DATA PRIVACY IN SOFTWARE ENGINEERING

PhD Working Group 7

Swapneel Sheth, Csaba Nagy, Jonathan Bell,
Éva Ádámkó, Qi Dawei

Mentor: Mark Grechanik
INTERVIEW PROCESS

- Online Survey
- Primarily, qualitative and personal interviews
- 12 participants

Interviewee Affiliation:
- Student: 42%
- Faculty: 33%
- Industry: 17%
- Not indicated: 8%
INITIAL THOUGHTS
INITIAL THOUGHTS

• “Privacy is an important topic, but what does it have to do with Software Engineering?”
INITIAL THOUGHTS

• "Privacy is an important topic, but what does it have to do with Software Engineering?"

• "What does privacy have to do with requirements engineering?"
INITIAL THOUGHTS

• “Privacy is an important topic, but what does it have to do with Software Engineering?”

• “What does privacy have to do with requirements engineering?”

• “Privacy is probably more important to the databases community, but since they don’t have the experience with applications that SE has, we need to work together”
INITIAL THOUGHTS

• “Privacy is an important topic, but what does it have to do with Software Engineering?”

• “What does privacy have to do with requirements engineering?”

• “Privacy is probably more important to the databases community, but since they don’t have the experience with applications that SE has, we need to work together”

• “Privacy is important [...] but software (and software engineering) will follow the money”
INITIAL THOUGHTS

• “Privacy is an important topic, but what does it have to do with Software Engineering?”

• “What does privacy have to do with requirements engineering?”

• “Privacy is probably more important to the databases community, but since they don’t have the experience with applications that SE has, we need to work together”

• “Privacy is important [...] but software (and software engineering) will follow the money”

• “Privacy is important for Software Engineering, not just the databases community - programs need to work with data, so SE should be involved”
PRIVACY IN SE: RE

Hi, build us a system!
PRIVACY IN SE: RE

“Hi, build us a system!”
PRIVACY IN SE: RE

“Hi, build us a system!”    “OK!”
PRIVACY IN SE: RE

“Hi, build us a system!”  “OK!”
PRIVACY IN SE: RE

“Hi, build us a system!” “OK!”
PRIVACY IN SE: RE

• Should contractor “blue” reuse code between clients?

• Can purple reverse engineer orange’s private requirements from the reused code?
• How do we **mine code** when parts of it have been **redacted**?

• How do you **localize faults** when you **do not have access** to all source? [due to privacy or licensing concerns]

• How do we **preserve private information** in stack traces?

• Is **anonymization** enough?

• Could we **mine** PC reviews with MSR and NLP techniques to **identify** “anonymous” reviewers?
FINAL THOUGHTS

• “This problem is **not as straight forward** as I initially thought - it seems there are many open problems”

• “Current state-of-the-art seems to be **lacking** in expertise”

• “I hadn’t thought of these problems before... it seems that the **state-of-the-art may not work**”

• “There seem to be **many open and interesting problems** for SE and privacy”

• “Since 9/11 there have been **many grants** and **lots of money** in this area (in the US)”
Global Software Development

MSR

PRIVACY
Source Code Analysis

Requirements Engineering
Fault Localization

Software Testing

PWG Group 7:
Swapneel Sheth, Csaba Nagy, Jonathan Bell, Éva Ádámkó, Qi Dawei

Mentor: Mark Grechanik