lero THE IRISH SOFTWARE ENGINEERING RESEARCH CENTRE

Gathering and Using Empirical Evidence for Software Architecture Research

Muhammad Ali Babar

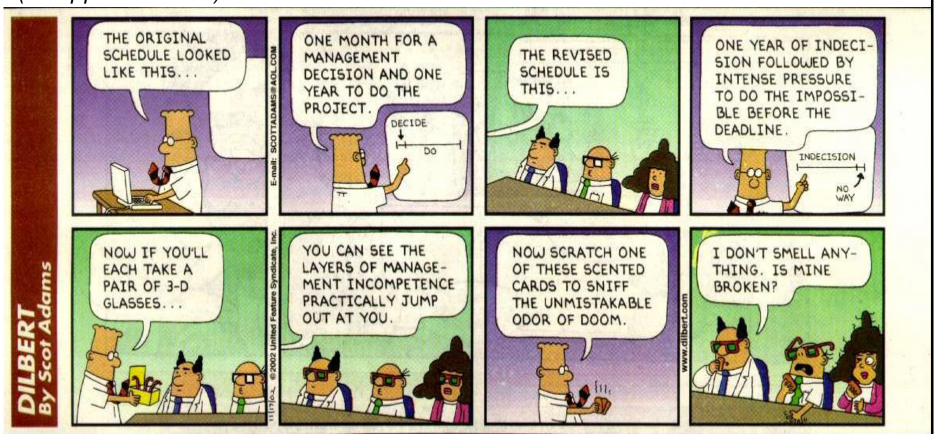
Outline

- Research context
- Motivation
- Different approaches in empirical research
- Snapshots of our use of different approaches
- Reflections on various aspects of empirical studies conducted by us
- Concluding remarks
- What are we doing now!

Designing Architecture is Hard!

"..The life of a software architect is a long (and sometimes painful) succession of sub-optimal decisions made partly in the dark…"

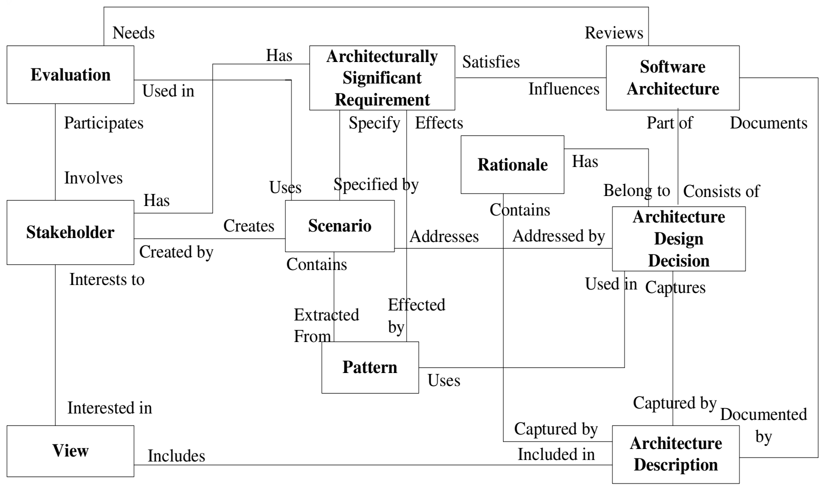
(Philippe Krutchen)



Why is Architecture Design Hard?

- Architecture design typically takes place at an early stage of the project
  - hard, often impossible, to thoroughly reason about the consequences of many design decisions
- Involves making design decisions that are difficult/costly to change downstream if they are discovered to be flawed
- Complex design trade-offs needed to meet competing architectural requirements
- Put very simply – architecture aims to address any issues that will be expensive/impossible to change once the project progresses

Contextualizing Research in Software Architecture



Motivation

- Software Architecture community has developed several methods, techniques, and tool to support the architecture process
- However, not much effort to rigorously assess the research output
- Anecdotal or “common-sense” based evidence is not good enough for supporting industrial decisions
- Systematically gathering and widely disseminating evidence is vital for successful technology transfer
- Empiricism provides scientifically valid approaches to systematically gather and use evidence
- Our goal is to develop and/or empirically assess various methods, techniques, and tools

