Research Methods in Computer Science

(Serge Demeyer — University of Antwerp)

AnSyMo
Antwerp Systems and software Modelling
http://ansymo.ua.ac.be/

1. Research Methods

Introduction
- Origins of Computer Science
- Research Philosophy

Research Methods
- 1. Feasibility study
- 2. Pilot Case
- 3. Comparative study
- 4. Observational Study [a.k.a. Ethnography]
- 5. Literature survey
- 6. Formal Model
- 7. Simulation

Conclusion
- Studying a Case vs. Performing a Case Study
  - Proposition
  - Unit of Analysis
  - Threats to Validity

How to perform research? 
(and get "empirical" results)

How to write research? 
(and get papers accepted)

How many of you have done / will do a case-study?

Helicopter View
**Computer Science**

All science is either physics or stamp collecting (E. Rutherford)

We study artifacts produced by humans

Computer science is no more about computers than astronomy is about telescopes. (E. Dijkstra)

**Mathematical Origins**

- **Turing Machines**
  - Halting problem

- **Algorithmic Complexity**
  - $P = \text{?} NP$

- **Compilers**
  - Chomsky hierarchy

- **Databases**
  - Relational model

(Inductive) Reasoning

- logical argumentation
- theorem proving, ...
- axioms & lemma's
- foo, bar type of examples
- "deep" and generic universal knowledge

Gödel theorem: consistency of the system is not provable in the system.

⇒ A complete and consistent set of axioms for all of mathematics is impossible

**Science vs. Engineering**

Science

- Physics
- Chemistry
- Biology
- Mathematics
- Geography

Engineering

- Civil Engineering
- Electronics
- Chemistry and Materials
- Electro-Mechanical Engineering

1. Research Methods

**Science vs. Engineering**

- Computer Science
- Software Engineering

**Engineering Origins**

- **Computer Engineering**
  - Moore’s law: “the number of transistors on a chip will double about every two years”
  - Self-fulfilling prophesy
  - Hardware technology
  - RISC vs. CISC
  - MPSoC
  - Compiler optimization
  - peephole optimization
  - branch prediction

- **Empirical Approach**
  - Tom De Marco: “you cannot control what you cannot measure”
  - quantify
  - mathematical model
  - Pareto principle
  - 80% - 20% rule
    - (80% of the effects come from 20% of the causes)

As good as your next observation.

Premise: The sun has risen in the east every morning up until now.

Conclusion: The sun will also rise in the east tomorrow. … Or Not?
1. Research Methods

Influence of Society

Lives are at stake (e.g., automatic pilot, nuclear power plants)

Huge amounts of money are at stake (e.g., Ariane V crash, Denver Airport Baggage)

Software became Ubiquitous... it's not a hobby anymore

Corporate success or failure is at stake (e.g., telephone billing, VTM launching 2nd channel)

Interdisciplinary Nature

Science

Engineering

“Hard” Sciences

Science

“Soft” Sciences

Psychology

Sociology

Economics

Computer Science

Action Research

The Oak Forest
Robert Zünd - 1982

Franz and Luciano
Franz Gertsch - 1973
Objective ↔ Subjective

- Plato’s cave
- Scientific Paradigm (Kuhn)
  + Dominant paradigm / Competing paradigms / Paradigm shift
  ⇒ Normal science vs. Revolutionary science

1. Research Methods

Dominant view on Research Methods

**Physics**
(“The” Scientific method)
- form hypothesis about a phenomenon
- design experiment
- collect data
- compare data to hypothesis
- accept or reject hypothesis
- ... publish (in Nature)
- get someone else to repeat experiment (replication)

**Medicine**
(Double-blind treatment)
- form hypothesis about a treatment
- select experimental and control groups that are comparable except for the treatment
- collect data
- commit statistics on the data
- treatment → difference (statistically significant)

Cannot answer the “big” questions
... in timely fashion
- smoking is unhealthy
- climate change
- darwin theory vs. intelligent design
- ...
- agile methods

2. Pilot Case, Demonstrator

Cannot answer the “big” questions ...

Case studies - Spectrum

case studies are widely used in computer science
⇒ “studying a case” vs. “doing a case study”

1. Feasibility study
  • is it possible?

2. Pilot Case, Demonstrator
  • is it appropriate?

3. Comparative study
  • What is “it”?

4. Observational Study
  • what is known/unknown?

5. Literature survey
  • underlying concepts?

6. Formal Model
  • what if?

7. Simulation
  • what if?
Feasibility Study

Here is a new idea, is it possible?

