Datenschutzaspekte des Projektes

(Stand: September 2013)

Michael Predeschly, Prof Dr. Markus Schäffter

Inhaltsverzeichnis:

Vorwort:

Im Nachfolgenden werden aufbauend auf dem „Merkblatt zum Datenschutz bei medizinischen Studien mit Patientendaten“ des Bayrischen Datenschutzbeauftragen [1] die Aspekte der oben genanten Studie beleuchtet. Hierbei wird stets auf die Datenvermeidung und Datensparsamkeit nach §3a des Datenschutzgesetzes geachtet.

1. Inhalt und Zweck der Studie

Soziale Netze und in besonderem der Marktführer Facebook werden tagtäglich von vielen Millionen Menschen genutzt. In Deutschland sind derzeit ca. 25 Millionen Menschen aktive Nutzer des Netzes. Mit mFG (**m**obiles **F**acebook **G**estationsdiabetes) soll untersucht werden, inwieweit Soziale Netze bei der Bewältigung von Krankheit hilfreich eingesetzt werden können. Die Form der Unterstützung soll dabei in Form von Online-Interventionen erfolgen, welche schon in vielen Bereichen erfolgreich eingesetzt wurden [8, 10,12,13]. Die Nutzung von Facebook ist liegt nahe, da hierbei sowohl eine Einarbeitung in ein neues System entfällt [15], als auch bereits ein reger Austausch über Krankheiten dort stattfindet [14].

Alle Fragestellungen dieser Studie sollen am Modell des Gestationsdiabetes untersucht werden. Dazu sollen Patientinnen mit Gestationsdiabetes die Möglichkeit bekommen, Vitalparameter, wie z.B. Gewicht, Blutzucker, und Gewohnheiten, wie körperliche Aktivität, Zigarettenkonsum, transparent zu machen und Informationsangebote zur Bewältigung der Krankheit innerhalb von Facebook mit ausgewählten "Facebook-Freunden" zu teilen und zu kommentieren. Gestationsdiabetes wurde deswegen gewählt, da die Krankheit sehr unvermittelt im Leben der Patientin erscheint und meist eine Umstellung des Verhaltens und der Lebensgestaltung nach sich zieht. Ein weiterer Punkt ist, dass die Nutzung von Facebook und Smartphones in der Zielgruppe eine starke Verbreitung hat. Smartphones sind deswegen wichtig, da so eine Echtzeit-Intervention erfolgen kann [12,13], da Benachrichtigungen direkt auf dem Smartphone anzeigt werden.

Dabei ist nicht nur interessant, wie die Patientin in Facebook unterstützt werden kann [7], sondern auch, inwieweit die Freunde diese Unterstützung gewähren und in welcher Form diese erfolgt. Sind es aufbauende Kommentare oder Angebote für ein Offline-Treffen? Werden eventuelle weitere Materialien zu der Krankheit gesammelt und der Patientin zugänglich gemacht? Auch ist interessant, wie schnell eine Reaktion erfolgt, wenn die Patientin etwas in der Gruppe, die sie unterstützen soll, schreibt. Ebenfalls soll untersucht werden, wie stark sich die Freunde an einem solchen Experiment beteiligen und sich dazu über die Krankheit informieren. Hierzu werden regelmäßig Informationen bereitgestellt und den Patientinnen und ihren Freunden zum Selbststudium vorgelegt [9]. Die Freunde werden hierzu regelmäßig auf im Kontext des Krankheitsverlaufes interessante Inhalte hingewiesen. Dies dient auch zur Verbesserung des Verständnisses der Krankheit [11].

Zur Validierung dieser Annahmen wird eine Anwendung mobiles Facebook Gestationsdiabetes (mFG) entwickelt, die zum einen Daten aus Facebook exportiert und nach Facebook exportiert, aber auch selbst Daten über das Nutzungsverhalten erfasst. In der Anwendung selbst kann die Patientin Daten über sich und die Krankheit eintragen, die dann graphisch aufbereitet den Freunden zur Verfügung gestellt werden, um einen

erleichterten Einstieg in die Beurteilung des Zustandes der Patientin zu erlauben.

2. **Betroffener Personenkreis**

Es sollen im Rahmen einer ersten Machbarkeitsstudie, 10 Patientinnen gewonnen werden, die nach der Diagnose des Schwangerschaftsdiabetes, bis zur Geburt ihres Kindes, ihre Daten in mFG eintragen.

Jede dieser Patientinnen soll darüber hinaus die Möglichkeit haben, eine beliebige Anzahl von Freunden in eine Facebook-Gruppe einladen, in der dann während der Studie diskutiert werden kann. Hierbei entscheidet die Patientin selbst, welche Facebook-Freunde in die Gruppe eingeladen werden und welche Personen dann über den dort bekanntgegebenen Zugang Zugriff auf die personenbezogenen Daten in mFG erhalten.

3. **Beteiligte, speichernde Stelle**

Es gibt zwei Stellen, an denen Daten hinterlegt werden:

1. Die Hochschule Ulm, auf deren Server das mFG-System betrieben wird. Die personenbezogenen Daten der Patientin werden ausschließlich auf dem Server der Hochschule Ulm, gespeichert.
2. Facebook beinhaltet alle weiteren Daten, die in der Diskussion der Freunde mit der Patientin in Facebook in so genannte „Geheime-Gruppen“ eingestellt werden. Dies bedeutet man kann nur auf Einladung Mitglied der Gruppe werden.
   Diese Daten werden nach Beendigung der Studie exportiert und pseudonymisiert gespeichert. Mit Hilfe der Daten soll vor allem die Art und die Häufigkeit der Interaktion der Freunde mit der Patientin nachvollzogen werden.
   Nach diesem Export wird die Gruppe in Facebook aufgelöst (ehemalige Gruppenmitglieder haben keinen Zugriff mehr auf die Daten) und vom Administrator gelöscht.
   Nach Angaben von Facebook sollen die Daten nach 90 Tagen (bezogen auf Profil) auch aus allen Backups bei Facebook gelöscht sein.

Die Daten werden dabei ausschließlich im Rahmen der Promotion des Autors und gegebenenfalls später darauf aufbauender Forschungsarbeiten genutzt.

4. **Erhobene Daten sowie deren Erforderlichkeit**

Einmalig zu Beginn der Studie werden Daten der Patientin bzw. des Kindes erfasst, um Voreinstellungen in mFG vornehmen zu können. Diese Daten werden nach Beendigung der Studie wieder gelöscht.

Daten die in mFG erfasst werden:

- Größe in cm
- initiales Gewicht in kg
- Raucherin oder Nichtraucherin
- Schwangerschaftswoche
- voraussichtliches Geburtsdatum

Angaben zur Größe und des Gewichts sind erforderlich, um den BMI der Patientin zu bestimmen, der für einen Vergleich mit den Normwerten aus Patientenrichtlinien benötigt wird, sowie für Normangaben bzgl. des Blutzuckers [16]. Außerdem soll den Freunden der Patientin im Laufe der Schwangerschaft anhand eines Diagrammes der relative-Gewichtsverlauf angezeigt werden (vgl. Abbildung 1).

Durch die Angabe, ob die Patientin eine Raucherin ist, wird eine weitere Pflegemaske in mFG freigeschaltet, in der die Anzahl der täglich gerauchten Zigaretten erfasst werden. Hier soll, ähnlich wie bei zu großen Gewichtsschwankungen, der „soziale Druck“ untersucht werden, der durch die Freunde erzeugt wird.

