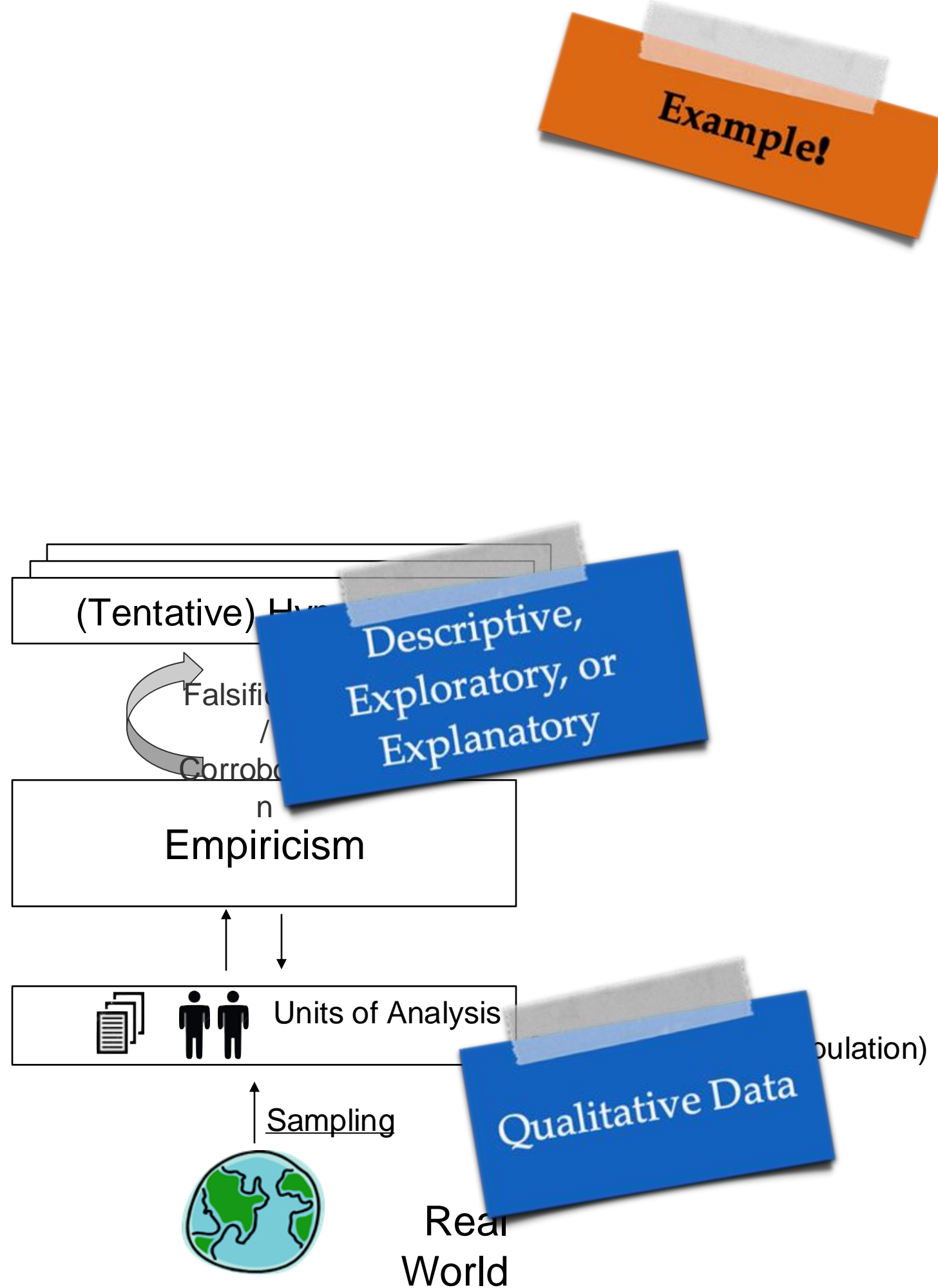
From real world to theories... and back

Principles, concepts, terms

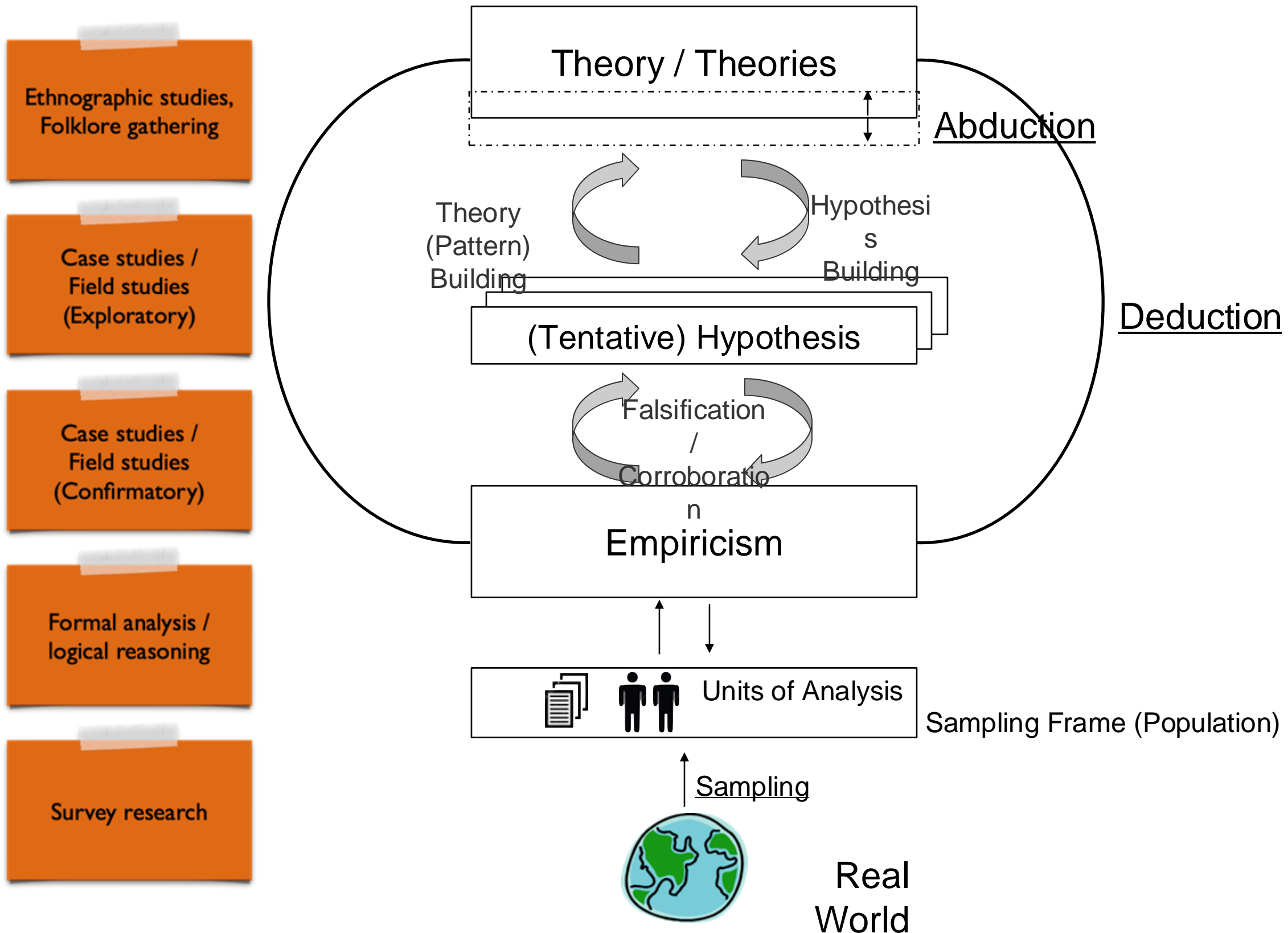


(Empirical) methods

- Each method...
 - ...has a specific purpose
 - ...relies on a specific data type
- Purposes
 - Exploratory
 - Descriptive
 - Explanatory
 - Improving
- Data Types
 - Qualitative
 - Quantitative



(Empirical) methods - where do they belong?



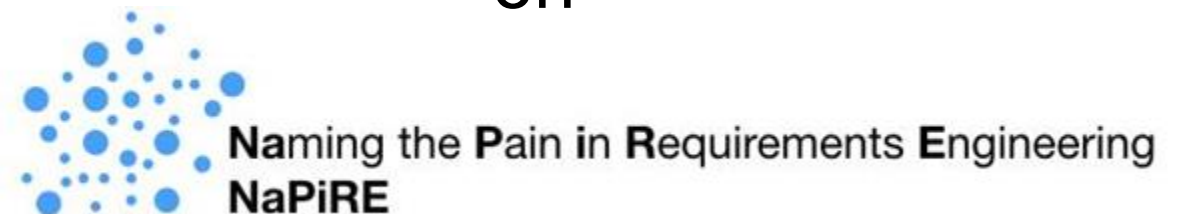
Survey research in a nutshell

Surveys

- allow for observational studies
- can have different purposes
- rely on both quantitative and qualitative data
- can employ different data analysis methods
- are (often) used in combination with multiple empirical methods (“research programmes”)

A very simplified process for survey research

* And some examples based
on



Survey Planning

Defining Research Objectives

Characterising Target Population

Sampling

Questionnaire Design

Recruiting & Measuring

Survey Execution

Data Coding & Editing

Post-survey Adjustments

Data Analysis & Interpretation

Packaging & Reporting

Data Curation & Disclosure

Reporting

Survey Planning

Defining Research Objectives

Characterising Target Population

Sampling

Questionnaire Design

Recruiting & Measuring

Survey Execution

Data Coding & Editing

Post-survey Adjustments

Data Analysis & Interpretation

Packaging & Reporting

Data Curation & Disclosure

Reporting

Survey Planning

Defining Research Objectives

Characterising Target Population

Sampling

Questionnaire Design

Recruiting & Measuring

Survey

Execution

Data Coding & Editing

Post-survey Adjustments

Data Analysis & Interpretation

Packaging & Reporting

Data Curation & Disclosure

Reporting

- Scoping overall endeavour via objectives and (research) questions
- Basis for:
 - target population
 - questionnaire design



Example!

Objective

Understand what problems practitioners experience in their Requirements Engineering.

Research

Questions

- (1) Which contemporary problems exist in RE?
- (2) What are observable patterns of problems and context characteristics?
- (3) What are their perceived causes and effects?

Survey Planning

Defining Research Objectives

Characterising Target Population