- Metaphor: Christopher Columbus and western route to India

- Is it possible to solve a specific kind of problem ... effectively?
  - computer science perspective ($P = NP$, Turing test, ...)
  - engineering perspective (build efficiently; fast — small)
  - economic perspective (cost effective; profitable)

- Is the technique new / novel / innovative?
  - compare against alternatives
    - See literature survey; comparative study

- Proof by construction
  - build a prototype
  - often by applying on a “CASE”

- Conclusions
  - primarily qualitative; "lessons learned"
  - quantitative
    - economic perspective: cost - benefit
    - engineering perspective: speed - memory footprint

Pilot Case (a.k.a. Demonstrator)

Here is an idea that has proven valuable; does it work for us?

- Metaphor: Portugal (Amerigo Vespucci) explores western route

- proven valuable
  - accepted merits (e.g. "lessons learned" from feasibility study)
  - there is some (implicit) theory explaining why the idea has merit

- does it work for us
  - context is very important

- Demonstrated on a simple yet representative “CASE”
  - “Pilot case” ≠ "Pilot Study"

- Proof by construction
  - build a prototype
  - apply on a "case"

- Conclusions
  - primarily qualitative; "lessons learned"
  - quantitative; preferably with predefined criteria
    - compare to context before applying the idea!!
Comparative Study

Here are two techniques, which one is better?

- for a given purpose!
  + (Not necessarily absolute ranking)
- Where are the differences? What are the tradeoffs?

- Criteria check-list
  + predefined
    - should not favor one technique
  + qualitative and quantitative
    - qualitative: how to remain unbiased?
    - quantitative: represent what you want to know?
  + Criteria check-list should be complete and reusable!
    ➡ If done well, most important contribution (replication!)
    ➡ See literature survey

- Score criteria check-list
  + Often by applying the technique on a "CASE"

- Compare
  + typically in the form of a table

Observational Study [Ethnography]

Understand phenomena through observations

- Metaphor: Diane Fossey "Gorillas in the Mist"

- systematic collection of data derived from direct observation of the everyday life
  + phenomena is best understood in the fullest possible context
    ➡ observation & participation
    ➡ interviews & questionnaires

- Observing a series of cases "CASE"
  + observation vs. participation?

- example: Action Research
  + Action research is carried out by people who usually recognize a problem or limitation in their workplace situation and, together, devise a plan to counteract the problem, implement the plan, observe what happens, reflect on these outcomes, revise the plan, implement it, reflect, revise and so on.

- Conclusions
  + primarily qualitative: classifications/observations/...
Literature Survey

What is known? What questions are still open?


Systematic
- "comprehensive"
  - precise research question is prerequisite
  - defined search strategy (rigor, completeness, replication)
  - clearly defined scope
  - criteria for inclusion and exclusion
  - specify information to be obtained
  - the "CASES" are the selected papers

- outcome is organized

<table>
<thead>
<tr>
<th>classification</th>
<th>taxonomy</th>
<th>conceptual model</th>
</tr>
</thead>
<tbody>
<tr>
<td>table</td>
<td>tree</td>
<td>frequency</td>
</tr>
</tbody>
</table>

1. Research Methods
Formal Model

How can we understand/explain the world?
- make a mathematical abstraction of a certain problem
  - analytical model, stochastic model, logical model, re-write system, ...
  - often explained using a "CASE"
- prove some important characteristics
  - based on inductive reasoning, axioms & lemma’s, ...

Motivate
- which factors are irrelevant (excluded) and which are not (included)?
- which properties are worthwhile (proven)?
  ➡ See literature survey

Simulation

What would happen if ...?
- study circumstances of phenomena in detail
  + simulated because real world too expensive; too slow or impossible
- make prognoses about what can happen in certain situations
  + test using real observations, typically obtained via a "CASE"

Motivate
- which circumstances are irrelevant (excluded) and which are not (included)?
- which properties are worthwhile (to be observed/predicted)?
  ➡ See literature survey

Examples
- distributed systems (grid); network protocols
  + too expensive or too slow to test in real life
- embedded systems — simulating hardware platforms
  + impossible to observe real clock-speed / memory footprint / ...
  ➡ Heisenberg uncertainty principle

Case studies - Revisited

case studies are widely used in computer science
⇒ "studying a case" vs. "doing a case study"

1. Feasibility study
   Proof by construction; often by applying on a “CASE”

2. Pilot Case, Demonstrator
   Demonstrated on a simple yet representative “CASE”

3. Comparative study
   Score criteria check-list; often by applying on a “CASE”

4. Observational Study
   Observing a series of “CASES”

5. Literature survey
   “CASES” = selected papers

6. Formal Model
   often explained using a “CASE”

7. Simulation: test prognoses with real observations obtained via a “CASE”
**Case Study Research**

1. **Research Methods**
   - Origins of Computer Science
   - Research Philosophy
   - Case Study Research
     - Feasibility study
     - Pilot Case
     - Comparative study
     - Observational Study [a.k.a. Ethnography]
     - Literature survey
     - Formal Model
     - Simulation

2. **Conclusion**
   - Studying a Case
   - Performing a Case Study
   - + Unit of Analysis
   - + Threats to Validity

---

**Case study — definition**

A case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between the phenomenon and context are not clearly evident.

