Replication and Ethics in Experimentation

Eduardo Figueiredo

http://www.dcc.ufmg.br/~figueiredo

Experiment Replications

- Replication is repeating the experiment under similar conditions
  - Varying subjects, for instance
- Replication helps to increase confidence
  - If randomization is correct, replications with samples of a population are expected to show the same results

Classification

- Close replications
  - Follow as close as possible the original procedures
- Differentiated replications
  - Study the same research questions, but using different procedures
  - They may also vary one or more conditions

Close Replications

- They increase confidence on results
  - But, they may require the same researchers to replicate
- Close replications are actually rare in software engineering
  - Many factors may vary in the complex setting of a SE experiment

Differentiated Replications

- They are used for more exploratory studies
  - More knowledge may be gained
- Example of factors to change
  - Site: where the experiment is conducted
  - Instrumentation: forms and data collection
  - Subjects: samples from different population
  - Variable and measures, etc.

Aggregating Evidence

- Several empirical studies may together give answers to questions
  - Collection and synthesis of evidence must meet scientific standards
- Systematic literature reviews are means to collect and synthesize evidence
  - They target a specific research question
  - If question is more general, a mapping study can be used

Ethics in Experimentation

Ethics

- Every empirical strategy involving humans must consider ethical aspects
  - Some are regulated by laws
- Key principles
  - Subjects must give informed consent
  - Researchers must maintain confidentiality
  - The study should have value over risks

Informed Consent

- Subjects are participating voluntarily
  - They should have enough information to decide whether participate or not
- Consent form typically include
  - Research project title
  - Consent and comprehension (by subject)
  - Confidentiality promised by experimenters
  - Clarification, if the subject asks for
  - Signatures (both subject and experimenter)

Confidentiality

- Subjects must be sure that all information remains confidential
- Aspects of confidentiality
  - Data privacy: restricted data access to, for instance, private information
  - Data anonymity: data must not identify a subject
  - Anonymity of participation: decision of participation should be kept secret

Inducement (Benefits)

- Subjects may receive benefits to motivate their participation
  - They are going to learn something new
  - Some monetary inducement may also be involved

- Experimenters should be fair to treat all participants

Example of Ethical Issues

- Mining code repositories retrieving personal data (like social networks)
  - Location
  - Programming habits
  - Comments in commits
- Source code written by humans may reveal information about them

Bibliography

- C. Wohlin et al. **Experimentation in Software Engineering**, Springer. 2012.
  - Chapter 2 - Empirical Strategies (Sections 2.6, 2.8, and 2.11)