Sampling

Questionnaire Design

Recruiting & Measuring

Survey Execution

Execution

Data Coding & Editing

Post-survey Adjustments

Data Analysis & Interpretation

Packaging & Reporting

Data Curation & Disclosure

Reporting

- Target population: abstract definition of set of units to be studied
- Frame population: All units in the (envisioned) sampling frame
- Sample: Actual set of (eligible) future respondents

Example!

Target Population

Practitioners working with requirements:

- Requirements Engineers
- Software Architects
- ...

in domain X and region Y

Frame Population

- Requirements Engineers in domain X and region Y, registered in association Z
- ...

Sample

- [Role] working with requirements for [x] years
- ...

Survey Planning

Defining Research Objectives

Characterising Target Population

Sampling

Questionnaire Design

- Design of questionnaire used to answer the research questions
- Implementation / Realisation

Example!

Re

Parts	No.	Question	Type
Demographics	Q 1	What is the size of your company?	Closed(SC)

D

Po

Data Analysis

Packaging

Data Curatio

Rep

400193 NaPiRE 2017/18 Participants can take part.

+ Page + Filter + Dynamics Scoring Preview

Survey language: "English" Search

TITLE	ID	INFO	ACTIONS
Language Selection	2420013	!	+ [eye] [trash]
111 Language Selection	4423789	! V	[eye] [trash]
Instructions	2401449		+ [eye] [trash]
998 Start Page and Consent	4381586		[eye] [trash]
Metadata	2401450	!	+ [eye] [trash]
131 Country Selection	4381885	! V	[eye] [trash]
141 Team Size	4381710	! V	[eye] [trash]
111 Sector	4381702	! V	[eye] [trash]
111 System Class	4381714	! V	[eye] [trash]
121 Quality Attributes	4381722	V	[eye] [trash]
111 Team Distribution	4381731	! V	[eye] [trash]
111 Main Role Respondent	4381764	! V	[eye] [trash]
141 Experience Respondent	4381823	! V	[eye] [trash]
141 Certification Role Respondent	4381849	V	[eye] [trash]

Survey Planning

Defining Research Objectives

Characterising Target Population

Questionnaire

Recruitment

Survey Execution

Experiment

Data Collection

Post-survey

Data Analysis

Packaging & Reporting

Data Curation & Disclosure

Reporting

- Actual data collection conducted through the survey

Respondents invitation

Control

Example!

