Notice of the Final Oral Examination
for the Degree of Master of Science

of

KARAN TONGAY

BE (Savitribai Phule Pune University, 2017)

“Privacy Preserving Software Engineering for Data Driven Development”

Department of Computer Science

Friday, December 4, 2020
9:30 A.M.
Remote Defence

Supervisory Committee:
Dr. Neil Ernst, Department of Computer Science, University of Victoria (Supervisor)
Dr. Sean Chester, Department of Computer Science, UVic (Member)

External Examiner:
Dr. Issa Traore, Department of Electrical and Computer Engineering, UVic

Chair of Oral Examination:
Dr. Pooja Parmar, Faculty of Law, UVic

Dr. Stephen Evans, Acting Dean, Faculty of Graduate Studies
Abstract

The exponential rise in the generation of data has introduced many new areas of research including data science, data engineering, machine learning, artificial intelligence to name a few. It has become important for any industry or organization to precisely understand and analyze the data in order to extract value out of the data. The value of the data can only be realized when it is put into practice in the real world and the most common approach to do this in the technology industry is through software engineering. This brings into picture the area of privacy oriented software engineering and thus there is a rise of data protection regulation acts such as GDPR (General Data Protection Regulation), PDPA (Personal Data Protection Act), etc. Many organizations, governments and companies who have accumulated huge amounts of data over time may conveniently use the data for increasing business value but at the same time the privacy aspects associated with the sensitivity of data especially in terms of personal information of the people can easily be circumvented while designing a software engineering model for these types of applications. Even before the software engineering phase for any data processing application, often times there can be one or many data sharing agreements or privacy policies in place. Every organization may have their own way of maintaining data privacy practices for data driven development. There is a need to generalize or categorize their approaches into tactics which could be referred by other practitioners who are trying to integrate data privacy practices into their development. This qualitative study provides an understanding of various approaches and tactics that are being practised within the industry for privacy preserving data science in software engineering, and discusses a tool for data usage monitoring to identify unethical data access. Finally, we studied strategies for secure data publishing and conducted experiments using sample data to demonstrate how these techniques can be helpful for securing private data before publishing.