Aggregating Empirical Evidence about the Benefits and Drawbacks of Software Reference Architectures

S. Martínez-Fernández, P. S. Medeiros dos Santos, C. P. Ayala, X. Franch, G. H. Travassos

9th ACM/IEEE ESEM, Beijing (China), October 22-23, 2015
Motivation: problem

Can you think of a limitation of single empirical studies?

A view from selected reviewers’ comments on our case studies

"The authors [...] do not attempt to generalize their results – this makes the paper uninteresting"

ECSA 2013

"All the stakeholders involved in the empirical study are from the same company. [...] The results cannot be generalized beyond the specific company and across different SRAs"

ICSE 2014
Motivation: problem

"My last concern is about the applicability/generalizability of the results. The results may be biased by how the organization interprets the concept and the role that SRAs play."

EASE 2014

"The authors rightly point out the limitations of their work, which are based on a single organization and therefore may not be generalizable."

TOSEM 2015
Motivation: problem

How can we generalize the results from single empirical studies?

• The results are tied to the context of the single empirical study
• Even if we define well the context, it is impossible to generalize without further work
The good news

- Available evidence from single primary empirical studies is increasing!
  - Other studies gathering evidence in many contexts
- Heterogeneity of studies in Software Engineering
  - Many strategies, e.g., surveys, case studies, and experiments
Main goal

• Research synthesis and aggregation
  - *What?* Benefits and drawbacks of Software Reference Architectures (SRA)
  - *Why?* To evaluate the adoption of SRAs
  - *How?* The Structured Synthesis Method (SSM)
What? our research interest

• Evidence on Software Reference Architectures
  ▪ “an architecture that encompasses the knowledge about how to design concrete architectures of systems of a given application domain” [Nakagawa et al. 2011]
Why? Decide on SRA acquisition

• Benefits and drawbacks of acquiring SRAs to construct a family of software systems are essential to make informed decisions whether or not to adopt an SRA in an organization(s)

• Types of SRAs:
  ▪ to standardize / to facilitate
  ▪ for a single organization / for multiple organizations

• Example: AUTOSAR
How? Why the SSM?

- Organization and development of concepts to describe contextual aspects (integrative)
- Cause-effect & moderation relations (interpretive)
- Qualitative & quantitative evidence
- Uncertainty estimated for each evidence
  - Heterogeneity of studies
- Diagrammatic models → tend to improve comprehension
- Tool support: evidencefactory.lens-ese.cos.ufrj.br
The rest of the talk

- How we applied the SSM to aggregate evidence of SRAs?
  - Step 1: selecting studies
  - Step 2: extraction & representation
  - Step 3: aggregation & analysis
## Step 1 – five studies

<table>
<thead>
<tr>
<th>Id.</th>
<th>Study Type: Instruments</th>
<th>Participants</th>
<th>SRA Application Domain</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Expert meeting: presentations, discussions</td>
<td>Architects from the System Architecture Forum</td>
<td>Defense and commercial equipment</td>
<td>2010</td>
</tr>
<tr>
<td>S2</td>
<td>Case study: interviews, questionnaires, docs.</td>
<td>28 sw. architects and developers from IT consulting</td>
<td>Banks, insurers, public administration, utilities, and industries</td>
<td>2013</td>
</tr>
<tr>
<td>S3</td>
<td>Survey: questionnaires</td>
<td>90 sw. architects and developers from worldwide</td>
<td>n/a</td>
<td>2013</td>
</tr>
<tr>
<td>S4</td>
<td>Case study: interviews, docs., meetings</td>
<td>20 sw. architects, managers and experts from local e-government</td>
<td>Variability-intensive service-oriented systems</td>
<td>2013</td>
</tr>
<tr>
<td>S5</td>
<td>Survey: questionnaires</td>
<td>51 practitioners from AUTOSAR partners</td>
<td>Automotive systems</td>
<td>2015</td>
</tr>
</tbody>
</table>
Step 2 – quality evaluation

• Four subranges according to study type:
  - unsystematic observations [0.00, 0.25]
  - observational studies [0.25, 0.50]
  - quasi-experiments [0.50, 0.75]
  - randomized controlled [0.75, 1]

• Quality assessment
  - Two checklist to assess the quality of each study → 0.25 subrange

<table>
<thead>
<tr>
<th>Id.</th>
<th>Belief</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.25+0.10=0.35</td>
</tr>
<tr>
<td>S2</td>
<td>0.25+0.19=0.44</td>
</tr>
<tr>
<td>S3</td>
<td>0.25+0.15=0.40</td>
</tr>
<tr>
<td>S4</td>
<td>0.25+0.15=0.40</td>
</tr>
<tr>
<td>S5</td>
<td>0.25+0.17=0.42</td>
</tr>
</tbody>
</table>
Step 2 – representing evidence

Interoperability was highlighted as another major motivation for using Reference Architectures. The Reference Architecture was aimed at interoperability to improve compliance for a given context.