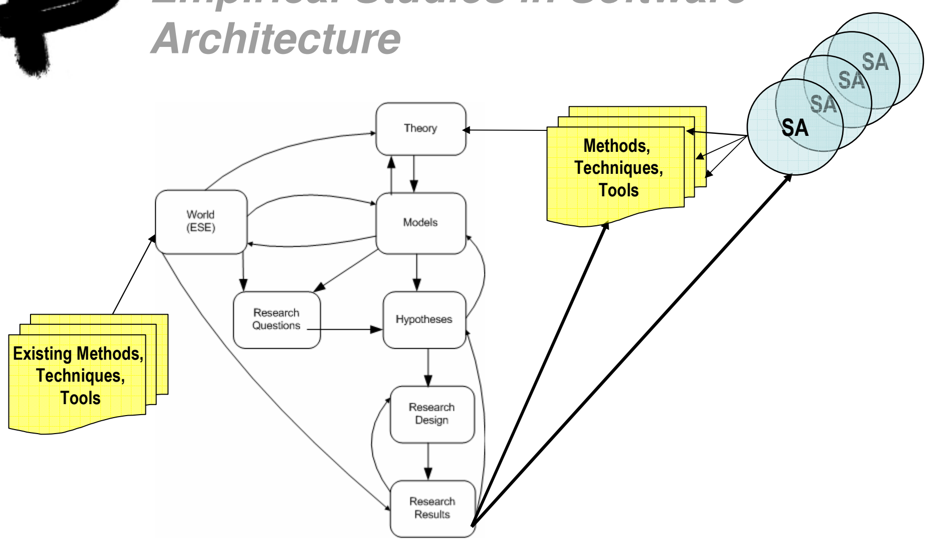
Empirical Research

- What is an empirical research?
  - Empirical research in software engineering is the scientific use of quantitative and qualitative data to understand and improve software product and software development process (Vic Basili).
  - Data is central element to address any research question
  - Issues related to the validity form the core
- Why is empirical research difficult?
  - Getting data describing real SE on realistic projects is hard
  - Involvement of people
  - Large undertaking in complex organizational settings
  - Research needs to be cost-justified & timely results
  - Non-standardized and rapidly changing technology

Main Research Approaches

- Controlled experiments
- Surveys – interviews and questionnaires
- Focus groups
- Expert opinion
- Observational study
- Case study
- Post mortem analysis
- Systematic reviews

Empirical Studies in Software Architecture



A Generic Process of Empirical Studies

- Definition
- Planning
- Operation
- Analysis and Interpretation
- Conclusions
- Presentation and Packaging

Methodological Support

- Methodological support is required to systemize architecture evaluation process
- Several methods have been proposed but no basis for a systematic comparison for selection
- We proposed FOCSAAM as comparative criteria based on 17 elements
- Rick et al also proposed a basis for comparison using 4 elements
- How to know whether or not elements provide a basis for comparing different methods?

Empirical Assessment

- Approach
  - Expert Opinion
- Participants selection
  - 10 architects with extensive experience in evaluation
- Instrument
  - Questionnaire based on FOCSAAM's elements
- Results
  - Seven elements were supported by all participants
  - Other elements were supported by a majority with some disagreements on the value of a few elements

Architecture Evaluation Practices

- No empirically found knowledge about the factors influencing industrial practices
- Exploring the perceptions and experiences of practitioners from different organizations
- Focus group discussion
  - Two sessions with 5 participants each
  - Facilitators experienced in the topic area
- Identified 19 factors and organized them under 5 categories
- Also identified the strategies to deal with the challenges posed by the identified factors

Observations

- Selection of expert is a challenge
- Seeking agreement can be frustrating
  - Personal contacts & organizational affiliation
- Sources of valid data are usually hard to find so plan to maximize each source
- We planned to use same participants for both expert opinion and factor identification studies
- Planning should be thorough and extensive
- Execution needs well developed skills in running discussion workshops
- Data transcription is laborious and time consuming
- Try to involve more than one young researchers

Stakeholders' Participation

Stakeholders' Participation

- Involvement of stakeholders in methods like ATAM, SAAM, ALMA
- Collocating stakeholders is time-consuming and expensive
- Little support to address issues like conformity pressures, dominating personalities and cultural differences
- An evaluation may involve 2-20 stakeholders

Proposed Solutions

- Groupware supported process
  - Effects of groupware use on the quality of the deliverables
  - Effects on the satisfaction of the users
  - Features required by a groupware
- Use of mixed mechanism – collocated teams and groupware supported teams
- Which size of group can be optimal?

