The Shared Language of Crowds:
A Crowd-Sourced Approach to Mapping RSE and SER Terminology

Timo Kehrer
University of Bern

Robert Haines
University of Manchester

Guido Juckeland
Helmholtz-Zentrum Dresden

Shurui Zhou
University of Toronto

David E. Bernholdt
Oak Ridge National Laboratory

The Origin: Dagstuhl Seminar 24161

Research Software Engineering: Bridging Knowledge Gaps

(Apr 14 - Apr 19, 2024)

Organizers:

- **Stephan Druskat** (German Aerospace Center (DLR), Berlin)
- **Lars Grunske** (HU Berlin)
- **Caroline Jay** (University of Manchester)
- **Daniel S. Katz** (University of Illinois Urbana-Champaign)

The Origin: Dagstuhl Seminar 24161




Working Group: Developing a Common Language

Members:

- **David E. Bernholdt** (Oak Ridge National Laboratory)
- **Robert Haines** (University of Manchester),
- **Guido Juckeland** (Helmholtz-Zentrum Dresden-Rossendorf),
- **Timo Kehrer** (Universität Bern)
- **Shurui Zhou** (University of Toronto)

Motivation and Goals

Software Engineering Research (SER)

Systematic investigation of SE methods, tools, processes, practices

SE Fundamentals

Research Software Engineering (RSE)

Informal understanding of SE fundamentals through practical experience of creating research software

RSE Practices

Motivation and Goals

Software Engineering Research (SER)

Systematic investigation of SE methods, tools, processes, practices

SE Fundamentals

Research Software Engineering (RSE)

Informal understanding of SE fundamentals through practical experience of creating research software

RSE Practices

- Addressing inconsistent terminology for shared concepts.
- Establishing a shared understanding of SE fundamentals.
- Improving awareness and knowledge transfer across communities.

Emerging Questions from Initial Discussions

Emerging Questions from Initial Discussions

Scope & Boundaries

- What is the scope, where to draw boundaries?

Emerging Questions from Initial Discussions

Scope & Boundaries

- What is the scope, where to draw boundaries?

Fundamentals & Core Concepts

- What counts as fundamental?

Emerging Questions from Initial Discussions

Scope & Boundaries

- What is the scope, where to draw boundaries?

Fundamentals & Core Concepts

- What counts as fundamental?

Abstraction & Granularity

- What level of granularity is appropriate?

Emerging Questions from Initial Discussions

Scope & Boundaries

- What is the scope, where to draw boundaries?

Fundamentals & Core Concepts

- What counts as fundamental?

Abstraction & Granularity

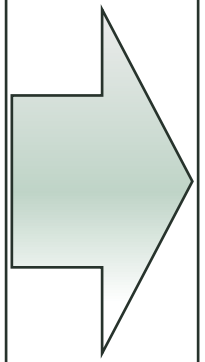
- What level of granularity is appropriate?

Consensus & Alignment

- How to reach consensus?

Approach

- Fundamentals & Core Concepts?
- Scope & Boundaries?
- Abstraction & Granularity?
- Consensus & Alignment?



- ➢ Systematic mapping based on SWEBOK as reference
- ➢ Crowdsourced collection across RSE and SER
- ➢ Web-based platform for community contributions
- ➢ Interactive discussion forum for qualitative insights

Starting Point: SWEBOK 4.0

Section	Heading	Page	Mapping
01	**Software Requirements**	**1-1**	
01.01	Software Requirements Fundamentals	1-2	
01.01.01	Definition of a Software Requirement	1-2	
01.01.02	Categories of Software Requirements	1-3	

Section	Heading	Page	Mapping
05.02	Test Levels	5-6	
05.02.01	The Target of the Test	5-6	
05.02.01.01	Unit Testing	5-6	mapping
05.02.01.02	Integration Testing	5-7	mapping
05.02.01.03	System Testing	5-7	mapping
05.02.01.04	Acceptance Testing	5-7	mapping
05.02.02	Objectives of Testing	5-7	
05.02.02.01	Conformance Testing	5-7	
05.02.02.02	Compliance Testing	5-8	
05.02.02.03	Installation Testing	5-8	mapping
05.02.02.04	Alpha and Beta Testing	5-8	mapping
05.02.02.05	Regression Testing	5-8	mapping

