UNIVERSITEIT TWENTE.

LOG DATA ANALYSIS
SUMMERSCHOOL “BIG DATA IN CLINICAL MEDICINE”
2 JULY 2014 – ENSCHEDE
SASKIA AKKERSDIJK & FLOOR SIEVERINK
A QUICK INTRODUCTION

Saskia Akkersdijk, MSc.
PhD candidate

Floor Sieverink, MSc.
PhD candidate

Supervisor: Lisette van Gemert-Pijnen, PhD
Department of Psychology, Health & Technology
UNIVERSITEIT TWENTE.
OUTLINE

1. Introduction to logging
2. Practical examples
3. How to make eHealth more persuasive
4. Methods and techniques for analysis
PERSUASIVE TECHNOLOGY

• Technology that is designed to change attitudes or behaviors of the users
• But not through coercion!
PROBLEMS OF EHEALTH

• Not all users…
  • … complete an intervention
  • … use all available components

• There is a dose-response relationship, but…

… there is little understanding how people use an eHealth intervention!
THE BLACK BOX PHENOMENON
• Therefore: monitoring the use of applications

• Not only the quantity of use (e.g. # log-ins, page-views, time spent on an application):
  ▪ Does not necessarily lead to better results
  ▪ Can be an indicator for non-strategical and unfocused use

=> Inefficient systems?

LOG DATA (2)

We look at:

- How do users navigate over an application?
- Which components are used mostly?
- At which points do users drop-out?
- When is the application used?
- Etc. etc. etc.

Understanding the use => Improving the (persuasiveness of) the application
LOG DATA

A history of actions that have taken place
VARIOUS APPEARANCES OF LOGDATA
SHORT OVERVIEW

- Google Analytics
- Database
- CSV/Excel
VARIOUS APPEARANCES OF LOGDATA (2)

GOOGLE ANALYTICS

UNIVERSITEIT TWENTE.
VARIOUS APPEARANCES OF LOGDATA (3)

DATABASES

12 databases
247 tables

UNIVERSITEIT TWENTE.
# VARIOUS APPEARANCES OF LOGDATA (4)

CSV/EXCEL FILES

<table>
<thead>
<tr>
<th>SessionID</th>
<th>User</th>
<th>Role</th>
<th>Disease</th>
<th>DateTimeTag</th>
<th>Device</th>
<th>Code</th>
<th>ExtraInformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>304626</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 13:00</td>
<td>PC</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>304626</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 13:08</td>
<td>PC</td>
<td>91</td>
<td>Apotheek: Cholesterol</td>
</tr>
<tr>
<td>2</td>
<td>414791</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 13:09</td>
<td>PC</td>
<td>91</td>
<td>Gezond leven in tien minuten</td>
</tr>
<tr>
<td>3</td>
<td>404954</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 13:12</td>
<td>PC</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>404954</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 13:14</td>
<td>PC</td>
<td>91</td>
<td>Apotheek: Hoge bloeddruk</td>
</tr>
<tr>
<td>3</td>
<td>404954</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 13:14</td>
<td>PC</td>
<td>91</td>
<td>Gezond leven in tien minuten</td>
</tr>
<tr>
<td>4</td>
<td>335253</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 13:19</td>
<td>PC</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>335253</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 13:20</td>
<td>PC</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>335253</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 13:21</td>
<td>PC</td>
<td>91</td>
<td>Fabels en feiten</td>
</tr>
<tr>
<td>5</td>
<td>404954</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 14:09</td>
<td>PC</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>404954</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 14:09</td>
<td>PC</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>404954</td>
<td>Patient</td>
<td>Diabetes</td>
<td>23-7-2013 14:09</td>
<td>PC</td>
<td>90</td>
<td></td>
</tr>
</tbody>
</table>
VARIOUS APPEARANCES OF LOGDATA (5)
THINGS TO CONSIDER

• What do you want to know?
• What do you want to do with your data?

⇒ How do you collect your data?
  • Google Analytics?
  • Server side?
  • Client side?

⇒ Is the application new or existing?
  Consequences for amount of work and missing data
PRACTICAL EXAMPLES

- My Health Platform (MHP)
- Living to the Full
- e-Vita
MY HEALTH PLATFORM

Mijn zorgdossier
Via Mijn zorgdossier houdt u uw persoonlijke gegevens bij. U kunt uw gegevens bekijken, wijzigen, aanvullen en verwijderen.