[Robert K. Yin. Case Study Research: Design and Methods; p. 13]

- empirical inquiry: yes, it is empirical research
- contemporary: (close to) real-time observations
  - + incl. interviews
- boundaries between the phenomenon and context not clear
  - + as opposed to "experiment"

---

**Spectrum of cases**

- created for explanation
- • foo, bar examples
- • simple model; illustrates differences
- • "textbook example"
- competition (tool oriented)
  - • approved by community
  - • comparing
- Benchmark
- real-life example
  - • industrial system,
  - open-source system
  - • context is difficult to grasp
- Exemplar
- mining Software Repositories Challenge.
  [Yearly workshop where research tools compete against one another on a common predefined case.]
- Benchmark
- approved by community
- known context
- "planted" issues
- Case
- Benchmark

---

**Case Study — Counter evidence**

- many more variables than data points
- multiple sources of evidence; triangulation
- theoretical propositions guide data collection
  - (try to confirm or refute propositions with well-selected cases)

---

**Sources**
## Misunderstanding 2: Generalization

One cannot generalize on the basis of an individual case; therefore the case study cannot contribute to scientific development. — [Bent Flyvbjerg, “Five Misunderstandings About Case Study Research.”]

- Understanding
  + The power of examples
  + Formal generalization is overvalued
    - dominant research views of physics and medicine

- Counterexamples
  + one black swan falsifies “all swans are white”
  - case studies generate deep understanding; what appears to be white often turns out to be black

- sampling logic vs. replication logic
  + sampling logic: operational enumeration of entire universe
    - use statistics: generalize from “randomly selected” observations
  + replication logic: careful selection of boundary values
    - use logic reasoning: presence of absence of property has effect

## Sampling Logic vs. Replication Logic

<table>
<thead>
<tr>
<th>Random selection</th>
<th>Selection of (boundary) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>⇒ generalize for entire population</td>
<td>⇒ understand differences</td>
</tr>
<tr>
<td>• propositions</td>
<td></td>
</tr>
<tr>
<td>• units of analysis</td>
<td></td>
</tr>
</tbody>
</table>

## Research questions for Case Studies

<table>
<thead>
<tr>
<th>Existence:</th>
<th><strong>Exploratory</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Does X exist?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description &amp; Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>• What is X like?</td>
</tr>
<tr>
<td>• What are its properties?</td>
</tr>
<tr>
<td>• How can it be categorized?</td>
</tr>
<tr>
<td>• How can we measure it?</td>
</tr>
<tr>
<td>• What are its components?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Descriptive-Comparative</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How does X differ from Y?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency and Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How often does X occur?</td>
</tr>
<tr>
<td>• What is an average amount of X?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Descriptive-Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How does X normally work?</td>
</tr>
<tr>
<td>• By what process does X happen?</td>
</tr>
<tr>
<td>• What are the steps as X evolves?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relationship</th>
<th><strong>Explanatory</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Are X and Y related?</td>
<td></td>
</tr>
<tr>
<td>• Do occurrences of X correlate with occurrences of Y?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Causality</th>
</tr>
</thead>
<tbody>
<tr>
<td>• What causes X?</td>
</tr>
<tr>
<td>• What effect does X have on Y?</td>
</tr>
<tr>
<td>• Does X cause Y?</td>
</tr>
<tr>
<td>• Does X prevent Y?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Causality-Comparative</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Does X cause more Y than does Z?</td>
</tr>
<tr>
<td>• Is X better at preventing Y than is Z?</td>
</tr>
<tr>
<td>• Does X cause more Y than does Z under one condition but not others?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>• What is an effective way to achieve X?</td>
</tr>
<tr>
<td>• How can we improve X?</td>
</tr>
</tbody>
</table>

### Proposition (a.k.a. Purpose)

Where to expect boundaries?
⇒ Thorough preparation is necessary!
⇒ You need an explicit theory.