InvitationTextAnonymous.txt

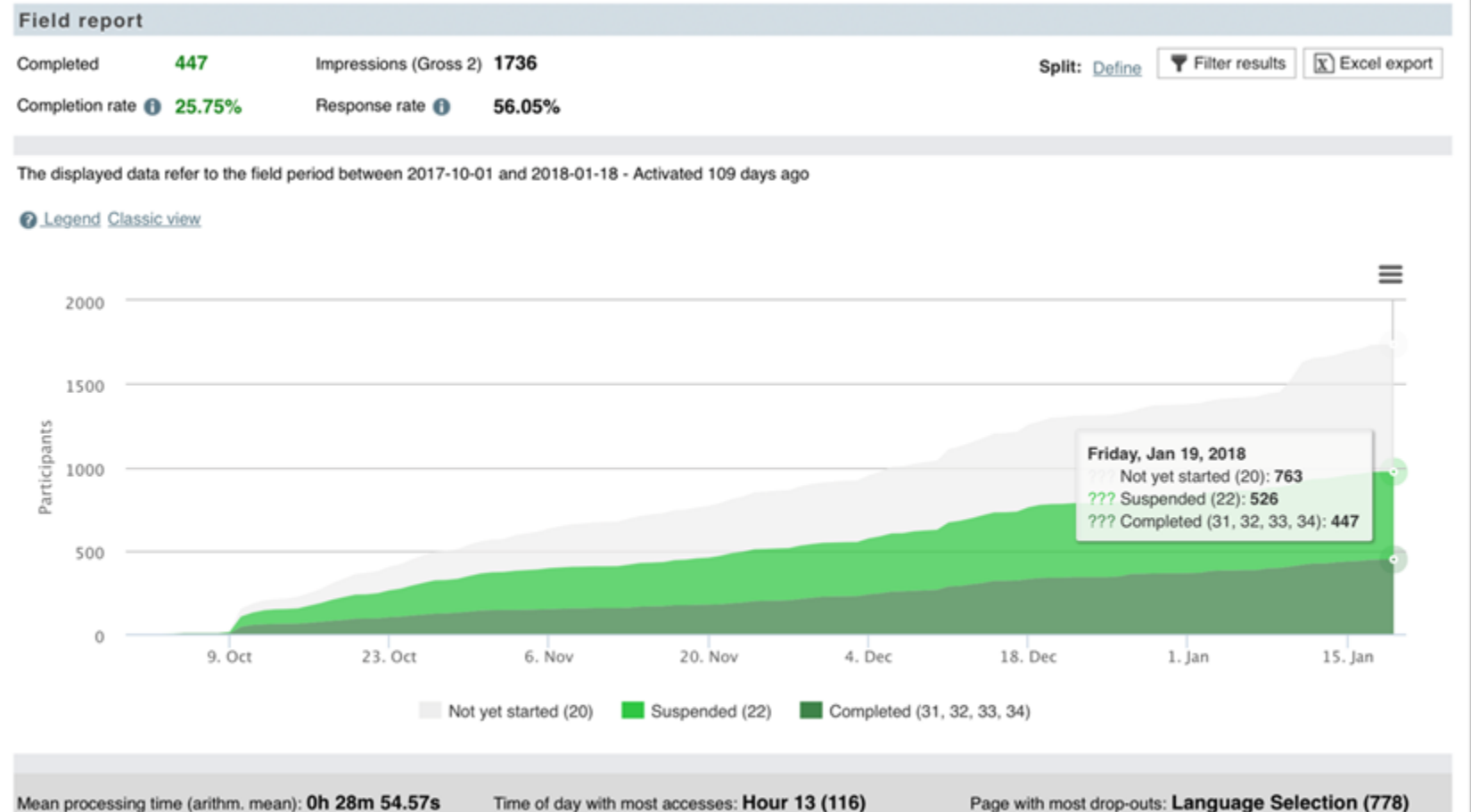
Dear colleagues,

We are launching the 3rd replication of our bi-yearly conducted survey on Requirements Engineering (RE) to explore trends in industrial practices and problems in RE. The survey forms part of the large-scale Naming the Pain in Requirements Engineering (NaPiRE) initiative [1], started in 2012, and it is conducted by an internationally distributed alliance of software engineering researchers with the goals to

(1) help the research community getting a better understanding of general industrial trends in Requirements Engineering (RE) and problems faced therein and, thus, strengthen the practical relevance of research outcomes.
(2) help practitioners assessing their current RE situation in context of overall trends.

We would like to cordially invite practitioners to participate in this survey, and researchers to further distribute the invitation.