Mijn gezondheid
Hier kunt u uw aandoeningen en allergieen noteren.

Mijn leefstijl
Hier kunt u de kenmerken van uw leefstijl bijhouden. Denk hierbij aan roken, bewegen en ontspannen.

Mijn notities
Hier kunt u uw aantekeningen noteren.

Mijn medicijnen en vaccinaties
Hier kunt u bijhouden welke medicijnen u gebruikt en hebt gebruikt. Ook kunt u uw vaccinaties en overige middelen noteren.

Mijn gegevens
Hier kunt u uw persoonlijke gegevens noteren, zoals naam, adres en e-mailadres.

Mijn zorgdoelen
Hier kunt u uw aangemaakte adviezen inzien. Ook kunt u uw behandelplannen, inhalatieplannen en leefstijlplannen inzien en wijzigen.

Mijn meetwaarden
Hier kunt u uw meetwaarden bijhouden, zoals uw gewicht. U kunt hier ook zelf meetwaarden toevoegen.
MHP
GOOGLE ANALYTICS BASED

Sessions per month

Session distribution over the week
Session distribution over the day
MHP
GOOGLE ANALYTICS BASED

MHP use
- Measurements
- Reference books
- Help
- Advice
- Nutrition coach
- Utility program tools
- Utility program tips
- Assignments Fitness Coach
Google Analytics can be very useful if:

- You want to know about *all* users
- You are mostly interested in numbers (quantity of use)
- You want a quick overview of the use

Things to consider:

- Results of Google Analytics are not always precise
- Individuals cannot be identified (use as a group)
LIVING TO THE FULL
LIVING TO THE FULL
ADHERENCE
LIVING TO THE FULL
NOTABLE USAGE PATTERNS OF INDIVIDUAL USERS

- Login and logout within one minute
- Feedback often not viewed when available

Adherers:

- ↑ logins at the start
- ↑ sessions to complete lessons (later, ↓ sessions to complete lesson, learning effect?)
- Female, ↑ need for cognition,

Mindfulness exercises are an important part of the intervention.

However, these exercises are often not viewed/downloaded as intended.

How do the effects of the intervention correlate to the use of the intervention?
With log data:

• Insight in adherence, no. actions, use patterns; adherers
• Prompts for persuasion in treatment program: willingness, awareness, adoption of new behaviors/skills

Next steps:

• Smart ways to handle log data are needed
• Real-time => personalised adaptive interfaces
• Intervene for people ‘at risk’
• Persuasive triggers => profiles, timing, dose, duration,
Attrition starts when users “get lost” in the intervention (Kelders et al., 2013; Nijland et al., 2011)

Thus, the first impression of the application is important.

Therefore:

Log data analysis of the first use


E-VITA
RESULTS

1. Users often follow the global menu structure (n=9)

2. 93% ended their session when they visited the education service as the first step after the log in.

3. Most users visited either 1 or 5 or more services

4. The distribution of the routes was diffuse.

E-VITA
CONCLUSIONS

• Log file analysis can provide valuable prompts for improving the system design of eHealth applications

• Needed:
  • Automated ways to handle the data
  • Validation of the results, e.g. via interviews, usability tests, questionnaires, …, …
How can we use this information to make eHealth applications more persuasive?
INTERPRETATION OF THE RESULTS