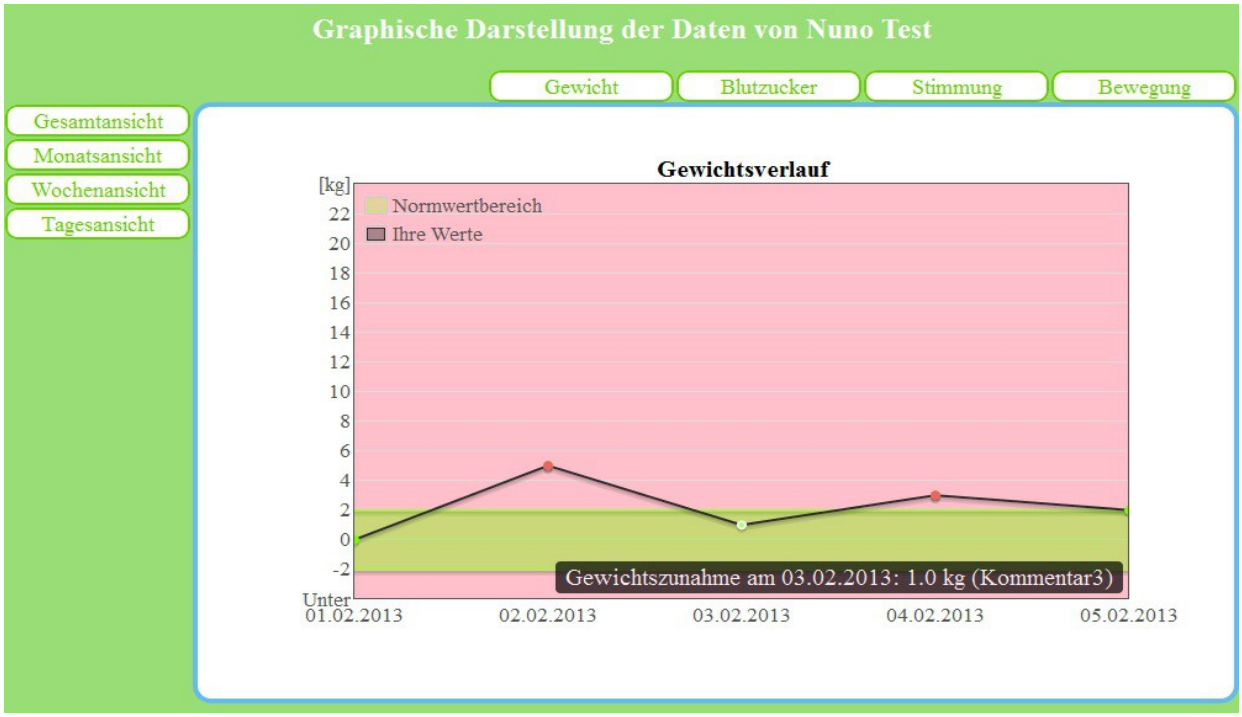


Abbildung 1: Relativer Gewichtsverlauf graphische Darstellung in mFG

Das voraussichtliche Geburtsdatum des Kindes bzw. die Schwangerschaftswoche helfen bei der Einordnung des Gewichtes der Schwangeren und ermöglicht es, zielgerichtete Informationen, abhängig vom Fortschritt der Schwangerschaft, an die Patientin und ihren

Freundeskreis herauszugeben.

Nach Beginn der Studie soll die Patientin regelmäßig Daten erfassen. Die hierfür ausgewählten Daten entstammen der Patientenrichtlinie der Deutschen Diabetes Gesellschaft (DDG) [16]. Bei allen Eintragungen kann die Patientin einen optionalen Kommentar zu dem jeweiligen Wert hinterlassen, um beispielsweise zu begründen, warum das Gewicht besonders stark angestiegen ist:

- Gewicht:
  Das Gewicht soll einmal pro Woche erfasst werden. Hierbei wird den Freunden allerdings immer nur die relative Gewichtszunahme angezeigt und nicht die absoluten Gewichtszahlen. Dies erscheint sinnvoll, da es in der Schwangerschaft nur um die Gewichtszunahme und nicht um das Ausgangsgewicht der Schwangeren geht. Optional wird es möglich sein, dass einzelne Gewichtswerte oder auch das ganze Gewicht ausgeblendet wird, sollte die Patientin dies wünschen.

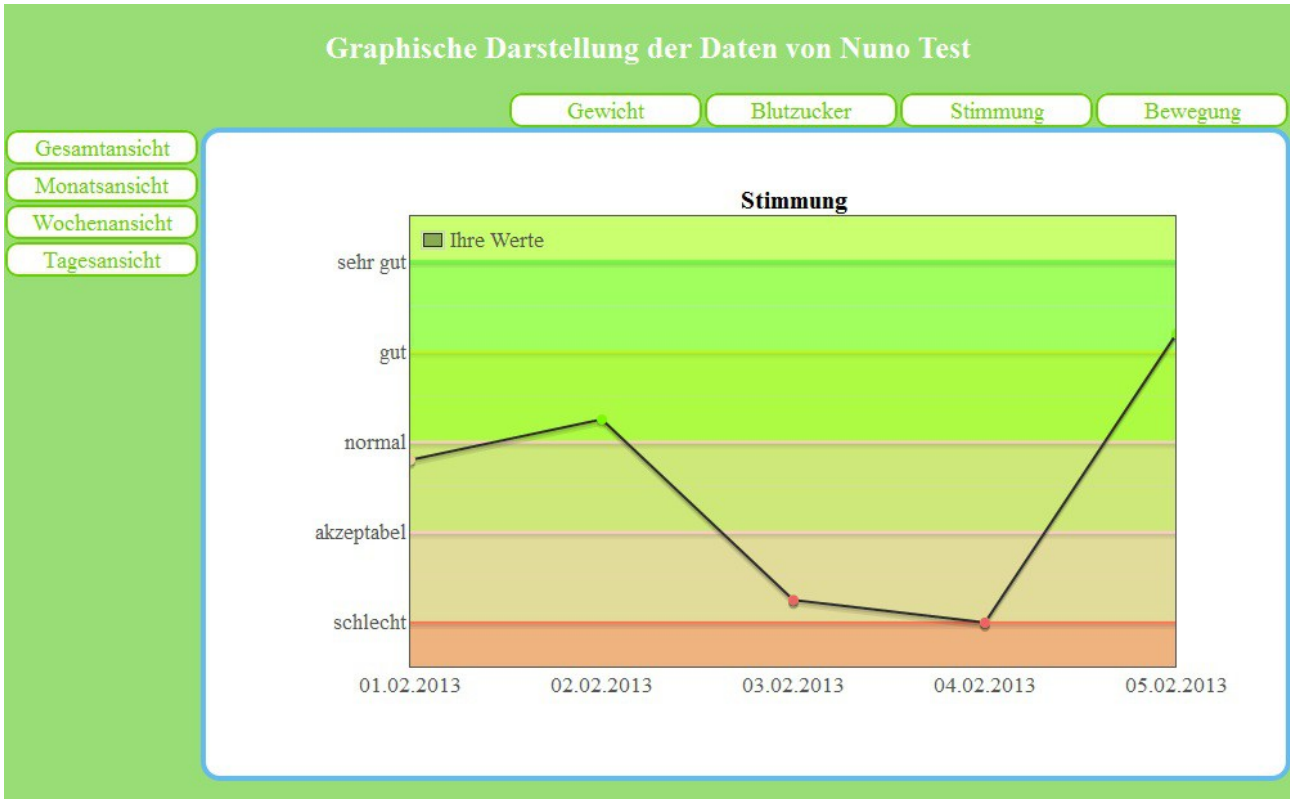


Abbildung 2: Stimmungsverlauf der Patientin mit fünf möglichen Werten

- Blutzucker:
  Hier werden pro Tag maximal sechs Werte erfasst, je einmal vor und nach jeder Mahlzeit. Die Auswertung erfolgt hierbei in zwei graphischen Darstellungen. Nüchternwerte und postprandiale Werte werden getrennt voneinander betrachtet.