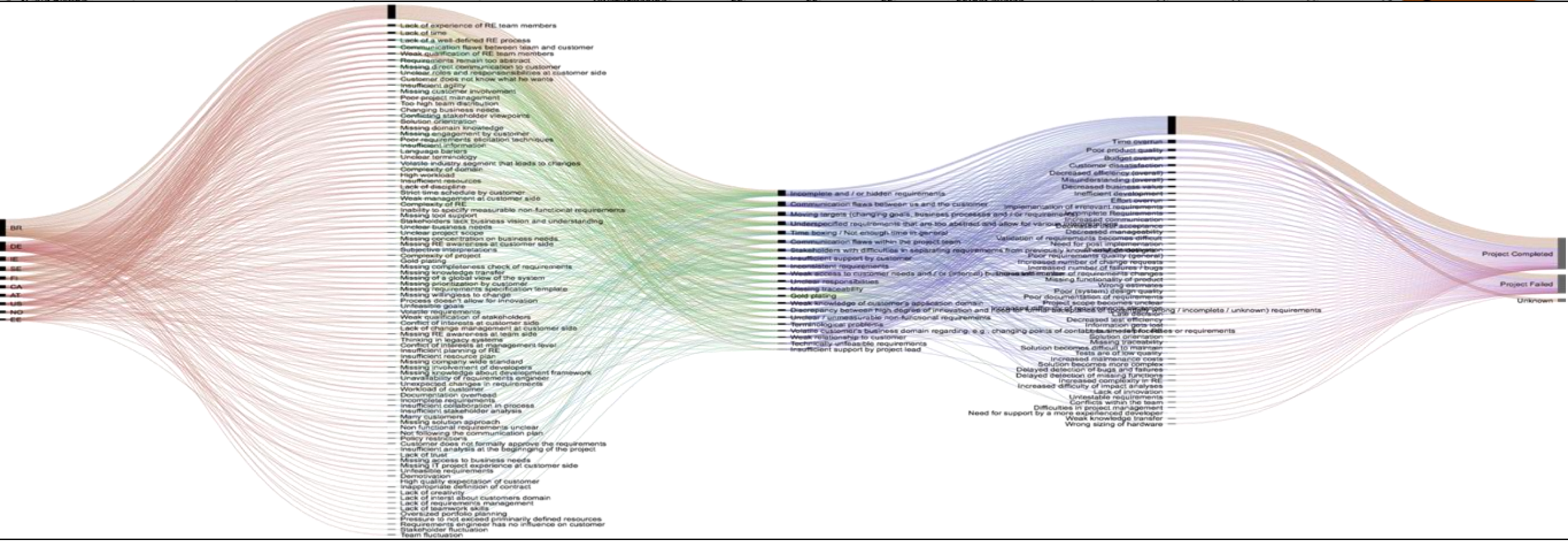
Most important facts in a nutshell:
* This survey is conducted worldwide.
* We pose a set of questions structured according to the categories:
-- General information about the respondent
-- Status Quo on requirements elicitation practices
-- Current problems experienced in the process
* The survey is anonymous, because we are interested in your critical opinions.
* The focus of the questions is on experience and opinions.
Redistribution of the survey invitation is highly appreciated.
* The overall initiative is run by the NaPiRE team, which is curated and made publicly available after the survey.
The link to the survey is <http://participate.napire.org>
We would highly appreciate your consideration.
Thanks and best regards,
The NaPiRE Team



Review



RAW DATA														
Top 5 Problems in RE					Manifestation in Process					Relation to Project Failure				
Problem #1	Problem #2	Problem #3	Problem #4	Problem #5	Problem #1	Problem #2	Problem #3	Problem #4	Problem #5	Problem #1	Problem #2	Problem #3	Problem #4	Problem #5
Missing traceability	Unclear non-functional reqs.	Inconsistent reqs.			overview lost	not enough sup	often detected			not quoted	quoted	not quoted		
Incomplete / hidden reqs.					Have to change	-66	-66			quoted		-77	-77	-77
Communication flaws	Communication flaw	Inconsistent reqs.	Incomplete / hidden reqs.	Unclear non-functional	the communication	no time, much	requirements v	no requirements	most requirements	quoted	quoted	quoted	quoted	quoted
Incomplete / hidden reqs.	Time boxing	Volatile domain	Moving targets	Separation reqs. from	feedback while	request for cha	Number of cha	Number of cha	recognizing in	quoted	quoted	not quoted	not quoted	not quoted
Moving targets	Time boxing				as system chan	Development c	-66	-66	-66	quoted	quoted		-77	-77
Underspecified reqs.	Terminological prob	Unclear responsibilities	Communication flaws to	Communication flaws	Agreed design	Overlengthy pr	agreed upon de	repeated meeti	repeated work	quoted	quoted	not quoted	not quoted	not quoted
Communication flaws	Communication flaws in team					-99	-99	-66	-66	quoted	not quoted		-77	-77
Incomplete / hidden reqs.	Inconsistent reqs.	Time boxing	Technically unfeasible reqs.			-99	-99	-99	-99	quoted	quoted	not quoted	not quoted	not quoted
Communication flaws	Incomplete / hidden reqs.	Inconsistent reqs.	Technically unfeasible reqs.		Sending docum	wrong implem	wrong implem	wasted money		quoted	not quoted	not quoted	quoted	-77
Moving targets	Time boxing	Underspecified reqs.	Unclear non-functional r	Inconsistent reqs.	leads to change	the timepoint c	leads to change	customers ofte	often detected	quoted	not quoted	not quoted	not quoted	not quoted
Incomplete / hidden reqs.	Missing traceability	Moving targets	Gold plating	Inconsistent reqs.	Discussion in C	Increasing Sys	Discussion; Ci	unfair treatmer	Discussion,	quoted	not quoted	not quoted	not quoted	quoted
Moving targets	Separation reqs. from	High degree of innovati	Time boxing	Inconsistent reqs.	features that di	sticking to wha	prototypes for	no time for cus	don't have a go	quoted	not quoted	not quoted	quoted	not quoted
Incomplete / hidden reqs.	Unclear responsibili	High degree of innovation	vs. need for formal acceptance		surprise of bus	nothing is done	-99	-66	-66	quoted	not quoted	not quoted		-77
Underspecified reqs.	Incomplete / hidden reqs.	Time boxing	Insufficient support by c	Weak relationship to c	leads to many	gaps come up	requirements a	customers are	customer dom	not quoted	quoted	not quoted	quoted	not quoted
Communication flaws	Weak access to cust	Underspecified reqs.	Incomplete / hidden reqs.	Unclear responsibilities	Clarification o	If no direct cor	A common sit	This is quite ne		not quoted	quoted	quoted	not quoted	not quoted



