

Klaus-Georg Deck
Dr. sc. hum.

Integrierte Datenmanagementsysteme für multizentrische, multidisziplinäre Studien zu genetisch assoziierten Krankheiten

Geboren am 06. 06. 1966 in Landau i. d. Pfalz

Allgemeine Hochschulreife am 19. 06. 1984 in Neustadt a. d. Weinstraße

Studiengang der Fachrichtung Mathematik (NF Philosophie) vom WS 1984/85-SS 1989

Vordiplom am 20. 08. 1986 an der Universität Heidelberg

Diplom am 23. 05. 1989 an der Universität Heidelberg

Promotionsfach: Epidemiologie

Referent: Prof. Dr. rer. nat. Heiko Becher

Die Dissertation behandelt Datenmanagementsysteme von multizentrischen, multidisziplinären Familienstudien und verfolgt die folgenden Ziele:

1. Beschreibung der Anforderungen an ein Datenmanagementsystem für solche Studien.
2. Etablierung eines allgemeinen Modells eines Datenmanagementsystems für diesen Studientyp.
3. Implementierung des Datenmanagementsystems FAMOS für die Studie "Hereditäres Mamma- und Ovarialkarzinom".

Anforderungen

Die Anforderungen an ein Datenmanagementsystem für den betrachteten Studientyp sind aus der Datensicht eine den Studienvariablen adäquate und möglichst redundanzfreie, umfassende Datenmodellierung, die der Multizentrik und der Multidisziplinarität der Studie gerecht wird, die korrekte und vollständige Repräsentierbarkeit von genetischen Stammbäumen sowie die Bereitstellung von parametrisierbaren Routinen zum Datenmonitoring, zur Erstellung deskriptiver Statistiken und zum Datenexport für zentrumsübergreifende Datenauswertungen. Aus Sicht der mit dem System arbeitenden Benutzer sind dies die typischen funktionalen und ergonomischen Anforderungen zum Erfassen und Verwalten von Studiendaten, die jedoch um Integritäts- und Plausibilitätsroutinen sowie eine benutzerfreundliche, durchgängige Änderungssensitivität erweitert wurden. Hinzu kommt die Möglichkeit einer intuitiven Stammbaumerfassung.

Unter der Perspektive der Systemverwaltung und -administration muss ein solches System an zentrumspezifische Anforderungen konfigurierbar, an unterschiedlich große Datenvolumina skalierbar und grundsätzlich releasefähig auslieferbar sein. Um Rekonstruierbarkeit und Authentizität der erfassten Daten zu gewährleisten, sind neben einer integrierten Benutzerverwaltung eine automatische Lös- und Änderungshistorie sowie ein Backup- und Archivierungskonzept der Daten zu implementieren.

Hinsichtlich Datenschutz- und Datensicherheit müssen neben den grundsätzlich für ein im klinischen Umfeld erforderlichen Kontrollmechanismen die folgenden Anforderungen implementiert werden: Die zentrumsübergreifende Datenauswertung erfordert eine automatische und integrierte Anonymisierung bzw. Pseudonymisierung der Transferdaten, bei der die Re-Identifikation eines Pseudonyms ausschließlich im jeweiligen Zentrum möglich ist. Da an dem System Vertreter verschiedener wissenschaftlicher Fachdisziplinen arbeiten, ist ein integriertes, rollenbasiertes Benutzerkonzept notwendig, das den unterschiedlichen Zugriffsberechtigungen zwischen den beteiligten Bereichen Rechnung trägt.

Modell

Anschließend wird ein Modell eines Datenmanagementsystems innerhalb einer Client-Server Architektur dargestellt, das die drei Ebenen Präsentation, Anwendung und Datenverwaltung umfasst. Die Präsentationsebene realisiert die Mensch-Computer Schnittstelle, insbesondere die Fenster-basierte Art und Weise der Dateneingabe und -darstellung. Die Anwendungsebene beinhaltet die eigentliche Programm- und funktionale Logik des Datenmanagementwerkzeugs, wie etwa den Ablauf bei der Dateneingabe, Plausibilitätsroutinen, Datenschutz, Benutzerverwaltung, Pseudonymisierung, Datensicherung und Datenexport. Als Bindeglied zwischen Präsentation und Datenverwaltung leistet diese Ebene die Umsetzung benutzerfreundlicher Darstellungsstrukturen in effizient verwaltbare Datenstrukturen.

Neben einem Programmiermodell für diese drei Ebenen und deren Schnittstellen wird eine allgemeine Datenmodellierungsmethodik präsentiert, die es erlaubt, beliebige Studienvariablen bereichsspezifisch innerhalb eines relationalen Datenmodells darzustellen und in relationalen Datenbanksystemen zu realisieren. Für die einzelnen Kategorien von Studienvariablen wird neben der datenbankseitigen Umsetzung auch deren Datenstruktur in der Anwendungsebene sowie die Art der Darstellung in der Präsentationsebene aufgezeigt.

Implementierung

Die Beschreibung einer konkreten Implementierung des Client-Server Datenmanagementsystems FAMOS für die Studie des Konsortium der Dt. Krebshilfe "Hereditäres Mamma- und Ovarialkarzinom" erfolgt im letzten Kapitel. Das Programmpaket FAMOS besteht aus einem Microsoft 32-Bit Windows Programm, das auf verschiedene relationale Datenbanksysteme zugreifen kann. In den einzelnen Studienzentren wird als Datenbank Microsoft-Access 97 verwendet, in der Datenzentrale am DKFZ existiert zusätzlich eine Implementierung mit ORACLE 8.0. Design und Layout des Programms folgen dem Microsoft-Windows Application Style Guide.