Empirical Assessment – Groupware Support

- Cross-over controlled experimental design for assessing the effects of groupware use
- Two questionnaires used to study the effects on the satisfaction of the participants
- Modified version of TAM used for determining the features required of tool
- Results showed groupware supported meetings were superior to face-to-face meetings
- Participants preferred face-to-face meetings
- Identified features can be used for developing a tool

Empirical Assessment – Group Size

- This study investigated the impact of group size on the quality of process outcome
- A randomized experimental design with 165 participants assigned to groups of 3, 5, and 7
- All groups used same material and performed same task, developing scenario profiles
- Results showed a strong non-linear relationship between group size and performance
- Participants also favored smaller groups
- Results enabled us to propose a mixed means for architecture evaluation – face-to-face & groupware supported

Pattern-Based Knowledge

- We developed a mechanism for extracting (*pattern-mining)* architecturally significant information from patterns (*ASIP*)
  - General process model
    - Steps involved in mining patterns with guidelines
  - Templates to record
    - Generic architecture information
    - Project related information relating to concrete scenarios
- Intended to reduce the time and expertise needed to extract information
- To provide structured information to support tasks during architecture process

Three-Part Research Project

Time scale	Questionnaire for self-reported data	Purpose	Type of study
Week 6 of academic calendar	First part of the questionnaire	To find out an average time required to mine a pattern Feedback on the effectiveness of the process, guidelines, templates	Observational study
Week 10 of academic calendar	Second part of the questionnaire	Feedback on the usefulness of ASIP in creating scenarios	Controlled experiment
Week 14 of academic calendar	Third part of the questionnaire	Feedback on the usefulness of ASIP in design SA	Controlled experiment

Assessment of the Process & guidelines

- An observation study, with no control group and subjective & self-recorded data
  - After lecture on mining patterns
    - Students given a home-work exercise
      - Were required to record data about the exercise
  - Populated a template with architectural information from standard pattern documentation
    - 30-40 minutes approx
    - All participants thought the process and guidelines were easy to follow and ASIP would be useful in developing scenarios and understanding the use of pattern in architecture design and comprehension

Evidence of the Usefulness of ASIP in Scenarios Development

- A randomised experiment with treatment & control groups
  - After lecture on scenario development
    - Students participated in scenario development workshop
      - Students assigned at random to two groups
      - One group was given ASIP, second group was not
  - Participants with ASIP developed better quality scenarios ($p<0.023$).
  - Questionnaire based data revealed that participants found ASIP helpful in developing scenarios

Evidence of the Usefulness of ASIP in Architecture Design

- Two-part experiment
  - Understanding the use of patterns in a SA
  - Selecting suitable patterns for SA design
  - 20 participants used templates for one exercise & standard pattern documentation (J2EE design pattern) for the other
    - Students were assigned to two groups at ***random***
- Response variable
  - Quality of answer to SA question
  - Questionnaire based self-reported data
- Participants given ASIP were better able to identify patterns in a SA (Mann-Whitney, $p=0.0375$)
- Participants given ASIP were better able to select suitable patterns (Mann-Whitney, $p=0.035$)
- Comparing ASIP with standard pattern documentation
  - 18 (90%) thought ASIP better
  - 2 (10%) thought standard pattern info better

Observations

- Process assessment is hard
  - Need participants who are going to use it
  - Be prepare to provide extensive training
  - Attempt to minimize the biased feedback
- Experiments are hard to design, conduct, and analyze – be aware of the reported lessons
- Randomization may be problematic – report the randomization method
- Identify and deal with the validity threats
- Procedural compliance is hard – with volunteers it becomes hardest, so have monitoring mechanism
- A lot of effort required to design and validate data collection instruments
- Running a pilot study is always an advantage
- Report by following the standards or norms in the discipline but don't follow everything

Use and Documentation of Design Rationale for Maintenance

- Several papers reported the lack of design rationale documentation but provided no evidence except personal observations of the authors
- We decided to gather empirical evidence of the degree of use and documentation of design rationale
- Also intention was to explore practitioners' perception about the importance and usefulness of architectural design rationale for software maintenance
- Designed and executed an online survey
- There were 30 questions on the topic and 10 demographic related questions
- Using availability sampling invited 547 people

What Did We find?

- Based on 15% response rate
- Practitioners perceived design rationale as an important input to support maintenance and enhancement tasks
- Large number of respondents believe they document design rationale
- A positive perception bodes well for introducing systematic capture and use of DR
- Lack of Methodological and Tooling Support
- Further required research
  - Determine the level of documentation based on the context
  - Find the mechanics of cost/benefits of documenting rationale
  - Identify the types and amount of design rationale required for different types of systems