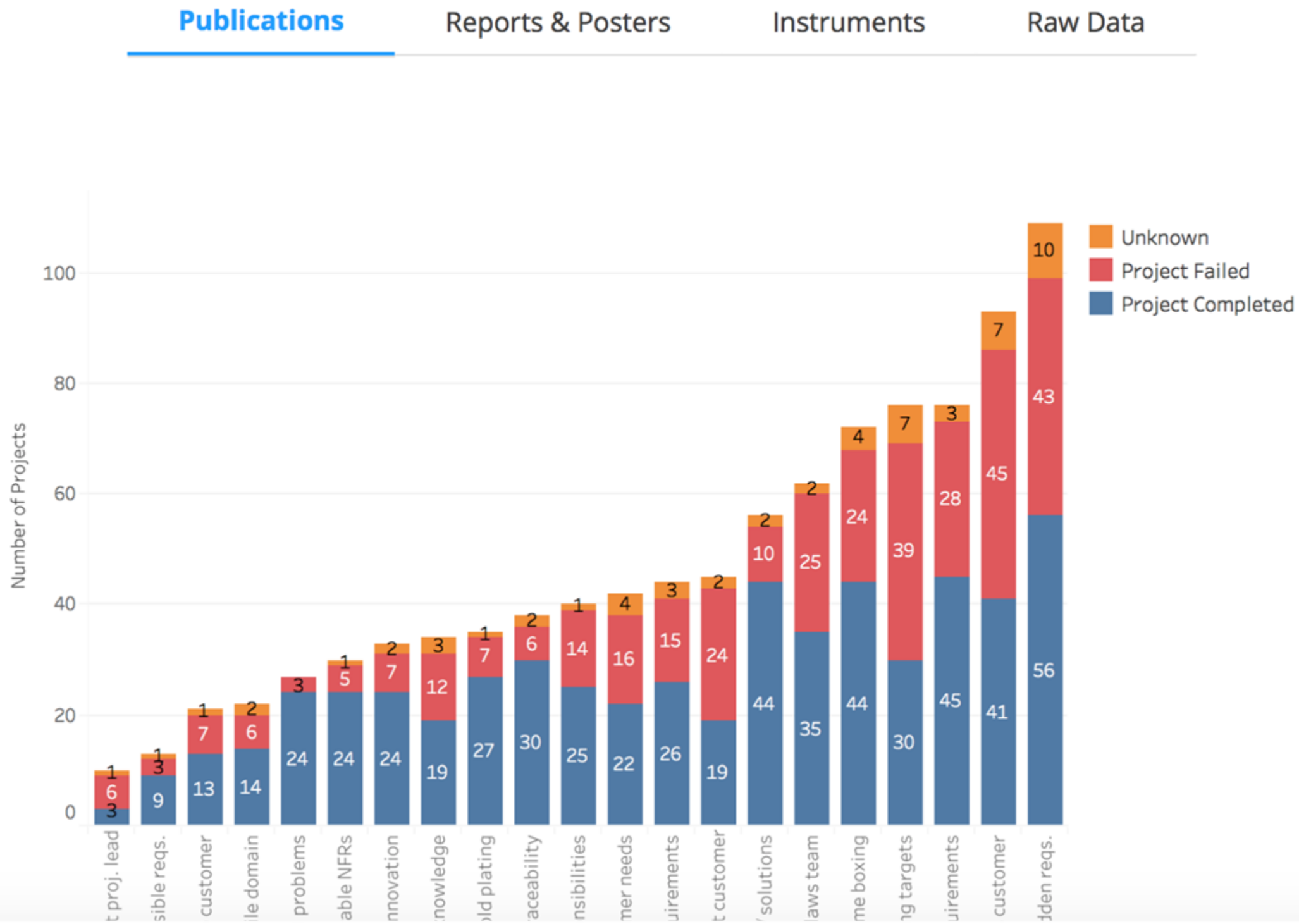
Survey Planning

Defining Research Objectives

Characterising Target Population

Sampling

- Data curation and disclosure: Preparation of data for reproducible disclosure (e.g. anonymisation, codebooks)



technical

sample!

Questions?



Outline

- A brief introduction into survey research
- (Selected) Best practices
- Lean coffee

Example - What could be wrong here?

Dear Dr. Fernández,

We are conducting a survey to evaluate an updated standard model that will be used to [REDACTED]. This survey will help to ensure the suitability and validity of this model and ensure its usefulness in meeting a specific class of problems in the analysis of multi-agent systems. Also, it could point to some important missing details and weak points in the updated model structure.

We are inviting you to participate in this research study. You will be asked to go through the model sections using a few real scenarios and then take a survey to answer a few questions regarding your experience using the updated model. On this email, you will find attached the updated model file.

This survey has no risk to participants, and it will not require any personal information. It will be only known by my advisor and me. Your participation is voluntary, and you are not likely to have any direct benefit from being in this research study. The expected time to complete the entire task is 20 minutes.

The following is the online survey that needs your participation: [REDACTED]

Thank you for participating.

[REDACTED]

--

Best Regards

[REDACTED]

Who are they?