Term Mapping Schema

```yaml
# array of synonyms or adjacent concepts
se_fundamental:

# Short (1 sentence) description of the SE fundamental
fundamental_description:

# array of section identifiers
swebok_section:

# array of synonyms or adjacent concepts
rse_concept:

# Text, a brief description of the typical realizations of the fundamental, in RSE practice
rse_practice:

# General level of awareness of the fundamental in the research software community
# integers 0–3, 0=effectively no awareness, 3=widespread awareness
rse_awareness:
rse_awareness_source:

# General level of usage of the fundamental in the research software community
# integers 0–3, 0=effectively no usage, 3=widespread use
rse_usage:
rse_usage_source:

# Potential for SE research to improve use in research software
# integers 0–3, 0=effectively no opportunity, 3=significant SE research beneficial
ser_potential:
ser_potential_source:

# Reasons/opportunities for the SE research
ser_opportunities:

# References (external links, papers, etc., that may provide useful connections)
references:

# Date of last review by the editorial board (YYYY-MM-DD)
last_reviewed:
```

Term: The oracle problem

Source: SWEBOK section 05.01.02.07

SE Fundamental	RSE Equivalent
• The oracle problem	• Checking with the science lead • Checking prior papers • Gold result

Fundamental description	RSE practice
An oracle can be any human or mechanical agent that decides whether the system under test behaved correctly. If the oracle cannot decide, the test output is classified as inconclusive.	RSEs are not generally aware of the the SER terminology, but do practice the oracle concept. Often through an appeal to expert knowledge or prior results.

SER potential (source) and opportunities	RSE awareness (source) and usage (sources)
3 (expert judgement)	1 (expert judgement)
RSEs *should* know the term 'the oracle problem' and should be thinking about it more specifically as they work. The oracle problem will be increasingly important/challenging in AI-related settings, for example. Strategies for dealing with inconclusive outcomes are also of increasing importance.	3 (expert judgement)

Term Mapping Schema

```
# array of synonyms or adjacent concepts
se_fundamental:

# Short (1 sentence) description of the SE fundamental
fundam

# arra
swebo

# arra
rse_c

# Tex                                                    tice
rse_pr

# General level of awareness of the fundamental in the research software community
# integers 0–3, 0=effectively no awareness, 3=widespread awareness
rse_awareness:
rse_awareness_source:

# General level of usage of the fundamental in the research software community
# integers 0–3, 0=effectively no usage, 3=widespread use
rse_usage:
rse_usage_source:

# Potential for SE research to improve use in research software
# integers 0–3, 0=effectively no opportunity, 3=significant SE research beneficial
ser_potential:
ser_potential_source:

# Reasons/opportunities for the SE research
ser_opportunities:

# References (external links, papers, etc., that may provide useful connections)
references:

# Date of last review by the editorial board (YYYY-MM-DD)
last_reviewed:
```

Source (e.g., SWEBOOK Section)

Term: The oracle problem

Source: SWEBOK section 05.01.02.07

SE Fundamental	RSE Equivalent
• The oracle problem	• Checking with the science lead • Checking prior papers • Gold result

Fundamental description	RSE practice
An oracle can be any human or mechanical agent that decides whether the system under test behaved correctly. If the oracle cannot decide, the test output is classified as inconclusive.	RSEs are not generally aware of the the SER terminology, but do practice the oracle concept. Often through an appeal to expert knowledge or prior results.

SER potential (source) and opportunities	RSE awareness (source) and usage (sources)
3 (expert judgement)	1 (expert judgement)
RSEs *should* know the term 'the oracle problem' and should be thinking about it more specifically as they work. The oracle problem will be increasingly important/challenging in AI-related settings, for example. Strategies for dealing with inconclusive outcomes are also of increasing importance.	3 (expert judgement)

Term Mapping Schema

```yaml
# array of synonyms or adjacent concepts
se_fundamental:

# Short (1 sentence) description of the SE fundamental
fundamental_description:

# array
swebo

# arra
rse_c

# Tex                                                    n RSE practice
rse_p

# Ger                                                    munity
# inte
rse_aw
rse_awareness_source:

# General level of usage of the fundamental in the research software community
# integers 0–3, 0=effectively no usage, 3=widespread use
rse_usage:
rse_usage_source:

# Potential for SE research to improve use in research software
# integers 0–3, 0=effectively no opportunity, 3=significant SE research beneficial
ser_potential:
ser_potential_source:

# Reasons/opportunities for the SE research
ser_opportunities:

# References (external links, papers, etc., that may provide useful connections)
references:

# Date of last review by the editorial board (YYYY-MM-DD)
last_reviewed:
```

Collection of synonyms or adjacent concepts

Term: The oracle problem

Source: SWEBOK section 05.01.02.07

SE Fundamental	RSE Equivalent
• The oracle problem	• Checking with the science lead • Checking prior papers • Gold result

Fundamental description	RSE practice
An oracle can be any human or mechanical agent that decides whether the system under test behaved correctly. If the oracle cannot decide, the test output is classified as inconclusive.	RSEs are not generally aware of the the SER terminology, but do practice the oracle concept. Often through an appeal to expert knowledge or prior results.