"Study 1"
Step 2 – representing evidence

Reuse and commonality [of the Reference Architecture] throughout product generation. The potential benefit of this being shorter development cycles and reduced cost from not having to start from scratch.

Study 1
Step 2 – entire evidence model

One evidence model for each paper → 5 models in total
Step 3 – aggregation

• Aggregation Conflict Resolution
  ▪ Three options: Add/Remove/Join
    • Add: reputation, which is only in S5
    • Remove: benefits of software vendors of S2
    • Join: SRAs for different contexts, interoperability or development costs from many studies

• It applies the *Dempster’s Rule of Combination*
# Common effects in many contexts

<table>
<thead>
<tr>
<th>Study</th>
<th>Representation evidence from studies, shown as: intensity (belief value)</th>
<th>Aggregation Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effect</strong></td>
<td># Papers</td>
<td>Intensity</td>
</tr>
<tr>
<td>Inter-operability</td>
<td>S1</td>
<td>S2</td>
</tr>
<tr>
<td>Development costs</td>
<td>PO, SP (0.35)</td>
<td>PO (0.15)</td>
</tr>
<tr>
<td>Communication</td>
<td>PO (0.35)</td>
<td>PO (0.09)</td>
</tr>
<tr>
<td>Learning curve</td>
<td>SN, NE (0.36)</td>
<td>NE (0.13)</td>
</tr>
</tbody>
</table>
Take away #1

• SRA effects present in different contexts:
  ▪ improved interoperability
  ▪ reduced development costs
  ▪ better communication
  ▪ higher learning curve
# Effects without significant changes

## Aggregating Empirical Evidence

### Representation evidence, shown as: intensity (belief value)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Study</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th># Papers</th>
<th>Intensity</th>
<th>Belief</th>
<th>Conflict</th>
<th>Difference^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintenance cost</td>
<td></td>
<td>PO (0.35)</td>
<td></td>
<td></td>
<td></td>
<td>PO (0.14)</td>
<td>2</td>
<td>PO</td>
<td>44%</td>
<td>-</td>
<td>9%</td>
</tr>
<tr>
<td>Productivity</td>
<td></td>
<td>PO, SP  (0.30)</td>
<td></td>
<td></td>
<td></td>
<td>PO (0.11)</td>
<td>2</td>
<td>PO</td>
<td>38%</td>
<td>-</td>
<td>8%</td>
</tr>
<tr>
<td>Ease of developing</td>
<td></td>
<td>PO (0.30)</td>
<td>PO (0.07)</td>
<td></td>
<td>WP, PO  (0.03)</td>
<td></td>
<td>3</td>
<td>PO</td>
<td>35%</td>
<td>-</td>
<td>5%</td>
</tr>
<tr>
<td>Alignment</td>
<td></td>
<td>WP, PO  (0.19)</td>
<td></td>
<td></td>
<td>WP (0.07)</td>
<td></td>
<td>2</td>
<td>WP, PO</td>
<td>24%</td>
<td>-</td>
<td>5%</td>
</tr>
<tr>
<td>Restriction</td>
<td></td>
<td>NE (0.13)</td>
<td>NE (0.06)</td>
<td></td>
<td>NE, WN  (0.07)</td>
<td></td>
<td>3</td>
<td>NE</td>
<td>18%</td>
<td>-</td>
<td>5%</td>
</tr>
</tbody>
</table>
Take away #2

- Effects without significant changes in their belief value after the aggregation:
  - maintenance cost, productivity, ease of developing, alignment, restriction, standardization, latest technologies, investment, reliability, dependability, reputation, software quality, novel design solution
- These effects are relevant topics that need to be further studied → need to look for explanations
## Effects in certain contexts

**Study**  | **Representation evidence from studies, shown as: intensity (belief value)**  | **Aggregation Results**
|----------|--------------------------------------------------------------------------------|--------------------------
| Effect   |                                                                              | # Papers | Intensity | Belief | Conflict | Difference^a |
| Complexity | S1: WN (0.06)  | S2: SN, NE (0.27)  | S3  | S4  | S5  | 2 | SN, NE | 26% | 0.017 | -1% |
| Terminology conventions | S1: WN, IF (0.40)  | S2: WP, PO (0.35)  | S3  | S4  | S5  | 2 | WP, PO | 31% | 0.060 | -4% |
| Flexibility of suppliers | S1: WN, IF (0.40)  | S2: WP, PO (0.35)  | S3  | S4  | S5  | 2 | WN, IF | 31% | 0.140 | -9% |
Take away #3