17 pages attachment to
20 minutes to answer the

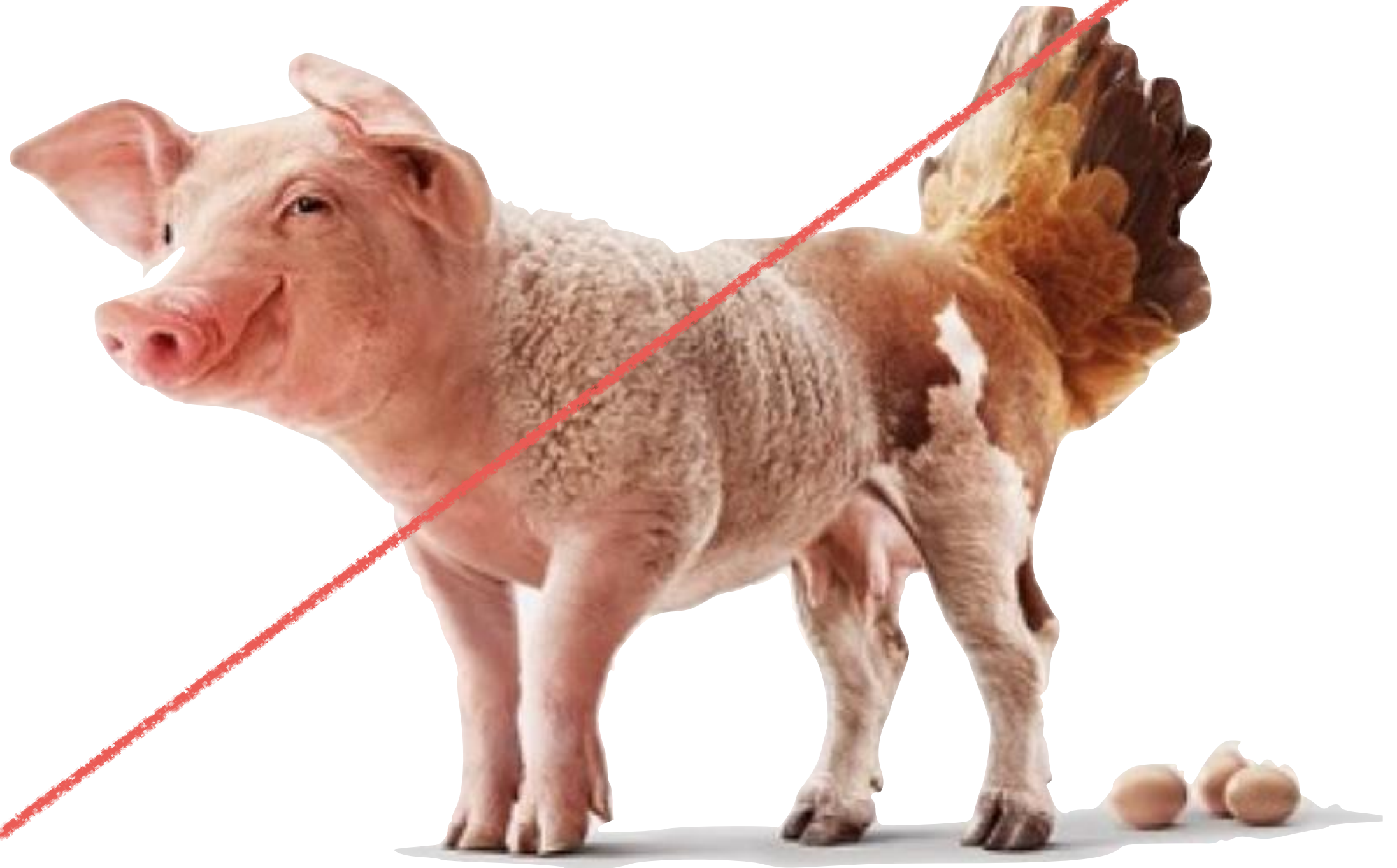
SURVEY



Disclaimer



Disclaimer (Bavarian edition)



Survey Planning

Defining Research Objectives

Characterising Target Population

Sampling

Questionnaire Design

Recruiting & Measuring

Survey Execution

Data Coding & Editing

Post-survey Adjustments

Data Analysis & Interpretation

Packaging & Reporting

Data Curation & Disclosure

Reporting

- There are too many pitfalls to be handled in a short tutorial.

» Recommended reading:

Torchiano, M., Méndez Fernández, D., Travassos, G.H., de Mello, R. M. (2017). Lessons Learnt in Conducting Survey Research. In: Proc. 5th International Workshop on Conducting Empirical Studies in Industry (CESI). ICSE 2017. Available at

<https://arxiv.org/abs/1702.05744>



Survey Planning

Defining Research Objectives

Characterising Target Population

Sampling

Questionnaire Design

Recruiting & Measuring

Survey Execution

Data Coding & Editing

Post-survey Adjustments

Data Analysis & Interpretation

Packaging & Reporting

Data Curation & Disclosure

Reporting

We concentrate on 4 issues, instead
(I personally find highly important)

Defining research objectives

Challenge

Know the limitations of survey research

- Survey research opts for answers that rely on experiences, opinions, and observations (folklore) of the respondents
- Respondents' bias is our natural environment
 - » Develop internal questions to help you depict the research objective and target population
 - » Opt for descriptive questions (“what is happening?”) or explanatory questions (“why is this happening?”) rather than normative questions (“what should we do?”)

Characterising the target population

Challenge

Identify the real target population

- Do not restrict the target population by factors like availability or expected number of responses!
 - » Based on the research objectives, answer the question:
“Who can best provide you with the information you need?”,
not
“Who are probably available to participate?”

Characterising the target population

Challenge

Identify the units of analysis (1/2)

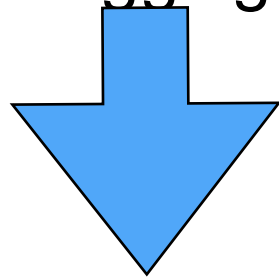
- Individuals? Groups? Teams? Companies?
- For instance, investigating Java developers' programming practice is a research objective different from investigating Company practice for Java programming

Characterising the target population

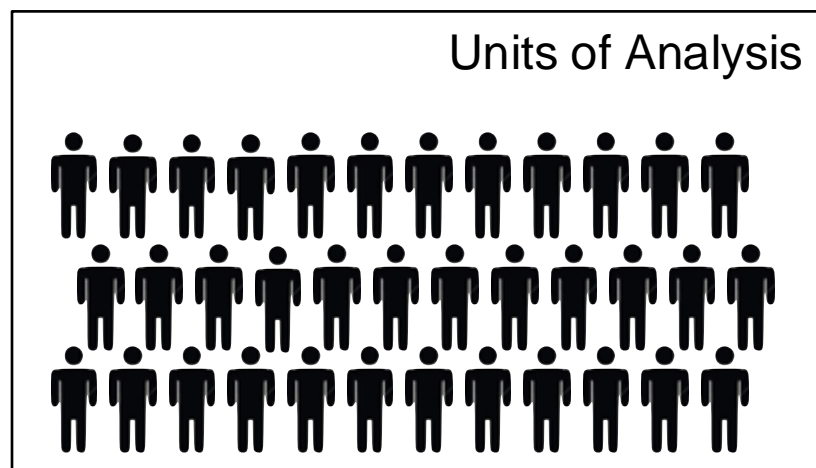
Challenge

Identify the units of analysis (2/2)

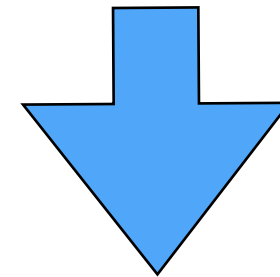
“How do developers perform code debugging?”



