PhD Thesis outline

[**General Planning 2**](#_htjirofzjpc)

[**Detailed progress 2**](#_ntwxawqkg9vc)

[**Chapter 1: Introduction 4**](#_etsray288stm)

[**Chapter 2: A systematic review of privacy-preserving distributed data mining 5**](#_84eexv4rrgod)

[**Chapter 3: A Privacy-Preserving Infrastructure for Analyzing Personal Health Data in a Vertically Partitioned Scenario 6**](#_a9btz3yro2e7)

[**Chapter 4: Studying the Association of Diabetes and Healthcare Cost on Distributed Data from The Maastricht Study and Statistics Netherlands using a Privacy-Preserving Federated Learning Infrastructure 7**](#_lb1b10jhgvby)

[**Chapter 5: Generalized Linear Models on Vertically Partitioned Data using Distributed Block Coordinate Descent 8**](#_wncdgjwvon7b)

[**Chapter 6: Improving correlation capture in generating imbalanced data using differentially private conditional GANs 9**](#_r7vrh4tsjo4c)

[**Chapter 7: Citizen-centric data management for using personal data responsibly in research 10**](#_wxmfxmef86mo)

[**Chapter 8: Discussion and Conclusion 11**](#_1byevsq11cca)

## 

## General Planning

| Items | Start | End | Duration |
| --- | --- | --- | --- |
| All main chapters ready | 03-Sep-2020 | 20-Feb-2021 | 6 months |
| Intro+discussion+connection | 01-Feb | 01-Mar | 1 month |
| Supervisor review+revision | 01-Mar | 20-Apr | 1.5 month |
| Ready for reading committee | 20-Apr | 01-May | 0.5 month |
| Send to reading committee | 01-May | 15-Jun | 1 .5 month |
| Thesis approval | 16-Jun | 01-Sep | 2.5 month |
| Defense | 01-Sep | 01-Oct | 1 month |
| **Target Time** | **31-December-2021** | | |
| Detailed progress | | | |
|
| **Chapters** | **Task** | **Deadline** | **Complete** |
| Chatper 1 - Introduction | Draft | 1-Feb | TO DO |
| Revison | 15-Feb | TO DO |
| Chatper 2 - title | Revision | 14-Sep | Completed |
| Feedback | 30-Sep | Completed |
| Submission | 25-Mar | Completed |
| Chapter 3 - title | Published | - | Completed |
| Chapter 4 - title | Get analysis result | 18-Jun | **Completed** |
| Result&Dicussion | 10-Jul | Completed |
| Draft | 4-Aug | Completed |
| Revison (1) | 25-Aug | In progress |
| Revison (2) | 16-Sep | In progress |
| Finalize and submit | 29-Sep | In progress |
| Chatper 5 - title | Revision | 15-Sep | Completed |
| Submission | 30-Sep | Completed |
| Chapter 6 - title | Experiments | 6-Nov | Completed |
| First draft | 30-Nov | Completed |
| Revision (1) | 3-May | Completed |
| Revision (2) | 30-Jun | Completed |
| Finalize | 15-Aug | Completed |
| Submission | 1-Sep | Completed |
| Chapter 7 - title | Exploration+literature | 30-Jun | Completed |
| Experiments | 1-Sep | Completed |
| First draft | 1-Sep | Completed |
| Revision (1) | 1-Oct | In progress |
| Revision (2) | 15-Oct | In progress |
| Finalize | 1-Nov | In progress |
| Submission | 7-Nov | In progress |
| Chapter 8 - Discussion & conclusion | Draft | 15-Feb | TO DO |
| Revison | 1-Mar | TO DO |

## 

## Chapter 1: Introduction

* Current picture (Statement)
* Problems (we are facing now: centralized data, privacy issues)
* Challenges (sharing data takes risks)
* Motivations and the need for a distributed privacy-preserving data mining platform
* Significance (how important at this moment for who?)
* Contributions (at the beginning and the end)
* Structure of this thesis

## 

## Chapter 2: A systematic review of privacy-preserving distributed data mining

**Status: Published**

**Paper title:** A systematic review of privacy-preserving distributed data mining

**Target Publisher:** [Data Science](https://content.iospress.com/journals/data-science)

**Submission:** 1st July 2021

**What is the research problem?**

Which issues each of the existing algorithms are tackling and which challenges remain yet unsolved in the literature?

**sub-questions to answer:**

1. What user scenario does each approach assume?

2. What privacy/ which level of privacy does each approach preserve?

3. What evidence does each approach provide to prove the privacy can be preserved?

4. What data mining problem does each approach solve?

5. What performance (accuracy, efficiency, scalability) does each approach achieve?

**What are the results (what will we learn)?**The result is a comprehensive overview of existing studies including their algorithms, privacy-preserving methods, evaluations, and other 7 metrics. Moreover, this survey should also provide a guideline for future research in this domain.

## 

## Chapter 3: A Privacy-Preserving Infrastructure for Analyzing Personal Health Data in a Vertically Partitioned Scenario

**Status: Published**

**Paper title:** XXX

**Conference:**  XXX

**Paper link:** XXX

**What is the research problem?**

**What is the approach?**

**How is the approach novel (what are the contributions)?**

**What are the results (what will we learn)?**

## 

## Chapter 4: Studying the Association of Diabetes and Healthcare Cost on Distributed Data from The Maastricht Study and Statistics Netherlands using a Privacy-Preserving Federated Learning Infrastructure

**Status: Revision**

**Paper title:** XXX

**Conference:**  XXX

**Paper link:** XXX

**What is the research problem?**

**What is the approach?**

**How is the approach novel (what are the contributions)?**

**What are the results (what will we learn)?**

## 

## Chapter 5: Generalized Linear Models on Vertically Partitioned Data using Distributed Block Coordinate Descent

**Status: Submitted** (second author)

**Paper title:** XXX

**Conference:**  XXX

**Paper link:** XXX

**What is the research problem?**

**What is the approach?**

**How is the approach novel (what are the contributions)?**

**What are the results (what will we learn)?**

## 

## Chapter 6: Improving correlation capture in generating imbalanced data using differentially private conditional GANs

**Status: Submitted**

**Paper title:** XXX

**Conference:**  XXX

**Paper link:** XXX

**What is the research problem?**

**What is the approach?**

**How is the approach novel (what are the contributions)?**

**What are the results (what will we learn)?**

## 

## Chapter 7: Citizen-centric data management for using personal data responsibly in Research

**Status: Revision**

**Paper title:** XXX

**Conference:**  XXX

**Paper link:** XXX

**What is the research problem?**

**What is the approach?**

**How is the approach novel (what are the contributions)?**

**What are the results (what will we learn)?**

## 

## 

## Chapter 8: Discussion and Conclusion

* Review of chapters
* Challenges in applying theories to practice
  + Data linkage in vertically partitioned data
  + A trusted third party in reality
  + The optimal privacy-preserving technology
  + Privacy measurement and proof
  + Explainability and transparency
* Generation and use of synthetic data
  + Privacy preservation in synthetic data
* Indispensability of ethical-legal support
* Privacy-preserving methods and (re)consenting
* Citizen control over their data
* Future perspectives