Developing a Data Management System for Obstetric Research Into Long Term Effect of In Vitro Fertilisation Techniques

Allard J. van Altena
University of Amsterdam
Email: a.j.vanaltena@amc.uva.nl

18th May 2015

Contents

Chapter 1

Introduction

2 pages

The domain and background What is the clinical background of the problem. Outcome of fertility treatment is unknown. Solved by linking PRN to IVF (introduce IVF-PRN project)?

What is big data? ***Allard:** *(Where do I include a small piece of base information about big data?)* Why is big data of importance for this problem? Introduce that data is being collected and should be shared/used in research. Move towards the point that there is a problem with data processes.

The data problem Talk about the research type in which data is being used (registration), say how other types of research (trials, RCT) might also benefit in the end. So what is the actual (data) problem we have?

Using IT as leverage Propose that we use IT to overcome the problem. How should the system help with the problem (link back to big data problems)?

Research questions

- How do we implement a user-friendly system in a IVF medical domain which covers problems concerning: data security, data access, data browsing, and data querying?
  - What are the legal and security aspects of this system?
  - What is the data model for this system?
  - What are the functions of this system and which parts of the research process should this system support?
  - Who are the users and what are the use-cases for these users?
  - What functions were actually implemented in the prototype?
  - To what extent does this system meet the expectations of users?
- What needs to be changed in the current attitude towards data usage to promote big data in a IVF medical domain?
  - What are the blocking aspects of data usage?

– What are the promoting aspects of data usage?
– What alignment needs to take place to promote data usage?
– How can IT be leveraged to achieve this goal?

Chapter 2

Requirement Analysis

3 pages

2.1 Process Analysis

Doing research and IVF-PRN project data How is research currently performed? What is data management in the context of the IVF-PRN system?

The system functionality concept Initial idea (before brainstorm), where does the system fit into the research workflow.

2.2 Brainstorm

The execution and results Idea after brainstorm, where does the system fit into the research workflow now?

The similarities and differences Differences between before and after brainstorm.

Analysis Functions transposed in research workflow, what points are supported/improved.

8 pages

2.3 Security

2.3.1 Literature Review

Why is security so important it has its own section.

Medical big data ethics and security Might move big data to another section (introduction?).

The 'how?' and 'why?' of security Why is security needed, background information. Integrate literature on security and introduce some concepts of security that were described.

The legal side of security Integrate Dutch law into the security, there are some requirements that need to be met for a system to be legal.

2.3.2 Interview

Set-up Description of the interviews taken and what information was gathered there.

2.3.3 Technical & Procedural Cornerstones of Security

Table of items that were distilled from literature and interviews.

Analysis Analysis, how do the security items apply to the system.

2 pages

2.4 Data Provenance

What is data provenance.

The 'why?' of provenance What is provenance good for?

The 'how?' of provenance How can provenance be applied?

Analysis What does provenance provide for the IVF-PRN system?

Chapter 3

System Design & Implementation

1 - 2 pages

3.1 Function Design

Visualised Show figures created by applying the brainstorm functions to the research workflow and describe in text what they are saying.

4 pages

3.2 Re-use

***Allard:** *Rosemary subsection first and start comparison against other systems from there, or order like it is now?*

3.2.1 Review

In-house development What is it now = neuroscience data management and processing system. Describe what Rosemary is and use that as a starting point to compare the other systems to (home-brew is our starting point in this).

External system evaluations Describe what systems were encountered and why they do not meet our needs and/or why Rosemary/home-brew was a better choice.

3.2.2 Rosemary

Goodness of fit Functions of Rosemary. Compare to brainstorm functions.

Why Rosemary is the better pick Arguments for use of this system: re-use, in-house project, expertise is here, data model fits with minimal change (next chapter).

Data model Describe Rosemary data model (datums, tagging, messaging, etc.), mention flexibility of the model.

3 pages

3.3 Implementation

3.3.1 Processes

Division of management tasks Describe user, data, request, and publication-management, which were implemented and which were left out for now. What functions for each of the tasks was implemented.

User roles What user roles were implemented and which (parts) of the management tasks are they allowed to perform.

3.3.2 Rosemary

Testing the data model's flexibility How was the Rosemary data model changed to fit to our functional needs. Show figure with differences.

Back-end and front-end Explain how the front and back-end were changed to meet each of the functions.

Chapter 4

System Evaluation

5 pages

4.1 Case Evaluation

4.1.1 Set-up

Describe set-up of case evaluation.

4.1.2 Cases

Researcher Evaluation with researcher user.

Committee Evaluation with committee user.

Administrator Evaluation with administrator user.

4.2 Analysis

Summary, positives and negatives. Where does the system fully support the users? What are the hard parts for the users?

System hits Positive points, successes, opportunities.

System misses Missed opportunities, missing functions.

Chapter 5

Discussion & Conclusion

4 pages

- What are the biggest blocking factors in doing research with sensitive data and how can these be overcome?
- What is the importance of supporting administrative tasks (provides security, better understanding, etc.)?
- How can this specific system be used in other settings/knowledge domains?

5.1 Overview and Implications

Give short overview of all the chapters (outcomes/results). What are the outcome of the requirement analysis, design, software implementation, cases, and evaluation steps?

5.2 Appraisal

Strengths and limitations One registry as example (no branching to other domains). Limited time to program, thus less implemented functionality. However, this example provides many different angles to approach the problem.

Comparison with existing systems Comparison with other existing software (OpenClinica, etc)

Future research and development considerations What functions need to be implemented to give the system an extra 'boost'. How can the system manage the expectations of users better. How can the system better support users in their data management tasks.

5.3 Conclusion

Answer the questions stated in the introduction.

Chapter 6

Data Management Position

10 pages

6.1 Introduction

Problem Describe data 'usage' phases and what the problems are now:

- Study preparation, data usage approval
- Data retrieval
- Data massaging/preparation
- Data analysis
- Data reuse

Position There should be more data re-use and cooperation between data producers/consumers. IT can help with the alignment of these two factions.

6.2 Background and Evidence

How is research performed now? Describe Dutch law on privacy and data usage. Explain how METC and commissions influence data usage. Talk about clinicians. Patient willingness to share their data for research. Big data and its general stance on data usage.

6.3 Discussion

(Limitations):

- How does Dutch law prescribe security and ethics to be taken into consideration and why does this matter, what is the fear?
- Is law taken into account in the right manner every time?
- How do METC, commissions, and other influential bodies limit the use of data?
- What are the clinician's opinion, thoughts, and 'holding back' in fear of rebuttal

- What are expected hold-ups for research now (data gathering can take extremely long)?
- Will patients withhold their data for use in research and why?

(Chances):

- What does existing big data research describe is necessary in order to conduct successful and meaningful studies?
- What are some necessities for a certain fluidity in using sensitive data for research?
- How to can limitations be aligned with the big data opportunities
- What needs to be changed in attitude in order to open up data reuse and sharing (introduce Scandinavian ideas)?
- How can online and offline (IT) be leveraged to bring convincing and easy-to-grasp ideas to inspire people to change their way of working?

Give both sides of the issue and discuss how these two can be aligned with each other to meet at a sweet spot for both factions.

6.4 Conclusion

Course of action Figure out if IT can be leveraged to align medical and technical personnel.

Possible solutions IT leverage in the form of a system (like the one in this thesis).

References go here

Appendix A

Abbreviations

Appendix B

Security Checklists