- Stimmung:
  Die Stimmung sollte ebenfalls jeden Tag erfasst werden. Hierbei wird ein subjektiver Wert zwischen sehr schlecht und sehr gut von der Patientin vergeben (vgl. Abbildung 2).
- Bewegung:
  Bei der Bewegung kann die Patientin selbst eigene Arten der körperliche Aktivität angeben. Wichtig ist für die Auswertung dabei die gesamte körperliche Aktivität pro Tag. Diese Daten werden ebenfalls graphisch aufbereitet, den Freunden präsentiert.
- Nikotin: Sollte die Patientin angeben haben, dass sie raucht, soll die Anzahl der täglich gerauchten Zigaretten protokolliert werden.

Daten, die in der Facebook-Gruppe erfasst werden:

Neben den Daten, die in mFG gespeichert werden, gibt es noch die in der Facebook-Gruppe eingetragenen Daten. Der am Ende der Studie erfolgende Export umfasst dabei alle dort eingestellten Interaktionen.

Der Facebooknutzer kann zusätzlich bei Beiträgen verschiedene Ergänzungen angeben, die einen Beitrag um Zusatzinformationen erweitern z.B. die Angabe des Ortes der per GPS oder IP bestimmt wird, oder Personen, die bei einem Beitrag erwähnt wurden. Diese Daten werden nicht mit exportiert.

Die restlichen Daten, die exportiert werden, sind im Folgenden beschrieben:

- Facebook-ID:
  Eindeutige Nummer, über die der Facebook-Nutzer eindeutig identifiziert wird.
- Typ des Beitrages:
  Hier werden verschiedene Kategorien unterschieden: Post (Beitrag, welcher kommentiert werden kann), Kommentar, Like (Zustimmung signalisieren durch drücken eines „Daumen nach oben“-Symbols). Dies ist für die Auswertung interessant, da so gemessen werden kann, welche Art von Interaktion wie stark genutzt wurde.
- Datum/Zeit des Beitrages:
  Der Zeitpunkt, an dem ein Beitrag in der Gruppe eingestellt wurde, wird über einen so genannten Timestamp festgehalten, der wie folgt aussieht: z.B. 2012-11-01T15:46:28+0000 für den 1. November 2012 um 15:46.

  Eine genaue Zeiterfassung ermöglicht es, die Zeit zu messen, die zwischen den

einzelnen Interaktionen liegt. Gerade diese Reaktionszeit sollte durch die Wahl des Mediums Facebook in Kombination mit der Benachrichtigung an ein Smartphone deutlich besser sein, als in anderen, vergleichbaren Systemen.

- Inhalt des Beitrages:
  Bei einem Post oder einem Kommentar wird der Text gespeichert. Dieser wird bei der späteren Auswertung unterschiedlichen Klassen zugeordnet werden, um die Art der Kommunikation zu beurteilen u.a. Angebote zur Unterstützung der Patientin oder Ermahnungen.
- Rolle innerhalb der Gruppe:
  Hier wird unterschieden zwischen Administrator und nicht Administrator. Für die Gruppe lässt sich so allein anhand dieses Merkmales festhalten, ob die Patientin, die neben dem Ersteller der Gruppe (mFG-Administrator) der einzige Administrator ist, etwas geschrieben hat, oder ein Freund.

Wichtig bei allen aus Facebook exportierten Daten ist lediglich die Art und der Zeitpunkt an dem kommuniziert wurde. Daher werden aus diesem Grund nach dem Export der Daten die Nachrichten selbst anonymisiert (Pseudonymisierungskonzept siehe Abschnitt 9.).

Daten, die mittels User-Tracking in mFG erfasst werden:

Innerhalb von mFG werden die Nutzer des Systems und ihr Surfverhalten überwacht. Dies bedeutet es wird genau protokoliert, welche Seiten aufgerufen wurden. Im einzelnen wird dabei erfasst zu welcher Uhrzeit, Seiten aufgerufen wurden, die Verweildauer auf der jeweiligen Seite und die Häufigkeit der Aufrufe bestimmter Inhalte.

Daten, die über Fragebögen erfasst werden:

Die maschinell erfassten Daten werden am Ende der Studie durch Fragebögen und Interviews ergänzt. Diese dienen zum einen dazu, das System zu verbessern und Informationen über die Nutzungsweise zu erhalten. Zum anderen sollen so auch weiterführende Informationen gewonnen werden, die Effekte betreffen, welche außerhalb von Facebook und mfg aber im Zusammenhang mit der Nutzung des Systems erfolgt sind, z.B. Anrufe oder Besuche eines Freundes.

Die Fragebögen werden am Ende der Studie von der Patientin und in einer abgewandelten Form von den Freunden der Patientin ausgefüllt. Die Daten werden danach analog zu den bisher erfassten Daten in anonymisierter bzw. pseudonymisierter Form ausgewertet. Hierbei werden wiederum jeder Gruppe von Patentin und deren Freunden ein Buchstabe und den Freunden ihre bereits vergebene Nummer zugeordnet.

Wichtig ist für die wissenschaftliche Auswertung alleine die Zuordnung zur jeweiligen Patientin, da zwischen den einzelnen Gruppen Unterschiede zu erwarten sind. Interessant dürfte es hierbei auch sein, zu sehen, inwieweit bestimmte Verhaltensweisen oder Vorkenntnisse das Verhalten während des Studienverlaufes beeinflussen.

5. **Analyseergebnisse der Daten**

Während des Studienverlaufes werden die patientenbezogenen Daten für die Patientin selbst und deren Freunde in graphischer Form aufbereitet. Hierbei werden verschiedene Diagramme erstellt, die über mFG eingesehen werden können. Dabei wird den Freunden der Patienten kein Zugriff auf die Daten selbst gewährt. Es ist jedoch möglich, sich einzelne Werte der den Diagrammen zugrundeliegenden Daten anzeigen zu lassen.

Nach Abschluss der Studie wird das Nutzerverhalten der Teilnehmer analysiert. Hierbei wird betrachtet, welche Interaktionen erfolgten und wie die zeitlichen Zusammenhänge zwischen diesen Interaktionen waren. z.B. wie lange dauerte es bis zur ersten Reaktion auf einen Beitrag? Zu welchen Uhrzeiten erfolgten die meisten Reaktionen? Hierzu werden zum einen die aus Facebook exportierten Daten, aber auch die in mFG über das Nutzerverhalten gewonnen Daten aus dem User-Tracking herangezogen.

Die personalisierten in mFG gespeicherten Daten, wie z.B. Gewicht und Blutzucker werden nicht für die Auswertung genutzt. Es wird lediglich festgehalten, ob der Wert innerhalb oder außerhalb der zulässigen Grenzen gelegen hat. Außerdem wird festgehalten, zu welchem Zeitpunkt eine Eintragung erfolgte (siehe auch Beispiel in Abschnitt 9). Die Auswertung des Erfassungszeitpunktes ermöglicht es, im Zusammenspiel mit den Zeitpunkten der Reaktion der Freunde, Rückschlüsse auf die Schnelligkeit der Interaktion zu ziehen.

6. **Einwilligungserklärung, Patienteninformation**

Die Einwilligung der Patientin erfolgt nach §4a des BDSG. In einer Datenschutzerklärung, die vor Beginn der Studie abgegeben und mittels Unterschrift bestätigt werden muss, wird die Patientin darüber informiert, dass personenbezogene Daten von ihr erfasst werden und aus welchem Grund.