Managing Architectural Knowledge

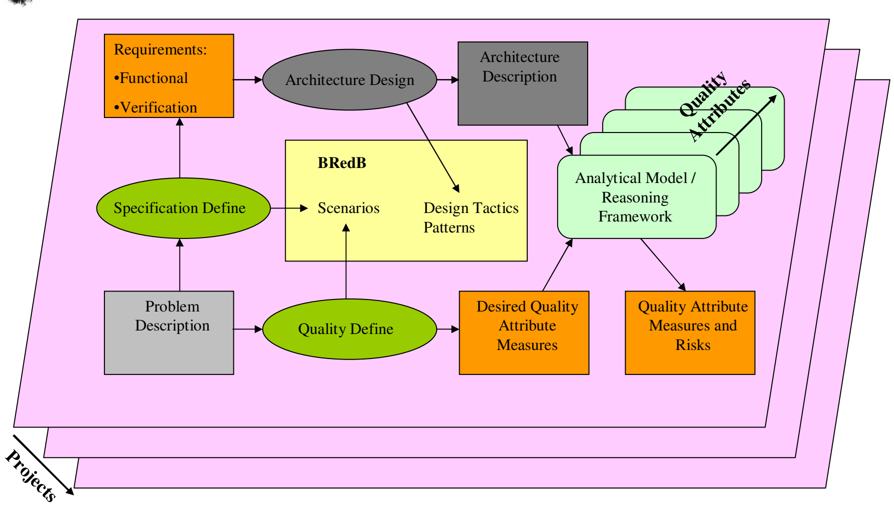
- An architecture embodies crucial design decisions
  - Rarely captured in architecture docs (if they exist!)
  - If rationale behind design decisions is lost:
- If rationale behind design decisions is lost:
  - System evolution becomes hard
  - Difficult to identify design errors
- Our survey on design rationale found:
    - 80% can't understand designs without adequate docs
    - 73% forget why they designed something!
  - Impediments to capturing design info:
    - 61% have no time/budget/tools
- We built BRedB

BRedB Case Study

- Australian Defense Science Technology Organization (DSTO)
  - Avionics architecture assessment
  - Long lived, high cost projects
- BRedB used for avionic architecture evaluation
  - Quality attributes
  - Evaluation framework
  - Risk management

LERO© 2006 THE IRISH SOFTWARE ENGINEERING RESEARCH CENTRE 29

BRedB Supported Architecture Evaluation Process



LERO© 2006 THE IRISH SOFTWARE ENGINEERING RESEARCH CENTRE 30

Initial Findings

BRedB added considerable rigour to the DSTO evaluation process

- Repository of expert knowledge in general scenarios
- Archive of past project experience valuable over time
- effective mechanism to organise and query large amounts of architecture knowledge

BRedB should be useful if an organization:

- Is outsourcing/off-shoring/purchasing its systems?
- has superhero architects prone to sudden departure or large pay increase demands

More studies and R&D needed

Some More Reflections

- Managing the team
- Study plan and approval
- Recruitment of participants
- Kinds of data to be collected
- Funding aspects
- Ethical issues
- Reporting the results

Managing the Team

- Identify stakeholders and their respective stakes
- Communicate clearly the goals, procedures, and expected outcomes
- Try to raise stakes of all team members
- Keep sponsors posted on issues
- Identify and resolve conflict swiftly
- Sometime resolution may not be possible – peaceful separation!

Collaborators

- Barbara Kitchenham - Keele University, UK
- Stefan Biffl - Vienna Technical University, Austria
- Antony Tang and Jun Han – Swinburne University of Technology Melbourne
- Len Bass, SEI, CMU
- Ian Gorton, PNNL, USA
- Torgeir Dingsoyr - SINTEF, Norway
- Odd Petter – NTNU, Norway
- Makoto Nonaka - Toyo University, Japan

Study Planning, Assessing, Executing

- Plan carefully and precisely document each aspect of the study as a protocol document.
- Describe the process and rationale of preparing each artifact, e.g., documents, questionnaire etc.
- Anticipate all possible scenarios when things can go wrong and have alternative plans
- Provide extensive briefing to all team members
- Run a formal pilot study or at least informal assessment by independent researchers
- Ensure study is executed according to the protocols
- Take notes of each event that goes wrong
- Be available to monitor and support executing team

Approval for Ethics

- Don't underestimate this step – it can be frustrating
- Prepare your application very carefully and thoroughly
- Application may be assessed using procedures developed for other disciplines like biomedicine
- Assessor may not be able to appreciate the mechanics of empirical research in software engineering
- So make sure to highlight that no one would become "psychopath" or "Disable" as a result of your study – But in carefully chosen words
- Include the examples of the tasks to be completed
- Highlight the benefits to the participants
- Emphasis the value of the research outcome to the community
- Get your application to the approving body well in advance