<table>
<thead>
<tr>
<th>Name</th>
<th>Date/Time</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>January 12, 13:14</td>
<td>Log in</td>
</tr>
<tr>
<td>Mary</td>
<td>January 12, 13:20</td>
<td>Log in</td>
</tr>
<tr>
<td>John</td>
<td>January 12, 13:22</td>
<td>Finish questionnaire</td>
</tr>
<tr>
<td>Mary</td>
<td>January 12, 13:50</td>
<td>Finish questionnaire</td>
</tr>
<tr>
<td>Mary</td>
<td>January 12, 13:41</td>
<td>Send email</td>
</tr>
<tr>
<td>George</td>
<td>January 21, 16:51</td>
<td>Log in</td>
</tr>
<tr>
<td>George</td>
<td>January 21, 16:30</td>
<td>Finish questionnaire</td>
</tr>
<tr>
<td>George</td>
<td>January 21, 16:31</td>
<td>Complete action</td>
</tr>
<tr>
<td>Mary</td>
<td>January 30, 21:20</td>
<td>Log in</td>
</tr>
<tr>
<td>George</td>
<td>January 22, 21:39</td>
<td>Finish questionnaire</td>
</tr>
<tr>
<td>Mary</td>
<td>February 02, 12:30</td>
<td>Log in</td>
</tr>
<tr>
<td>Mary</td>
<td>February 02, 12:35</td>
<td>Complete action</td>
</tr>
<tr>
<td>Mary</td>
<td>February 03, 12:40</td>
<td>Send email</td>
</tr>
<tr>
<td>George</td>
<td>February 04, 22:00</td>
<td>Log in</td>
</tr>
<tr>
<td>George</td>
<td>February 04, 22:12</td>
<td>Finish questionnaire</td>
</tr>
<tr>
<td>George</td>
<td>February 04, 22:13</td>
<td>Complete action</td>
</tr>
<tr>
<td>Mary</td>
<td>February 05, 16:46</td>
<td>Log in</td>
</tr>
<tr>
<td>Mary</td>
<td>February 05, 17:20</td>
<td>Finish questionnaire</td>
</tr>
<tr>
<td>Mary</td>
<td>February 05, 17:24</td>
<td>Send email</td>
</tr>
</tbody>
</table>
How do users ‘move’ over an application?

=> The use of what (combinations of) services correlates with adherence and effects?

=> How can we make this routes visible/logical?

=> Do users find all services?

When do users drop out?

=> Why? How can we prevent this?
LOG DATA HANDS-ON FOR PERSUASIVE TECHNOLOGY (2)

• (How) do users react to persuasive triggers?
  => How can we use these triggers in such a way that users are persuaded to return to the application on the long term?

• What is the influence of the practical context and the content on the use of the application?
  => How can we adjust interventions in such a way that they are an added value in a certain context?
  => How can we make the content visible?

METHODS AND TECHNIQUES FOR ANALYSIS
FOR FUTURE RESEARCH

- ‘Simple’ statistics, charts, counts, etc.
- Followed pathways
- Markov Modeling
- Market-Basket analysis
- Machine learning
  - Supervised learning
  - Unsupervised learning
- Visualization of log data
MARKOV MODELING

Models a process where the state depends on previous states in a non-deterministic way.
MARKOV MODELING (2)
MARKET BASKET ANALYSIS

- Frequent itemset mining

- Business decision-making processes
  - Catalog design, cross-marketing, customer shopping behavior analysis.
MARKET BASKET ANALYSIS (2)

How can we improve the lay-out to make navigating over the system more logical and convenient?

Of transactions that included milk:
- 71% included bread
- 43% included eggs
- 29% included toilet paper
“Field of study that gives computers the ability to learn without being explicitly programmed.”
- Arthur Samuel (1959)

- Learn from data
- Examples:
  - Spam detection
  - Chess program
  - Speech recognition
- Two types:
  - supervised (teach a computer how to do something)
  - unsupervised (let the computer learn it by itself)
UNSUPERVISED LEARNING

- Clustering
  - Find structure in data
  - Given the behaviour of users of a system, divide them into different groups. Then interpret the groups to find out how these groups use the system.
UNSUPERVISED LEARNING (2)

- Clustering
  - Find structure in data

- Anomaly detection
  - Find outliers
  - Automatically find measurements of patients that are not 'normal'.
UNSUPERVISED LEARNING (3)

- Clustering
  - Find structure in data

- Anomaly detection
  - Find outliers

- Recommender system
  - Suggestions based on data
  - Suggest reading material for a patient based on what (s)he read before, or perhaps based on measurements.

UNIVERSITEIT TWENTE.
SUPERVISED LEARNING

- Give a computer labeled data (telling it what is what) and let it learn how to predict a label for new data

- Different methods, for example:
  - Linear regression
  - Neural networks
  - Support Vector Machines

- Example: Given the groups found in the unsupervised clustering approach, define different types of users and teach the computer to predict what kind of user a new user is based on his/her behaviour on the application. This enables the application to personalize its approach to that type of user.
VISUALISATION OF LOG DATA
QUESTIONS?

Contact: s.m.akkersdijk@utwente.nl / f.sieverink@utwente.nl