SER potential (source) and opportunities	RSE awareness (source) and usage (sources)
3 (expert judgement)	1 (expert judgement)
RSEs *should* know the term 'the oracle problem' and should be thinking about it more specifically as they work. The oracle problem will be increasingly important/challenging in AI-related settings, for example. Strategies for dealing with inconclusive outcomes are also of increasing importance.	3 (expert judgement)

Term Mapping Schema

```yaml
# array of synonyms or adjacent concepts
se_fundamental:

# Short (1 sentence) description of the SE fundamental
fundamental_description:

# array of section identifiers
swebok_section:

# array of synonyms or adjacent concepts
rse_concept:

# Text                                                    ctice
rse_p

# Ger
# inte
rse_av
rse_av

# Ger
# inte
rse_us
rse_usage_source:

# Potential for SE research to improve use in research software
# integers 0–3, 0=effectively no opportunity, 3=significant SE research beneficial
ser_potential:
ser_potential_source:

# Reasons/opportunities for the SE research
ser_opportunities:

# References (external links, papers, etc., that may provide useful connections)
references:

# Date of last review by the editorial board (YYYY-MM-DD)
last_reviewed:
```

Definition vs. typical realization

Term: The oracle problem

Source: SWEBOK section 05.01.02.07

SE Fundamental	RSE Equivalent
• The oracle problem	• Checking with the science lead • Checking prior papers • Gold result

Fundamental description	RSE practice
…can be any human or mechanical agent that decides whether the system under test behaved correctly. If the oracle cannot decide, the test output is classified as inconclusive.	RSEs are not generally aware of the the SER terminology, but do practice the oracle concept. Often through an appeal to expert knowledge or prior results.

SER potential (source) and opportunities	RSE awareness (source) and usage (sources)
3 (expert judgement)	1 (expert judgement)
RSEs *should* know the term 'the oracle problem' and should be thinking about it more specifically as they work. The oracle problem will be increasingly important/challenging in AI-related settings, for example. Strategies for dealing with inconclusive outcomes are also of increasing importance.	3 (expert judgement)

Term Mapping Schema

```yaml
# array of synonyms or adjacent concepts
se_fundamental:

# Short (1 sentence) description of the SE fundamental
fundamental_description:

# array of section identifiers
swebok_section:

# array of synonyms or adjacent concepts
rse_concept:

# Text, a brief description of the typical realizations of the fundamental, in RSE practice
rse_practice:

# General level of awareness of the fundamental in the research software community
# integers 0–3, 0=effectively no awareness, 3=widespread awareness
rse_awareness:
rse_awareness_source:

# General level of usage of the fundamental in the research software community
# integers 0–3, 0=effectively no usage, 3=widespread use
rse_usage:
rse_usa

# Pot
# inte
ser_p
ser_p

# Rea
ser_o

# Ref
referen

# Date of last review by the editorial board (YYYY-MM-DD)
last_reviewed:
```

General level of awareness

- 0=effectively no awareness;
- 3=widespread awareness

Term: The oracle problem

Source: SWEBOK section 05.01.02.07

SE Fundamental	RSE Equivalent
• The oracle problem	• Checking with the science lead • Checking prior papers • Gold result

Fundamental description	RSE practice
An oracle can be any human or mechanical agent that decides whether the system under test behaved correctly. If the oracle cannot decide, the test output is classified as inconclusive.	RSEs are not generally aware of the the SER terminology, but do practice the oracle concept. Often through an appeal to expert knowledge or prior results.
SER potential (source) and opportunities	**RSE awareness (source) and usage (sources)**
[illegible]	1 (expert judgement)
RSEs *should* know the term 'the oracle problem' and should be thinking about it more specifically as they work. The oracle problem will be increasingly important/challenging in AI-related settings, for example. Strategies for dealing with inconclusive outcomes are also of increasing importance.	3 (expert judgement)

Term Mapping Schema

```yaml
# array of synonyms or adjacent concepts
se_fundamental:

# Short (1 sentence) description of the SE fundamental
fundamental_description:

# array of section identifiers
swebok_section:

# array of synonyms or adjacent concepts
rse_concept:

# Text, a brief description of the typical realizations of the fundamental, in RSE practice
rse_practice:

# General level of awareness of the fundamental in the research software community
# integers 0–3, 0=effectively no awareness, 3=widespread awareness
rse_awareness:
rse_awareness_source:

# General level of usage of the fundamental in the research software community
# integers 0–3, 0=effectively no usage, 3=widespread use
rse_usage:
rse_usa

# Pot
# inte
ser_p
ser_p

# Rea
ser_o

# Ref
referen

# Date of last review by the editorial board (YYYY-MM-DD)
last_reviewed:
```