- Three effects of SRAs that were only present under certain contexts of the empirical studies:
  - SRA complexity
  - terminology conventions
  - flexibility of suppliers

- Considering the differences, we looked for explanations and made hypothesis
Some threats to validity

- Five studies $\rightarrow$ more studies are needed to reach definitive results
- Evidence modeling / extraction done by two researchers
- Glossary of terms $\rightarrow$ useful when studies refer to the same concept with different terms
- The SSM does not consider size of sampling
  - Use discount on surveys
Conclusion: take away #4

• Aggregating evidence helps to:
  ▪ Increase the confidence of the results of single empirical studies
    • Possibility to formulate new theories
  ▪ Reduce the effort of people interested in a topic
Thank you for your attention!
Comments and Questions

New aggregations? 😊 Use the tool!
evidencefactory.lens-ese.cos.ufrj.br

Silverio Martínez-Fernández
smartinez@essi.upc.edu
@silveriomf
The Dempster’s Rule of Combination

- the aggregated belief value for each hypothesis $C$ is equal to the sum of the product of the hypotheses belief values whose intersection between all hypotheses $A_i$ and $B_j$ of both evidence is $C$

$$m_3(C) = \sum_{A_i \cap B_j = C} m_1(A_i) \times m_2(B_j)$$

$$m_3(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) \times m_2(B_j)}{1 - K}, \text{ where } K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) \times m_2(B_j)$$
<table>
<thead>
<tr>
<th>Study</th>
<th>Representation evidence from studies, shown as: intensity (belief value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
<td></td>
</tr>
<tr>
<td>$m_1$</td>
<td>$N(0.13)$, $\Theta(0.87)$</td>
</tr>
<tr>
<td>$m_2$</td>
<td>${N^2, 0.0468}$</td>
</tr>
<tr>
<td>$m_3$</td>
<td>${SN, N^2}$, $0.3132$</td>
</tr>
<tr>
<td>$m_4$</td>
<td>$0.0832$, $0.5368$</td>
</tr>
<tr>
<td>$m_5$</td>
<td>$\Theta$, $\Theta$</td>
</tr>
<tr>
<td>$m_6$</td>
<td></td>
</tr>
<tr>
<td>Learning curve</td>
<td></td>
</tr>
<tr>
<td>$m_3(N) = 0.0468 + 0.0832 = 0.13$</td>
<td></td>
</tr>
<tr>
<td>$m_3{SN, N^2} = 0.3132$</td>
<td></td>
</tr>
<tr>
<td>Bel($N^2$)</td>
<td>0.93</td>
</tr>
<tr>
<td>Bel(${SN, N^2}$)</td>
<td>0.13 + 0.3132 = 0.44</td>
</tr>
</tbody>
</table>
Tool support for aggregation

Relationship history

<table>
<thead>
<tr>
<th>Proposition {PO, SP} with probability of 0.258</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposition {SN, NE, WN, IF, WP, PO, SP} with probability of 0.483</td>
</tr>
<tr>
<td>Proposition {PO} with probability of 0.259</td>
</tr>
<tr>
<td>Conflict Level = 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposition {PO, SP} with probability of 0.300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposition {SN, NE, WN, IF, WP, PO, SP} with probability of 0.561</td>
</tr>
<tr>
<td>Proposition {PO} with probability of 0.140</td>
</tr>
<tr>
<td>Conflict Level = 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposition {SN, NE, WN, IF, WP, PO, SP} with probability of 0.860</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposition {PO} with probability of 0.140</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposition {SN, NE, WN, IF, WP, PO, SP} with probability of 0.652</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposition {PO, SP} with probability of 0.348</td>
</tr>
</tbody>
</table>
Join example

Evidence based on the paper "A Survey on the Benefits and Drawbacks of AUTOSAR"

Evidence based on the paper "The Concept of Reference Architectures"

System
- Automotive Software (1)
  - Software Quality
  - Standardization
  - Interoperability
- Enterprise Software (2)
  - Interoperability

Technology
- Reference Architecture
  - AUTOSAR (1)

Evidence based on the paper "A Survey on the Benefits and Drawbacks of AUTOSAR"

Evidence based on the paper "The Concept of Reference Architectures"
Add example
Remove example

Evidence based on the paper "Benefits and Drawbacks of Reference Architectures"

Evidence based on the paper "Software Reference Architectures - Exploring Their Usage and Design in Practice"
Step 1 – selecting the studies

- Through a Systematic Review
  - Search string:
    - (“reference architecture?”)
    - AND
    - (“software architecture?” or “software structure?” or “software design?” or “system architecture?” or “system structure?” or “system design?”)
  - Inclusion criteria: any empirical study reporting findings based on evidence about the benefits and drawbacks of adopting an SRA