Recruitment of Participants

- Academic courses – But consider all drawbacks carefully
- Personal contacts – can't push too much
- Through professional bodies
- Industry collaborators
- Practitioners in return for training or nominal remuneration
- Be proactive and innovative but ethical

Kinds of Data

- Quantitative data
- Qualitative
  - Self-reported
  - Field notes
- Demographic data
- More is better than less but be mindful of the participants' time and effort
- We used three questionnaires with one experiment – all were sources of our additional analysis of the experimental data

Funding Issues

- Getting funding for empirical studies is always hard
- Provide cogent arguments and highlight significance
- Use appropriate terms – data entry / not analysis
- Prepare to do extra work or have a support network

Item	Task	Estimated hours	Rate/hour	Cost/item
1	GroupSize study	5	$25.78	128.9
2	GroupSize study	25	$25.78	644.5
3	Distributed SA evaluation study	10	$25.78	257.8
4	Distributed SA evaluation study	25	$25.78	644.5
5	Cross over phase	25	$25.78	644.5
6	Overall questionnaire	8	$25.78	206.24
	Experiment Data entry	**98**	**$25.78**	**2526.44**

Ethical Principles

- Informed consent
  - Disclosure
  - Comprehension and competence
  - Voluntariness
  - Right to withdraw
- Scientific value
  - Importance of research topic
  - Validity of research results
- Beneficence – Risk/Benefit Ratio
- Confidentiality
  - Anonymity
  - Confidentiality of data

Some of the Outcomes

- Research
  - 3 journal papaers
  - 14 conference papers
  - 1 workshop paper
- Education
  - Outcomes incorporated into course on software architecture and professional training
- Commercialisation
  - Development of commercially viable tool - BRedB
- Collaboration
  - Industrial partners - $A100,000 cash contribution
  - Participants' organisations

In Nutshell

- Empirical research consumes extensive human, physical, and material resources
- Needs skills and good understanding of several disciplines, e.g., stat, sociology etc. Require interdisciplinary collaboration
- Resource utilization needs careful thinking
- Interest in people is vital
- Appreciation isn't wide spread yet!
- But its rewarding and fun!

Current and Future Work

- Empirical Studies in Software Product Lines

Project Brief

- Project aim:
  - This project aims to carry out empirical research to gather and use evidence for identifying relevant research issues in SPL and rigorously assessing research outcomes.
- Project objectives:
  - Bridge the gap between research and practice in SPL area
  - Help achieve balance between rigor and relevance
  - Improve decision making in technology selection
  - Help build empirically assessed BOK about SPL practices
  - Train researchers and practitioners in empirical studies

Project Context

- Within SPL:
  - SPL 1: Visualization Expertise
  - SPL 2: Conformance Analysis (in Product Derivation)
  - SPL 3: Model-Based Product Derivation
  - SPL and Agile Approaches
- Outside of SPL:
  - GSD: SPL challenges in GSD
  - Open source: Challenges and opportunities for using FLOSS components in family of products
  - Empirical evaluation as a central Lero service

Potential Researchable Topics

- **Mechanics and approaches to calculating ROI**
- **Factors critical to successful SPL adoption**
- **Measuring the economic effects of SPL**
- **Current practices and tooling support in SPL**
- **Empirical evidence for organizational alternatives and key practice areas**
- **Relationship between CMMI and SPL**
- **Requirements for tooling to improve SPL practices and how to provide that support**
- **Integrating Agile practices in SPL**
- **Knowledge and experience management**

Some good references

- Experimentation in Software Engineering, Basili et al.
- Preliminary guidelines for empirical research in software engineering, Kitchenham et al.
- Studying Software Engineers: Data Collection Techniques for Software Field Studies, Lethbridge et al.
- Ethical Issues in Empirical Studies of Software Engineering, Janice Singer & Norman Vinson
- Qualitative Methods in Empirical Studies of Software Engineering, Carolyn B. Seaman
- Pfleeger & Kitchenham's series of papers in Softare engineering notes – 1995-96
- Evidence Based Software Engineering, Kitchanham et al.

Acknowledgements

- All these studies were designed and conducted while I was with NICTA
- Barbara Kitchenham provided extensive consultation on most of these studies
- Cynthia Wang helped us to prepare the data for analysis for several studies
- A big thank to all the participants of these studies

lero

THE IRISH SOFTWARE ENGINEERING RESEARCH CENTRE

Thank You

My Questions!

- Do you know Lero is recruiting? Researchers & Ph.D. students
- Do you know Lero has no competitor in Ireland?
- Do you know Lero is the only academic member on NESSI's board?

Get on board and build your future
Check out www.lero.ie/jobs

Lero – The Irish Software Engineering Research Centre