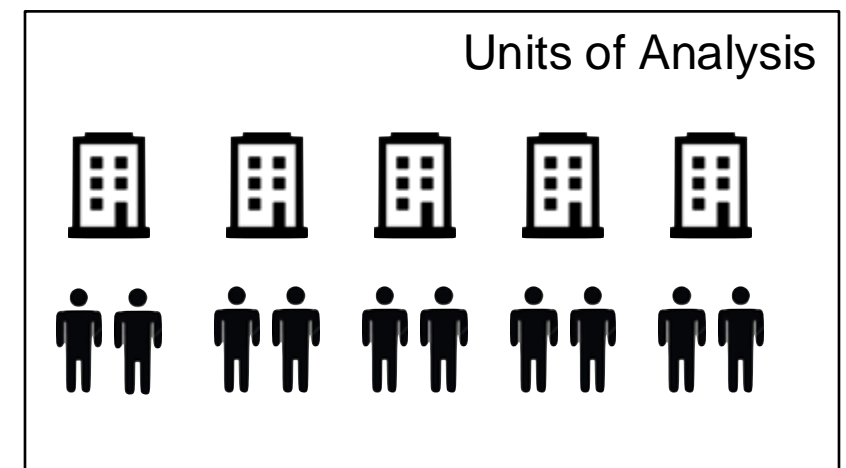
List of developers



“How is code debugging performed in companies?”



List of companies



Characterising the target population

Challenge

Characterise the subjects and units of analysis

- Different research objectives may demand different attributes to characterising individuals/ groups of individuals
- » Use standard reference models (e.g. in context of process tailoring):
 - Individuals: experience in the research context, experience in SE, current professional role, location and higher academic degree, ...
 - Project teams: team size, client/product domain (avionics, finance, health, telecommunications, etc.) and physical distribution, ...
 - Organisations: size, industry segment, location, type (government, private company, university, etc.), ...

Questionnaire design

Challenge

Design a clear, simple and consistent survey questionnaire

- Bad questionnaires can lead subjects initially willing to participate to give up!
 - » Use simple and appropriate wording for the survey questions
 - » Avoid vague sentences
 - » Avoid technical terms as much as possible or define them in the questionnaire, according to the survey target population
 - » Take preference to design short questions regarding a single concept
 - » Avoid double barreled questions

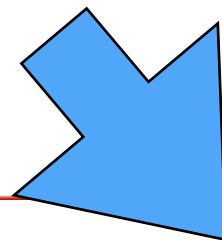
Questionnaire design

Challenge

Design a clear, simple and consistent survey questionnaire

In your opinion, do you agree or disagree that code refactoring is a need? And what about code smell detection?

- a) I strongly agree
- b) I partially agree
- c) I agree
- d) I disagree



Code refactoring is an essential practice for improving the understanding of object-oriented code.

- a) Totally agree
- b) Partially agree
- c) Neither agree nor disagree
- d) Partially disagree
- e) Totally disagree

Questionnaire design

Challenge

Design a clear, simple and consistent survey questionnaire

- » Avoid biased questions

Do you prefer working in projects following agile methods or those following usual non-agile approaches?

- » Avoid asking about events too far in the past

Considering the main characteristics of the last 10 software projects you have worked on, please answer the following questions:

Questionnaire design

Challenge

Design a clear, simple and consistent survey questionnaire

- » Avoid asking sensitive questions unless you really need to (and if you need to, make explicit why you ask these questions)

What is your gender?

What is your income?

What is your age?

Questionnaire design

Challenge

Design a clear, simple and consistent survey questionnaire

- » Avoid asking too demanding questions (w.r.t. time, effort)

After reading the attached papers regarding non functional requirements (NFR), please answer the following questions:

1. Which of the following NFR do you disagree are not relevant in the context of real-time systems?

...

Questionnaire design

Challenge

Design a clear, simple and consistent survey questionnaire

- » Select the appropriate response formats and scales

How much experience do you have in Java programming?

I have been working with Java programming at companies since 2011. Before, I got my first Java certification in 2009, when I started working in personal projects. But I have difficult with object-oriented parts... _____

How much experience do you have in Java programming?

 5 years

Do you have experience in Java programming?

Yes

No

Response formats and scales

	Nominal scale	Ordinal (and “Likert”) scales	Interval scale	Free-text responses	Numeric values
Characteristic	<ul style="list-style-type: none"> • Closed questions 	<ul style="list-style-type: none"> • Closed questions • Not always equally distributed • No distance measure 	<ul style="list-style-type: none"> • Closed questions • Considered equally distributed 	<ul style="list-style-type: none"> • Open questions • Allow “coding” 	<ul style="list-style-type: none"> • Open questions
Analysis	<ul style="list-style-type: none"> • Statistical analysis based on frequency 	<ul style="list-style-type: none"> • Significantly restricts statistical analysis 	<ul style="list-style-type: none"> • Statistical analysis less restrictive 	<ul style="list-style-type: none"> • Content analysis • High effort on data analysis 	<ul style="list-style-type: none"> • Allow a wide range of statistical analysis

Questionnaire design

Challenge

Design a clear, simple and consistent survey questionnaire

- Remember: you have one shot only!
(Once you started the survey, there is usually no way back.)
 - » Pilot the survey instrument with respondents characteristic for your population!
 - » Pilot the data (quantity and quality) by applying the planned data analysis techniques!

Data curation and disclosure

Challenge

A survey needs to be reproducible

- Reporting on a survey without background information is asking for too much credit from the reader (and the reviewers!)
 - » Report on details of all data collection with a study protocol
 - » Disclose the instrument used: Questionnaire, reference documents
 - » Disclose the anonymised data and the codebooks
 - » Do not use your institution websites (prone to changes)
 - » Instead, use open repositories like Figshare (providing DOIs)

Besides all methodological issues...
there is more.

Every survey needs a proper project plan

1. Plan for methodological challenges (✓)
2. Find a proper project organisation early
3. Set up a proper project infrastructure
4. Develop a good project dissemination plan
5. Organise an efficient data collection
6. Organise an efficient data curation and analysis
7. Develop a good packaging and reporting plan

Are there any particular best practices you would like to share?