Die Datenschutzerklärung umfasst folgende Punkte:

- Speicherungsort der Daten (Hochschule Ulm)
- Aufklärung, dass Daten nur pseudonymisiert gespeichert werden
- Art und Dauer der Speicherung von 10 Jahren
- Weitergabe der Daten gegebenenfalls in pseudonymisierter/ anonymisierter Form auch an dritte.
- Recht eines Widerrufs ihrer Einwilligung zur Teilnahme zur Studie und die Folgen

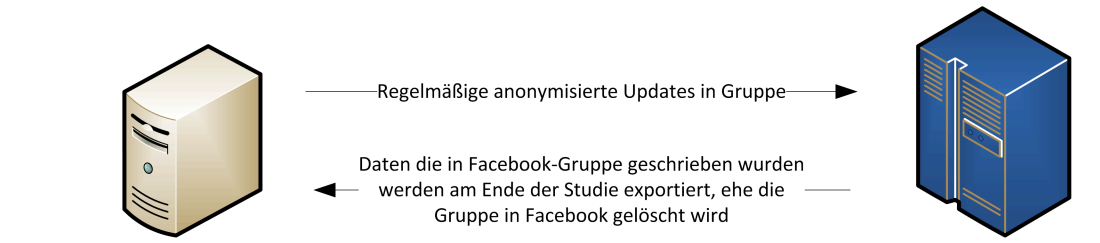
eines Rückzugs dieser Einwilligung.

-

Eine weitere Patienteninformation erfolgt, vor Beginn der Studie,in Form eines Gespräches, mit Hinweis auf die erfassten Daten. Nach diesem Gespräch erhält die Patientin die Datenschutzerklärung samt einem Merkblatt, welches die Nutzung und die im System erfassten Daten detailliert und exemplarisch vorstellt.

7. Datenflüsse und Speicherorte

Abbildung 3 zeigt die Aufteilung der Daten und deren Speicherungsort. Wichtig ist dabei zu beachten, dass alle kritischen personenbezogenen Daten der Patientin nicht bei Facebook, sondern auf dem Server der Hochschule Ulm gespeichert werden. Darüber hinaus werden alle Daten, die von mFG in Facebook eingestellt werden, vor der Übermittlung anonymisiert. Dies bedeutet, es wird nur eine Benachrichtigung verschickt, welche den abstrakten Zustand der Patientin anzeigt (z.B. in Form einer Ampel) und der Aufforderung sich die Visualisierung der Daten in mFG anzusehen.



Server der Hochschule Ulm

Daten:

- ID
- Gewicht
- Größe
- Raucher oder Nichtraucher
- Schwangerschaftswoche
- Blutzucker
- Stimmung
- Bewegung
- Nikotin
- Zeitpunkt des Aufrufes einer Seite
- Anzahl der Aufrufe einer Seite

Facebook-Server

Daten:

- FB-ID des Erstellers eines Beitrags
- Name des Erstellers eines Beitrags
- Typ des Beitrags
- Datum des Beitrags
- Inhalt des Beitrages (falls kein „Gefällt-mir“)
- Rolle in der Gruppe (Admin/kein Admin)
- (Facebook kann zusätzlich noch Ort und IP protokollieren)

Abbildung 3 Unterschiedliche Speicherorte der Daten

Die Daten, die ansonsten in Facebook von der Patientin und ihren Freunden eingestellt werden, werden nach Ende der Studie exportiert und die Gruppe anschließend gelöscht. Die exportierten Daten werden dabei in anonymisierter Form für die Auswertung gespeichert. Das Pseudonymisierungskonzept wird in Abschnitt 9. genauer beschrieben.

8. Archivierungskonzept und Löschfristen

Die Rohdaten, wie auch die anonymisierten/pseudonymisierten Daten werden nach Beendigung der Studie in einem passwortgeschützten Archiv auf CDs gebrannt. Die Daten auf den Archiv-CDs sind durch ein Passwort geschützt. Das Passwort des Archives ist in zwei Teile unterteilt, deren erster Teil nur dem Betreuer und deren zweiter Teil dem Autor bekannt ist.

Die Daten-CDs werden dabei in einem verschlossen Umschlag mit dem Datum der Vernichtung der Daten (10 Jahre nach Studienende), aufbewahrt. Die Daten dürfen nur in anonymisierter Form weitergegeben werden. Die Originaldaten sind nur zur Überprüfung der Ergebnisse durch die Leiter der Studie oder eine andere Forschungsgruppe zu nutzen. Es darf dabei kein anderer/neuer Sachverhalt untersucht werden.

Bis auf die am Ende der Studie erstellten Archiv-CDs werden alle Daten auf den Servern der Hochschule Ulm gelöscht. Die Facebook-Gruppen der jeweiligen Patientinnen werden ebenfalls aufgelöst und gelöscht, nachdem die Daten exportiert wurden.

9. Pseudonymisierungskonzept

Für die Studie ist der Zusammenhang zwischen Patientin und deren jeweiligen Freunden wichtig. Aus diesem Grund muss dieser Zusammenhang jederzeit erkenntlich sein. Um eine Anonymität der Patientin und der jeweiligen Freunde dennoch zu gewährleisten, wird die Patientin mittels eines Buchstabens substituiert (vgl Orientierungshilfe zur Pseudonymisierung in der medizinischen Forschung [6]).

Dabei ist der Buchstabe willkürlich festgelegt. Die Zuordnung von Buchstabe zu Patientin wird mittels eines Zufallsgenerators vorgenommen und im Falle einer Kollision so lange durchgeführt bis jeder Patientin ein eindeutiger Buchstabe zugeordnet ist. Die so erzeugte Zuordnung von Patientin zu Pseudonym bleibt so lange erhalten, bis die Daten der Patientin komplett anonymisiert wurden, was mit Abgabe des Fragebogens erfolgt ist. Anschließend wird die Zuordnung gelöscht. Eine Depseudonymisierung ist nicht vorgesehen.

a) Freunde der Patientin

Zur Verschleierung der Identität der Freunde der Studienteilnehmer werden diese durchnummeriert. Dabei muss der Zusammenhang der Freunde zur jeweiligen Patientin erhalten bleiben, um Rückschlüsse auf den jeweiligen Verlauf der Krankheit und die Reaktionen der zugehörigen Freunde auswerten zu können. Um diesen Zusammenhang wiederherzustellen, wird jedem Gruppenmitglied ein Buchstabe vorangestellt, welcher

derjenigen Patientin zugeordnet ist, z.B. für Freunde von Patientin Y wird einer Nummer ein Y vorangestellt.

Bsp. Freund von Patientin Y mit der Nummer 3

→ Y[3] ist das Pseudonym des Freundes

Sollte eine Person mit mehreren der Patientinnen befreundet sein, erhält sie für jede Patientin eine eigene ID und wird so in zwei Gruppen als Mitglied geführt. Hierdurch wird gewährleistet, dass die Person nicht durch seine Doppelzugehörigkeit identifiziert werden kann.

Bei den textuellen Beiträgen in der Facebookgruppe, den Kommentaren in Facebook oder mFG erfolgt eine Klassifizierung der Beiträge - der konkrete Inhalt wird nicht gespeichert, lediglich, ob es sich z.B. um einen ermutigenden Beitrag oder ein Hilfsangebot handelte. Darüber hinaus wird bei jedem Beitrag Uhrzeit, Datum und das Pseudonym des Erstellers gespeichert (vgl Tabelle 1).