General level of usage

- 0=effectively no usage;
- 3=widespread usage

Term: The oracle problem

Source: SWEBOK section 05.01.02.07

SE Fundamental	RSE Equivalent
• The oracle problem	• Checking with the science lead • Checking prior papers • Gold result

Fundamental description	RSE practice
An oracle can be any human or mechanical agent that decides whether the system under test behaved correctly. If the oracle cannot decide, the test output is classified as inconclusive.	RSEs are not generally aware of the the SER terminology, but do practice the oracle concept. Often through an appeal to expert knowledge or prior results.

SER potential (source) and opportunities	RSE awareness (source) and usage (sources)
3 (expert judgement)	1 (expert judgement)
RSEs *should* know the term 'the oracle problem' and should be thinking about it [illegible] as they work. The oracle problem will be increasingly important/challenging in AI-related settings, for example. Strategies for dealing with inconclusive outcomes are also of increasing importance.	3 (expert judgement)

Term Mapping Schema

```
# array of synonyms or adjacent concepts
se_fundamental:

# Short (1 sentence) description of the SE fundamental
fundamental_description:

# array of section identifiers
swebok_section:

# array of synonyms or adjacent concepts
rse_concept:

# Text, a brief description of the typical realizations of the fundamental, in RSE practice
rse_practice:

# General level of awareness of the fundamental in the research software community
# integers 0–3, 0=effectively no awareness, 3=widespread awareness
rse_awareness:
```

Potential for SER to improve usage in RSE

- 0=effectively no opportunity;
- 3=significant SE research beneficial

Term: The oracle problem

Source: SWEBOK section 05.01.02.07

SE Fundamental	RSE Equivalent
• The oracle problem	• Checking with the science lead • Checking prior papers • Gold result

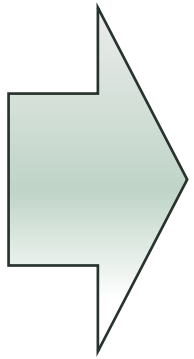
Fundamental description	RSE practice
An oracle can be any human or mechanical agent that decides whether the system under test behaved correctly. If the oracle cannot decide, the test output is classified as inconclusive.	RSEs are not generally aware of the the SER terminology, but do practice the oracle concept. Often through an appeal to expert knowledge or prior results.
SER potential (source) and opportunities	**RSE awareness (source) and usage (sources)**
3 (expert judgement)	1 (expert judgement)
RSEs *should* know the term 'the oracle problem' and should be thinking about it more specifically as they work. The oracle problem will be increasingly important/challenging in AI-related settings, for example. Strategies for dealing with inconclusive outcomes are also of increasing importance.	3 (expert judgement)

Initial Experience

- Pilot crowdsourcing sessions with mixed RSEs and SERs
- Participant-driven term selection and open discussions
- Insights on terminology granularity and SWEBOK structure
- Identification of shared, differing, and overlapping concepts

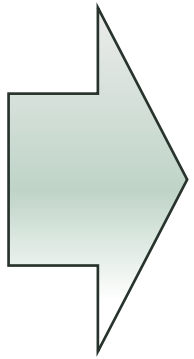
Envisioned Outcome

SE fundamentals unfamiliar to the RSE community



opportunities for education, training, and outreach to promote adoption by the RSE community;

SE fundamentals perceived as less useful by the RSE community



opportunities for adaptation and refinement by the SER community to make SE fundamentals more suitable for RS contexts.

Motivation and Goals

Software Engineering Research (SER)

Systematic investigation of SE methods, tools, processes, practices

SE Fundamentals

Research Software Engineering (RSE)

Informal understanding of SE fundamentals through practical experience of creating research software

RSE Practices

Emerging Questions from Initial Discussions

Scope & Boundaries

- What is the scope, where to draw boundaries?

Fundamentals & Core Concepts

- What counts as fundamental?

Abstraction & Granularity

Call for Contribution and Collaboration

A

- Fund
  Concepts?
- Scope & Boundaries?
- Abstraction & Granularity?
- Consensus & Alignment?

on SWEBOK as reference

- ➢ Crowdsourced collection across RSE and SER
- ➢ Web-based platform for community contributions
- ➢ Interactive discussion forum for qualitative insights

unfamiliar to the RSE community

ation,
ch to
promote adoption by the RSE community;

SE fundamentals perceived as less useful by the RSE community

opportunities for adaptation and refinement by the SER community to make SE fundamentals more suitable for RS contexts.