<table>
<thead>
<tr>
<th>Exploratory</th>
<th>Confirmatory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploratory case studies are used as initial investigations of some phenomena to derive new hypotheses and build theories.(*)</td>
<td></td>
</tr>
<tr>
<td>Confirmatory case studies are used to test existing theories. The latter are especially important for refuting theories: a detailed case study of a real situation in which a theory fails may be more convincing than failed experiments in the lab.(*)</td>
<td></td>
</tr>
</tbody>
</table>

1. Research Methods

Units of Analysis

What phenomena to analyze
- depends on research questions
- affects data collection & interpretation
- affects generalizability

Possibilities
- individual developer
- a team
- a decision
- a process
- a programming language
- a tool

Design in advance
- avoid "easy" units of analysis
  + cases restricted to Java because parser
    - Is the language really an issue for your research question?
  + report size of the system (KLOC, # Classes, # Bug reports)
    - Is team composition not more important?

Example: Clone Detection, Bug Prediction
- the tool/algorithim
  Does it work?
- the individual developer
  How/why does he produce bugs/clones?
- about the culture/process in the team
  How does the team prevent bugs/clones?
  How successful is this prevention?
- about the programming language
  How vulnerable is the programming language towards clones / bugs?
  (COBOL vs. AspectJ)

Threats to Validity (Experiments)

1. Conclusion validity
2. Internal validity
3. Construct validity
4. External validity

Threats to Validity — Examples (1/2)

1. Construct validity
- Do the operational measures reflect what the researcher had in mind?
- Time recorded vs. time spent
- Execution time, memory consumption, ...
  + noise of operating system, sampling method
- Human-assigned classifiers (bug severity, ...) + risk for "default" values
- Participants in interviews have pressure to answer positively

2. Internal validity
- Are there any other factors that may affect the results?
  ➤ Mainly when investigating causality!
- Were phenomena observed under special conditions
  + in the lab, close to a deadline, company risked bankruptcy, ...
  + major turnover in team, contributors changed (open-source), ...
- Similar observations repeated over time (learning effects)

Threats to Validity (Case Studies)

- Source: Runeson, P. and Höst, M. 2009. Guidelines for conducting and reporting case study research in software engineering.

1. Construct validity
- Do the operational measures reflect what the researcher had in mind?

2. Internal validity
- Are there any other factors that may affect the results?
  ➤ Mainly when investigating causality!

3. External validity
- To what extent can the findings be generalized?
  ➤ Precise research question & units of analysis required

4. Reliability
- To what extent is the data and the analysis dependent on the researcher (the instruments, ...)

Other categories have been proposed as well
- credibility, transferability, dependability, confirmability
1. Research Methods

Introduction

• Origins of Computer Science
• Research Philosophy

Research Methods

• 1. Feasibility study
• 2. Pilot Case
• 3. Comparative study
• 4. Observational Study [a.k.a. Etnography]
• 5. Literature survey
• 6. Formal Model
• 7. Simulation

Conclusion

• Studying a Case
• vs. Performing a Case Study
  + Proposition
  + Unit of Analysis
  + Threats to Validity

2. Propositions (a.k.a. Purpose)

• explanatory: where to look for evidence
• exploratory: rationale and direction
  + example: Christopher Columbus asks for sponsorship
  - Why three ships (not one, not five) ?
  - Why going westward (not south ?)
• role of "Theories"
  + possible explanations (how, why) for certain phenomena
    ➡ Obtained through literature survey

3. Unit(s) of analysis

• What is the case ?

4. Logic linking data to propositions

• Chain of evidence from multiple sources
• When does data confirm proposition ? When does it refute ?

5. Criteria for interpreting findings

• Chain of evidence from multiple sources
• When does data confirm proposition ? When does it refute ?

Threats to validity — Examples (2/2)

3. External validity
   • To what extent can the findings be generalized ?
   • Does it apply to other languages ? other sizes ? other domains ?
   • Background & education of participants
   • Simplicity & scale of the team
     + small teams & flexible roles vs. large organizations & fixed roles

4. Reliability
   • To what extent is the data and the analysis dependent on the researcher (the instruments, …)
   • How did you cope with bugs in the tool, the instrument ?
   • Classification: if others were to classify, would they obtain the same ?
   • How did you search for evidence in mailing archives, bug reports, …

Threats to validity = Risk Management

No experimental design can be “perfect”
... but you can limit the chance of deriving false conclusions

• manage the risk of false conclusions as much as possible
  + likelihood
  + impact
• state clearly what and how you alleviated the risk (replication !)
  + construct validity
    - precise metric definitions
    - GQM paradigm
  + internal & external validity
    - report the context consciously
  + Reliability
    - bugs in tools: testing, usage of well-known libraries, …
    - classification: develop guidelines & others repeat classification
    - search for evidence (mailing archives, bug reports, …):
      have an explicit search procedure