Bsp:

Facebook-Text: „Willst du heute spazieren gehen? → Angebot

Facebook-Text: „Finde ich nicht gut, wenn du so wenig Daten einträgst“ → Ermahnung

Die genauen Kategorien werden dabei nach Sichtung der eingestellten Facebook-Texte erstellt werden. Es sollen nicht mehr als 10 Kategorien existieren. Beispiele: Ermahnung, Ermunterung, Treffen …

Ein anonymisierter und pseudonymisierter Beispieldatensatz aus der Facebookgruppe könnte dann bei einem Kommentar bzw. einem Gefällt-mir-Klick wie folgt aussehen:

Freund	Typ	Timestamp	Inhalt
Y[42]	POST	2013-06-23T21:36:06+000	Ermahnung
Y[3]	LIKE	none	none
Y[14]	KOMMENTAR	2013-06-23T21:45:12+000	Einladung zu Aktivität

Tabelle 1: Anonymisierter und pseudonymisierter Beispieldatensatz aus der Facebookgruppe

b) Patientin

Die Patientin selbst erhält keine Nummer. Sie wird einfach mit dem Buchstaben Y substituiert. Hierdurch ist sie innerhalb der Datensätze eindeutig, aber ihre Identität bleibt verborgen.

Die regelmäßig erfassten personenbezogenen Daten der Patientin aus mFG werden ebenfalls mit dem ihr zugeordneten Buchstaben verbunden und gespeichert. Dabei wird der Name der Patientin durch den ihr zugeordneten Buchstaben ersetzt. Daten wie die Größe oder der erwartete Geburtstermin des Kindes werden gelöscht, da sie für die weitere Auswertung nicht notwendig sind und zur Identifizierung der Patientin beitragen könnten.

Ein Beispieldatensatz der Patientin, der nach den oben ausgeführten Maßnahmen entsteht, könnte für einen Tag wie folgt aussehen:

Gewicht: Angabe erfolgt in absoluten Zahlen. Es werden aber nur relative Angaben gespeichert.

Patientin	Datum	Gewichtsveränderung in kg
Y	23.06.13	+1,3

Tabelle 2: Pseudonymisierter Beispieldatensatz zum Gewicht

Blutzucker: (Zv1 steht dabei für **Z**eit **v**or Essen **1**, Wv1 für **W**ert **v**or Essen **1** bei Zn1 und Wn1 bedeutet das n dementsprechend **n**ach):

Pat.	Dat.	Zv1	Wv1	Zn1	Wn1	Zv2	Wv2	Zn2	Wn2	Zv3	Wv3	Zn3	Wn3
Y	23.06.	07:03	1	07:43	2	12:30	1	13:12	3	17:45	2	18:30	2

Tabelle 3: Pseudonymisierter Beispieldatensatz zum Blutzucker

Stimmung: (Stimmung wird in 5 Stufen angegeben, welche von 1-5 durchnummeriert werden):

Patientin	Datum	Stimmungskategorie
Y	23.06.13	3

Tabelle 4: Pseudonymisierter Beispieldatensatz zur Stimmung der Patientin

Bewegung: (Die Art der körperliche Aktivität wird in Form von Zahlen angegeben, die ersten Bewegungskategorien sind vorgegeben, alle selbstdefinierten bekommen darauf folgende Nummern) :

Patientin	Datum	Dauer in Minuten	Bewegungskategorie
Y	23.06.13	80	2

Tabelle 5: Pseudonymisierter Beispieldatensatz zum Bewegungsverhalten der Patientin

Nikotin:

Patientin	Datum	Anzahl Zigaretten pro Tag
Y	23.06.13	1

Tabelle 6: Pseudonymisierter Beispieldatensatz zum Nikotinkonsum der Patientin

c) Fragebögen

Es wird zwei unterschiedliche Varianten von Fragebögen geben, für die Patientin und für deren Freunde. Hierbei wird nach Abgabe der Bögen wieder lediglich der Buchstabe,der der Patientin zugeordnet wurde, vermerkt. Alle anderen Daten sind nicht personenbezogen. Somit ist lediglich erkennbar, das der Beantworter eines Freundefragebogens zu einer bestimmten Patientin gehört.

Beispiele für Fragen:

- Wie bewerten Sie das System hinsichtlich der Benutzbarkeit?
- Ergaben sich außerhalb des Systems Treffen, Kontakte die durch das System angeregt wurden?
- Haben Sie durch die Interaktion in der Facebook-Gruppe und mit mFG mehr über Gestationsdiabetes gelernt?
- Hat sich durch die Interaktion mit dem System bzw. mit Ihren Freunden über Facebook die Dauer ihrer wöchentlichen körperliche Aktivität erhöht? (nur Patientin)
- Würden Sie das System selbst nutzen? (nur bei Freunden)
- Würden Sie das System nochmals nutzen? (nur bei Patientin)

d) Künstliche Erweiterung der Anzahl der Studienteilnehmer

Da es sich bei der geplanten Studie um eine erste Machbarkeitsstudie handelt, sind nur geringe Teilnehmerzahlen vorgesehen. Aufgrund der geringen Anzahl der Studienteilnehmer von maximal 10 Personen d.h. 10 Facebook-Gruppen, mit einer ungleichen Anzahl von Freunden je Gruppe, müssen diese Zahlen aber erhöht werden, da einzelne Gruppen bzw. Personen sonst identifiziert werden könnten. Zu diesem Zweck werden die Daten der Patienten vervielfacht und mittels eines Rauschens künstlich

erweitert. Des weiteren werden die Patientinnen in Gruppen unterteilt, die nach Alter gegliedert sind. Je nach Zusammensetzung der Patientinnen werden diese Gruppen jeweils 2-5 Jahrgänge beinhalten.

e) Risikoabschätzung

Um abschätzen zu können, wie groß das datenschutzrechtliche Risiko bzgl. einer Verletzung der Persöhnlichkeitsrechte für die Betroffenen ist, sollte zuerst geklärt werden wie hoch der Schutzbedarf der erfassten Daten ist.

Zunächst wird hierfür die in [2] vorgestellte Vorgehensweise genutzt. Hierbei werden Daten klassifiziert und nach ihrer Schutzstufe, sowie der einzelnen Aspekte des Schutzbedarfs, unterschieden nach Verfügbarkeit, Vertraulichkeit und Integrität, bewertet. Hieraus errechnet sich dann der Gesamt-Schutzbedarf (vgl Tabelle 7).

Daten-kategorie	Personen-bezug ja/nein	Schutzstufe (A bis E)	Schutz-bedarf Verfügbar-keit (normal, hoch)	Schutz-bedarf Vertraulich-keit (normal, hoch)	Schutz-bedarf Integrität (normal, hoch)	Gesamtbe wertung Schutz-bedarf
Patientin						
Gewicht	ja	C	normal	hoch	normal	6
Blutzucker	ja	C	normal	hoch	normal	6
Bewegung	ja	C	normal	hoch	normal	6
Stimmung	ja	C	normal	hoch	normal	6
Nikotin	ja	C	normal	hoch	normal	6
Fragebog.	nein	B	normal	normal	normal	2
Freunde						
FB-Gruppe	ja	B	normal	normal	normal	2
Tracking	nein	B	normal	normal	normal	2
Fragebog.	nein	B	normal	normal	normal	2

Tabelle 7: Dokumentation des Schutzbedarfs

Wie man in der Tabelle sieht, sind die Daten, die von der Patientin in mFG erfasst werden, besonders schützenswert. Die Einstufung in Schutzstufe C lässt sich dabei wie folgt

begründen. Alle mit Schutzstufe C versehenen Daten könnten bei Bekanntwerden dem „Ansehen der Patientin schaden“ (vgl. [2]), da z.B. ein hoher Nikotinkonsum in der Schwangerschaft gesellschaftlich nicht für gut befunden wird. Dasselbe gilt für eine dauerhaft schlechte Stimmung oder eine zu geringe, kaum vorhandene körperliche Aktivität. Blutzucker und Gewicht lassen wiederum direkte Schlüsse auf die Lebensweise der Patientin zu und könnten somit auch negative Konsequenzen für das soziale Ansehen haben. Aus diesem Grund sind diese Art der Gesundheitsdaten durch §3(9) BDSG besonders geschützt.