Further reading

- Torchiano, M., Méndez Fernández, D., Travassos, G.H., de Mello, R. M. (2017). Lessons Learnt in Conducting Survey Research. In: Proc. 5th International Workshop on Conducting Empirical Studies in Industry (CESI). ICSE 2017.
Available at <https://arxiv.org/abs/1702.05744>
- Groves, Fowler, Couper, Lepkowski, Singer and Torangeau, (2009). “Survey Methodology – 2nd edition” John Wiley and Sons
- Conradi R., Li J., Slyngstad O. P. N., Kampenes V. B., Bunse C., Morisio M., Torchiano M. (2005). “Reflections on conducting an international survey of CBSE in ICT industry” IEEE 4th International Symposium on Empirical Software Engineering November.
- Linåker, J. Sulaman, S. M., de Mello, R. M. and Höst, M. (2015). Guidelines for Conducting Surveys in Software Engineering. TR 5366801, Lund University Publications. <http://lup.lub.lu.se/record/5366801/file/5366839.pdf>
- de Mello, R. M. and Travassos, G. H. (2016). Surveys in Software Engineering: Identifying Representative Samples. Proc. of 10th ACM/IEEE ESEM, Ciudad Real.

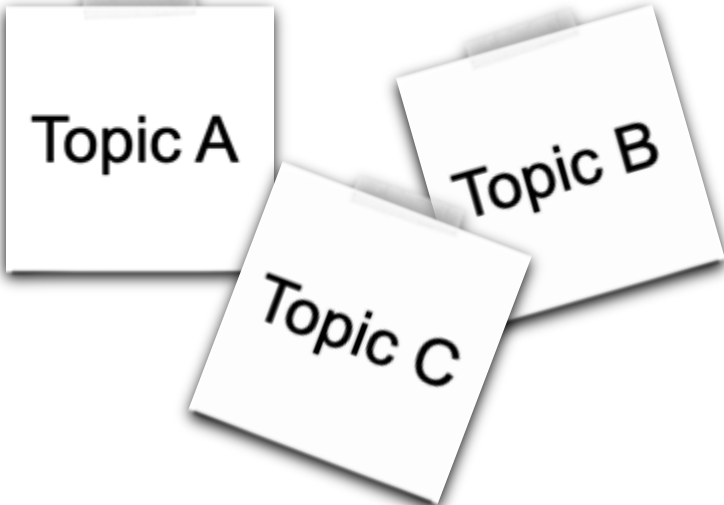


Outline

- A brief introduction into survey research
- (Selected) Best practices
- Lean coffee

Lean Coffee

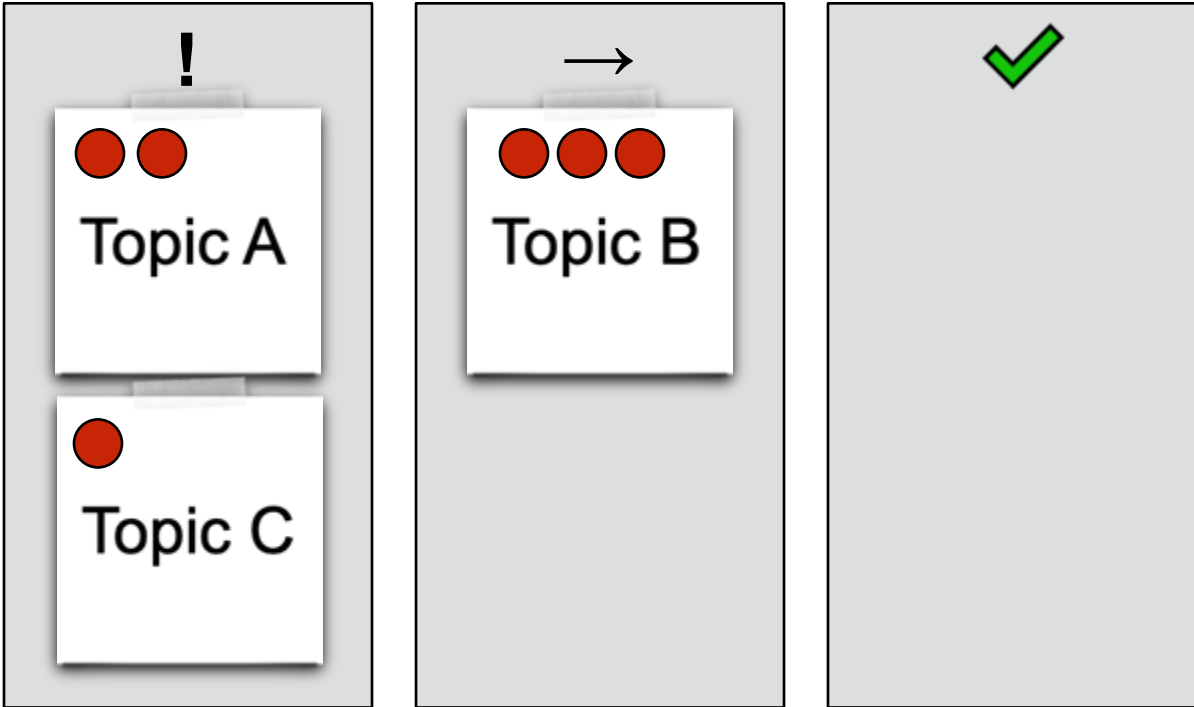
1. Collect topics



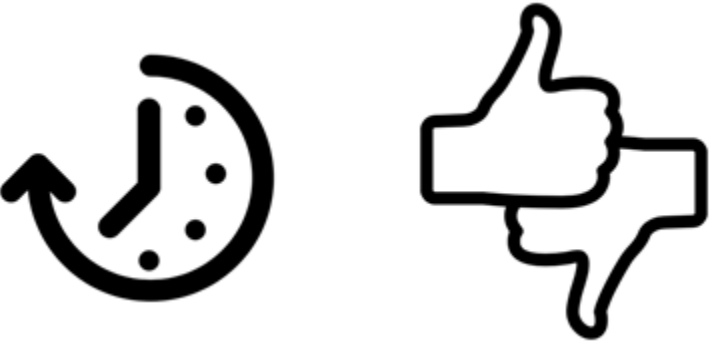
2. Rank by interest



3. Discuss topics in the group



4. Revise
(Continue with B or move on to A?)



Thanks for your participation!

In case of questions, approach me any time!