Alle anderen Daten, die in Fragebögen oder durch das User-Tracking erfasst werden, können keiner konkreten Person zugeordnet werden und lassen somit auch keine „besonderen Beeinträchtigungen“ erwarten (vgl. Beschreibung der Schutzstufen in [2]).

Neben der Betrachtung des Schutzbedarfes muss eine Analyse der möglichen Schwachstellen bzw. Angriffszenarien auf des System erfolgen (vgl. auch ISO 27005 [3]). Hierbei soll aufgezeigt werden, wie Angriffe auf das System erfolgen und wie diese verhindert werden können.

Wir fokussieren uns dabei auf die Schwachstellen, die sich durch das System mFG selbst und unserer Herangehensweise, bzw. die Art des Zugangs über das Internet und die Nutzung von Facebook ergeben. Eine Betrachtung der Schwachstellen in der Administration des Rechenzentrums, sowie deren Hardware, wird ausgeklammert, da dies außerhalb der Kontrollmöglichkeiten der Studienorganisation liegt.

Wir betrachten somit technische Schwachstellen, die durch das von uns zur Verfügung gestellte System bzw. die von uns vorgesehene Art der Pseudonymisierung/Anonymisierung Angriffe ermöglichen.

Dadurch ergeben sich drei Kategorien von Angriffspunkten und daraus resultierend verschiedene Angriffszenarien. Die drei unterschiedlichen Kategorien sind die mit dem System interagierenden Menschen, die Technik, die eingesetzt wird und höhere Gewalt.

Diese Klassifikation entspricht der vom BSI definierten Struktur [3,17] wobei wir uns nur auf die ersten beiden beschränken, da gegen die letzte Kategorie keine effizienten Schutzmaßnahmen getroffen werden können. Zusätzlich unterscheiden wir Angriffe auf das mFG-System und die Daten selbst.

a) mFG-System

Der erste Angriffspunkt ist technischer Natur in Form des mFG-System selbst, welches über ein Webfrontend verfügt und somit von jedem Client-System aus, auch weltweit, zugegriffen werden kann. Hierbei sind klassische Angriffe durch die meist genutzten Webangriffe SQL-Injection oder Cross-Site Scripting [5], bei denen alle Daten ausgelesen

werden, zu erwarten. Da der Hochschulserver für mögliche Angreifer ein sehr lohnendes Ziel ist, ist davon auszugehen, dass die Chance auf solche Angriffe hoch ist. Aber auch die Schutzmaßnahmen an der Hochschule sind dementsprechend hoch.

Der zweite große Punkt, in diesem Szenario, an dem Angriffe erfolgen können, sind die Nutzer des Systems selbst. Es existieren hierbei vier verschiedene Benutzer-Gruppen, die mit dem System interagieren und nur jeweils diejenigen Daten im Zugriff haben sollten, die für sie relevant sind:

- Rechenzentrums-Nutzer mit direkten Zugriff auf das Verzeichnis/die Datenbank. Diese können jederzeit alle Daten aus dem System herauslesen. Allerdings sollten gerade diese Nutzer (z.B. die Administratoren des Rechenzentrums) dementsprechend sensibilisiert sein. Aus diesem Grund stufen wir dieses Szenario als relativ unwahrscheinlich, wenn auch sehr gefährlich ein.
- Administrator-Nutzer-mFG, der vollen Zugriff auf die Daten des mFG Systems besitzt. Dieser Nutzer kann den Export der Daten aus der Datenbank vornehmen. Hierfür gibt es nur einen solchen Nutzer, bei dessen Kompromittierung allerdings ebenfalls ein hoher Schaden entsteht, da hier sowohl die Daten aller Patienten und auch die Tracking-Daten der Freunde ermittelt werden können, als auch die über den Fragebogen gewonnen Informationen.
- Die Patientin selbst, welche natürlich ebenfalls Zugriff auf ihre eingetragenen Daten besitzt, hier aber jeweils nur auf die von ihr selbst eingetragenen und nicht auf die anderer Patientinnen zugriff hat. Des weiteren hat sie keinen Zugriff auf das User-Tracking oder die Daten der Fragebögen.
- Freunde der Patientin, welche nur die Auswertungen der Daten der Patientin zu Gesicht bekommen. Somit ist hier der mögliche Schaden am geringsten.

Wir gehen sowohl bei Patientinnen, als auch Freunden der Patientinnen, davon aus, dass die Komprimierung dieser Zugänge eine mittlere Eintrittswahrscheinlichkeit hat, da die Passwörter vom Administrator vorgegeben werden und somit keine zu einfachen Passwörter genutzt werden können.

Neben den Risiken, die durch Angriffe auf das System mFG entstehen, existiert auch eine zweite Kategorie von Angriffen direkt auf die Daten, die im System existieren bzw. nach Abblauf der Studie vorliegen. Diese Art der Gefahren wird im folgenden Abschnitt näher betrachtet.

b) Erfasste und exportierte Daten:

Die zweite Kategorie von Angriffspunkten umfasst die exportierten und pseudonymisierten Daten selbst. Durch den Zusammenhang zwischen Patientin und Freunden, dürfte es möglich sein, bei Depseudonymisierung eines Freundes Rückschlüsse auf die Patientin zu

ziehen und umgekehrt (Vorausgesetzt man weiß von deren Schwangerschaft). Hiermit könnten dann wieder alle Daten zugeordnet werden. Aus diesem Grund ist der erwartete Schaden sehr hoch.

Schutzbedarf	normal	hoch	sehr hoch
Ereignisse	6	5	1,2,3,4,7

Tabelle 8: Schutzbedarfsfeststellung mit Einstufung der Risiken
Legende: *1. SQL-Injection, 2. Cross-Site-Scripting 3. Admin über Shell, 4. Admin von mFG, 5. Login Patientin, 6. Login Freund, 7. Depseudonymisierung*

Die hier vorgestellten Punkte wurden in Form einer Schutzbedarfsfeststellung [17] in Tabelle 8 klassifiziert. Man sieht deutlich, dass sich eine Reihe von möglichen Ereignissen ergibt, bei denen ein sehr hoher Schaden entstehen kann.

Um die in Tabelle 8 dargestellten Risiken bzw. deren Eintrittswahrscheinlichkeit zu minimieren, werden die im folgenden beschriebenen Vorsichtsmassnahmen getroffen. Eine Klassifikation, der nach den Vorsichtsmaßnahmen vorliegenden Eintrittswahrscheinlichkeit, in Form einer Risikomatrix, findet sich in Tabelle 9.

IT-Risikomatrix	Geringer Schaden	Mittlerer Schaden	Hoher Schaden
Geringe Eintrittsw.		1,2,4,5,7	3,7
Mittlere Eintrittsw.	6		
Hohe Eintrittsw.			

Tabelle 10: IT-Risikomatrix nach Verbesserungen am System
Legende: *1. SQL-Injection, 2. Cross-Site-Scripting 3. Admin über Shell, 4. Admin von mFG, 5. Login Patientin, 6. Login Freund, 7. Depseudonymisierung*

Zum Schutz der im System abgelegten Daten wurden folgende Schutzmaßnahmen getroffen:

In mFG werden lediglich prepared Statements genutzt, um Daten aus der Datenbank zu lesen. Hierdurch wird die Eintrittswahrscheinlichkeit eines solchen Angriffs sehr stark verringert [4].

Cross-Site Scripting wird durch den Einsatz von explizitem Erlauben von Werten in Eingabefeldern anstelle vom Verbieten einzelner Eingaben reduziert. Durch diese Maßnahmen reduziert sich also die Eintrittswahrscheinlichkeit des Angriffes [5].

Der Zugang des Administrators mit vollem Zugriff über die Shell kann leider, im Rahmen unserer Studie, nicht weiter abgesichert werden. Aus diesem Grund bleiben Eintrittswahrscheinlichkeit und Schaden gleich hoch. Wobei hier eine Überwachung durch

Logfiles gewährleistet, dass unberechtigtes Auslesen der Daten nicht unerkannt bleibt.

Bei Nutzung des Administratorzugangs von mFG, wird zusätzlich als Vorkehrung der Export der Patientendaten nur einmal am Tag für eine Patientin möglich sein. Dies ist bei unserer Studie keine große Einschränkung, da sich die Probandinnen auf mehrere Wochen verteilen werden. Somit verringert sich der anzunehmende Schaden.

Die Patientin erhält keine Möglichkeit, ihre Daten automatisiert auszulesen. Sollte sie Interesse an ihren Daten haben. muss sie diese über den Administrator erfragen. Somit müsste ein potentieller Angreifer die Daten von Hand erfassen, was die Eintrittswahrscheinlichkeit reduzieren dürfte.

Der Angriff über den Login der Freunde ist bereits unkritisch und bedarf somit keiner weiteren Betrachtung. Generell ist es gewährleistet, dass jeder der unterschiedlichen Nutzergruppen nur auf die für ihn relevanten Daten Zugriff erhält.

Als Absicherung der exportierten Daten werden diese pseudonymisiert gespeichert. Um einer Depseudonymisierung entgegenzuwirken, werden nur sehr wenige Daten erfasst, die Rückschlüsse auf die Identität zulassen. Zusammen mit der Verschleierung durch die künstliche Erweiterung der Probandenanzahl (vgl. Pseudonymisierungskonzept in Abschnitt 9) verringert sich zwar nicht der Schaden, aber die Eintrittswahrscheinlichkeit dürfte sinken.

Zusammen mit der Pseudonymisierung mittels der Zuordnung von Buchstaben über einen Zufallsgenerator erachten wir die Eintrittswahrscheinlichkeit hierfür als eher gering.

10. **Workflow, Beschreibung typischer Abläufe**

In Abbildung 4 ist der typische Abblauf der Studie als Interaktionsdiagramm dargestellt. Wichtig ist hierbei die Berücksichtigung der einzelnen Phasen der Studie. Die Anonymisierung und Pseudonymisierung erfolgt erst nach Abschluss der Studie und dem Export der entsprechenden Daten aus mFG bzw. Facebook.

Die wie unter 5 beschriebene Auswertung der Einzelgruppen soll nach Abschluss aller 10 Testgruppen veröffentlicht werden. Hierbei können die Teilnehmer jederzeit Widerspruch gegen die Veröffentlichung der Daten einlegen. Da die Daten nach der Anonymisierung nicht mehr einem einzelnen Teilnehmer zuzuordnen sind, muss somit die ganze Gruppe aus den Studienergebnissen entfernt werden.

Eine Auskunft der gespeicherten Daten kann nur so lange personenbezogen erfolgen, bis die Anonymisierung erfolgt ist.

Interaktionsdiagramm Stand 12.09.2013

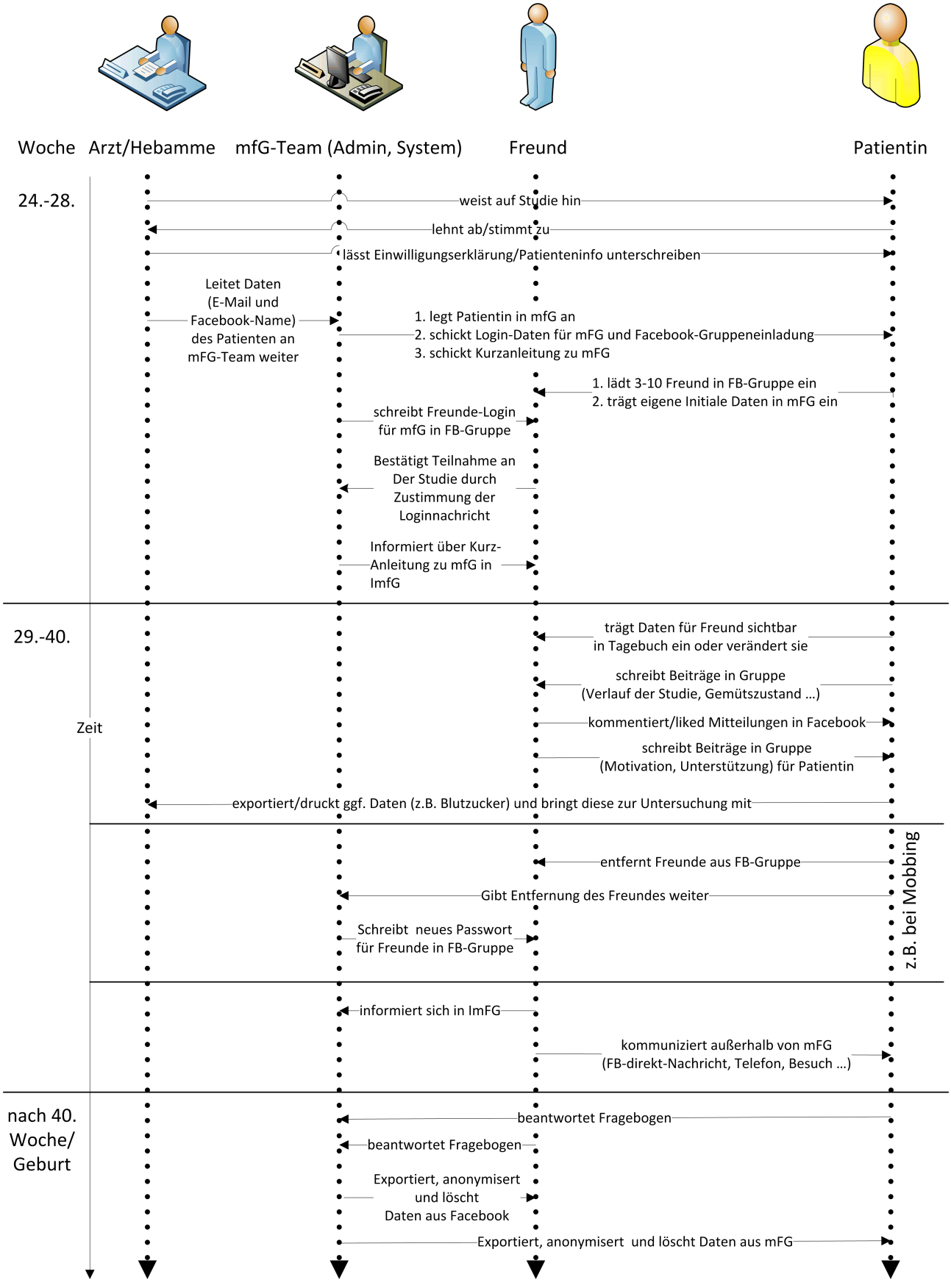


Abbildung 4: Abblauf der Studie als Interaktionsdiagramm

11. Qualitätssicherung, Monitoring

Während der Studie wird durch mFG selbst überwacht, dass die dort hinterlegten Daten in validem Format eingetragen werden. So soll es nicht möglich sein, z.B. ein Gewicht über 250 kg einzutragen oder Buchstaben anstelle von Zahlen für den Blutzucker anzugeben.

Zugriff auf die Daten, während der Studie, neben den Studienteilnehmern, ist nur drei Personen möglich:

- Betreuer
- Leiter der Studie
- Betreuender Arzt der jeweiligen Patientinnen an der Frauenklinik.

12. Technische Ausgestaltung

Auf dem Server der Hochschule läuft ein Apache Server, der zur Erzeugung der Darstellung der Informationen im Browser genutzt wird. MySQL und PHP werden für die Implementierung von mFG genutzt. Für die Visualisierung der eingepflegten Daten in mFG wird Flotr2 [18] als Framework genutzt. Die Datenerfassung kann sowohl über einen gängigen PC als auch über das Smartphone erfolgen, da eine speziell optimierte Smartphone-Ansicht existiert.

Die Verantwortung für die Server Infrastruktur an der Hochschule Ulm liegt beim IMZ der Hochschule Ulm

13. Zusammenfassung

Facebook soll zur Unterstützung von Patienten mit Hilfe des hier vorgestellten mfG-Systems eingesetzt werden. In diesem System sollen von der Patientin diverse Gesundheitsdaten erfasst werden (vgl. Kapitel 4) und anschließend so aufbereitet werden, dass die in einer geheimen Facebook-Gruppe zusammengefassten Freunde motivierend auf die Patientin einwirken können. Da gerade im Zusammenspiel mit dem amerikanischen sozialen Netzwerk aber große Datenschutzbedenken und Probleme bestehen, müssen hierbei diverse Sicherheitsmaßnahmen getroffen werden, um zu gewährleisten, dass keine personenbezogenen Daten in unbefugte Hände fallen.

Die Erhebung personenbezogener Daten wird deswegen auf ein Minimum beschränkt. Die erfassten personenbezogenen Daten werden noch während der Studie mittels eines Pseudonymisierungskonzeptes (vgl. Kapitel 9) in der Datenbank abgelegt. Somit wird gewährleistet, dass die Daten der Patientin gegen Missbrauch geschützt sind.

Die Daten die in der Facebook-Gruppe abgelegt werden, sind nur bis zum Ende der Studie

dort verfügbar. Danach werden diese von allen Mitgliedern getrennt und aufgelöst. Es bleibt lediglich ein anonymisierter Export der Daten, welcher nach einem Archivierungskonzept vorgehalten wird (vgl Kapitel 8).

Wichtig ist dabei auch die klare Trennung der Datenspeicherung in Gesundheitsdaten auf dem Server der Hochschule Ulm und anonymen Mitteilungen, welche in der unsichtbare Facebook-Gruppe der Patientin eingestellt werden (vgl Kapitel 3).

Die Daten selbst werden durch verschiedene Mechanismen geschützt, welche auch das Risiko minimieren sollen, dass Daten missbraucht werden. Dies wurde mit Hilfe einer Risikoanalyse in Kapitel 9 überprüft.

Aufgrund dieser Analyse und der getroffenen Massnahmen können die erfassten und gespeicherten Daten als unkritisch aufgefasst, sowie die Risiken als gering bezeichnet werden.

14. **Literaturliste**

1. Merkblatt zum Datenschut bei medizinischen Studien im Patientenbereich, Version 1.0, bayrischer Datenschutzbeauftragter vom 18.07.2007

   http://www.datenschutz-bayern.de/technik/orient/merkblatt_med_studien.html
2. Prozessorientierte Datenschutzpraxis, 2. Auflage, Eugen Ehmann, Oliver Schonscheck 201, ISBN-13: 978-3824590711
3. ISO/IEC 27005 International Standard, first edition 15.06.2008
4. SQL Injection Attaks and Defense, 1. Auflage, Justin Clarke, 2009, ISBN-13: 978-1597499637
5. A Survey on Web Application Vulnerabilities (SQLIA, XSS) Exploitation and Security Engine for SQL Injection, Johari, R.; Sharma, P., Communication Systems and Network Technologies (CSNT), 2012 International Conference pp.453-458, 11-13 May 2012 ISBN-978-1-4673-1538-8
6. Orientierungshilfen zur Pseudonymisierung in der medizinischen Praxis, bayrischer Datenschutzbeauftragter (letzter Besuch 22.02.2013) http://www.datenschutz-bayern.de/technik/orient/ohilfe_psn_03.html
7. Social Influence as a Driver of Engagement in a Web-Based Health Intervention, Josée Poirier, Nathan K. Cobb, J Med Internet Res 2012;14(1):e36
8. Persuasive Features in Web-Based Alcohol and Smoking Interventions: A Systematic Review of the Literature, Tuomas Lehto, Harri Oinas-Kukkonen, J Med Internet Res 2011;13(3):e46
9. Effects of Internet Use on Health and Depression: A Longitudinal Study, Katie Bessière, Sarah Pressman, Sara Kiesler, Robert Kraut, J Med Internet Res 2010;12(1):e6
10. Benefits of Peer Support in Online Japanese Breast Cancer Communities: Differences Between Lurkers and Posters, Yoko Setoyama, Yoshihiko Yamazaki, Kazuhiro Namayama, J Med Internet Res 2011;13(4):e122
11. Managing the Personal Side of Health: How Patient Expertise Differs from the Expertise of Clinicians, Andrea Hartzler, Wanda Pratt, J Med Internet Res 2011;13(3):e62
12. Harnessing Context Sensing to Develop a Mobile Intervention for Depression, Michelle Nicole Burns, Mark Begale, Jennifer Duffecy, Darren Gergle, Chris J Karr3, MA; Emily Giangrande, David C Mohr, J Med Internet Res 2011;13(3):e55
13. Real-Time Social Support Through a Mobile Virtual Community to Improve Healthy Behavior in Overweight and Sedentary Adults: A Focus Group Analysis, Yoshimi Fukuoka, Emiko Kamitani, Kemberlee Bonnet, Teri Lindgren, J Med Internet Res 2011;13(3):e49
14. Seeking Support on Facebook: A Content Analysis of Breast Cancer Groups, Jacqueline L Bender, Maria-Carolina Jimenez-Marroquin, Alejandro R Jadad, J Med Internet Res 2011;13(1):e16
15. Patient and Parent Viewson a Web 2.0 Diabetes Portal—the Management Tool, the Generator, and the Gatekeeper: Qualitative Study, Sam Nordfeldt, Lena Hanberger, Carina Berterö, J Med Internet Res 2010;12(2):e17

16. Leitlinie für Patientinnen, Schwangere und Interessierte zu Diagnostik, Behandlung u. Nachsorge der Deutschen Diabetes-Gesellschaft (DDG) und der Deutschen Gesellschaft für Gynäkologie und Geburtshilfe (DGGG) H.Kleinwechter, U.Schäfer-Graf, C.Bührer, I.Hoesli, F.Kainer, A.Kautzky-Willer, B.Pawlowski, K.Schunck, T.Somville, M.Sorger, 04/2012
www.deutsche-diabetes-gesellschaft.de/fileadmin/Redakteur/Leitlinien/Patientenleitlinien/GDM_Patienten_LL_END_2012_04_17.pdf
17. Bundesamt für Sicherheit in der Informationstechnik (BSI), Referat 114 - IT-Sicherheitsmanagement und IT-Grundschutz, Webkurs IT-Grundschutz – IT Grundschutz im Selbststudium Ausgabe April 2011
18. Flotr 2 Visualiesierungs-Framework http://humblesoftware.com/flotr2/