

# **Betreuung klinischer Langzeitstudien mit SAS am Methodischen Zentrum in Freiburg**

Ina Gäbler

Institut für Medizinische Biometrie und Medizinische Informatik,  
Methodisches Zentrum, Universitätsklinikum Freiburg  
Telefon: 0761 / 203 - 6682  
eMail: ig@imbi.uni-freiburg.de

## **Abstract**

In folgendem Beitrag soll die Arbeit des Methodischen Zentrums bezüglich der Entwicklung eines Datenmanagement-Konzeptes und der Umsetzung mit SAS dargestellt werden. Dazu wird zunächst kurz das Methodische Zentrum vorgestellt, bevor daraufhin in einem zweiten Teil auf die Entwicklung des Datenmanagement-Konzeptes und das erste Grobkonzept eingegangen wird, welches im Rahmen einer Diplomarbeit entwickelt wurde. Im dritten Teil wird auf den derzeitigen Stand der Entwicklung und die bisherige Umsetzung mit SAS eingegangen.

## **Das Methodische Zentrum in Freiburg**

Das Methodische Zentrum in Freiburg ist eine Untereinheit des Instituts für Medizinische Biometrie und Medizinische Informatik des Universitätsklinikums Freiburg und gehört zur Abteilung Medizinische Biometrie und Statistik. Das Methodische Zentrum beschäftigt sich seit über 10 Jahren mit der methodischen und biometrischen Betreuung bei der Durchführung klinischer Langzeitstudien, sowie deren Planung und Auswertung. Die dort durchgeführten Studien sind schwerpunktmäßig prospektive multizentrische klinische Studien der Phase III, bei denen es um den Therapievergleich bei Mammakarzinomen geht.

Das Methodische Zentrum arbeitet mit UNIX-Workstations unter dem Betriebssystem Solaris 2.7 der Firma SUN-Microsystems. Die Daten zu den einzelnen Studien werden auf einer speziell geschützten Workstation aufbewahrt und ausschließlich mit SAS 6.12 bearbeitet. Für die Textverarbeitung steht dem Methodischen Zentrum innerhalb des UNIX-Netzes ein PC zur Verfügung, der den MitarbeiterInnen die Verwendung des Office-Paketes von Microsoft erlaubt.

## **Entwicklung eines Datenmanagement-Konzeptes**

Bei der Betreuung der klinischen Studien am Methodischen Zentrum wurde das ursprüngliche Datenmanagement-Konzept [1] im Laufe der Zeit fortentwickelt, um dieses an hinzukommende Anforderungen bezüglich der Durchführung von klinischen Studien anzupassen. Dies geschah zunächst im Rahmen des normalen Arbeitsablaufes bezüglich der EDV-Betreuung, schließlich wurde aber ein kompletter Neuentwurf im Rahmen einer Diplomarbeit im Studiengang „Biowissenschaftliche Dokumentation“ beschlossen. Die Aufgabenstellung für die Diplomandin lautete „Entwurf eines Daten- und Prozedurmodells zur Dateneingabe und Datenhaltung in klinischen Lanzeitstudien“. Die Diplomarbeit wurde in der Zeit von Juni bis Oktober 1999 verfaßt.

Im folgenden werden zunächst kurz die Vorarbeiten dargestellt, die im Rahmen der sogenannten Analysephase dem Entwurf des neuen Konzeptes vorangingen. Im zweiten Teil des Abschnittes soll das grobe Ablaufmodell vorgestellt werden, welches in der Diplomarbeit entwickelt wurde.

### **Analysen zum Entwurf eines neuen Datenmanagement-Konzeptes**

Zunächst war es erforderlich, sich mit Methoden zu beschäftigen, die in der Regel angewendet werden, um ein System, wie beispielsweise ein Datenmanagement-Konzept entwerfen zu können. Als Ergebnis dieser Untersuchungen wurde nach Erkenntnissen der Wirtschaftsinformatik, genauer der Systemplanung vorgegangen, die den Entwurf eines Systems in eine Analyse-, eine Entwurfs- und eine Implementierungsphase aufteilt, in denen jeweils bestimmte Aufgaben anfallen [3].

Zur Dokumentation der Anforderungen an ein neues Datenmanagement-Konzept wurde im Rahmen der Analysephase vom Methodischen Zentrum bereits vor der Diplomarbeit ein Anforderungsprofil formuliert, welches drei Themenbereiche umfaßte. Als äußerer Rahmen der Neuentwicklung sollte dabei erstens beachtet werden, daß das neue Konzept mit den nationalen und internationalen Bestimmungen, insbesondere den ICH-GCP-Guidelines [4] abgestimmt ist. Desweiteren sollten zweitens gewisse organisatorische Rahmenbedingungen in den Entwurf einbezogen werden. Insbesondere sollte der gesamte Arbeitsablauf des Datenmanagements vereinfacht und das Datenmanagement-Konzept dazu möglichst modular aufgebaut werden, um möglichst viele Aufgaben unabhängig voneinander durchführen zu können. So viele einzelne Prozesse wie möglich sollten automatisiert durchgeführt werden. Zusätzlich sollte das neue Konzept eine schnelle elektronische Verfügbarkeit der Daten, sowie die präzise Kontrolle über den jeweils aktuellen Stand der Datenqualität erlauben. Dies sollte die Verfügbarkeit von Statusinformationen zu Dokumentationsbögen (Case Report Forms, kurz: CRFs), zu einzelnen Variablen auf den CRFs, sowie zu Rückfragen zu Patientendaten an Prüfarzte (Queries) einschließen. Sämtliche Korrekturen an den Daten sollten in Form eines Audit-Trails nachvollziehbar sein. Schließlich wurden drittens detaillierte Angaben zu konkreten Tätigkeitsbereichen beim Datenmanagement gemacht, die die Dateneingabe sowie die Kontrolle der Daten betreffen.

Zusätzlich zum Anforderungsprofil des Methodischen Zentrums mußte als Vorarbeit im Rahmen der Diplomarbeit zusätzlich das alte Datenmanagement-Konzept untersucht werden, um beim Neuentwurf die Schwachstellen des alten Konzeptes beachten und weitestgehend umgehen zu können. Zum Entwurf des neuen Datenmanagement-Konzeptes wurde die Betreuung von klinischen Studien in einzelne Tätigkeitsbereiche zerlegt, diese Arbeitsabläufe untersucht und mit Arbeitsabläufen anderer Datenmanagementzentren verglichen. Die Entwicklung des neuen Datenmanagement-Konzeptes sollte aus zwei Gründen bereits von Anfang an von Überlegungen zu einer eventuellen Umsetzung mit SAS begleitet sein. Zunächst war das ursprüngliche Datenmanagement-Konzept bereits mit SAS umgesetzt worden und desweiteren ist beim Entwurf von neuen Systemen im allgemeinen darauf zu achten, daß nicht zu viele gleichzeitige Veränderungen in einem laufenden Betrieb durchgeführt werden müssen, wenn die Ressourcen zur Umsetzung begrenzt sind.

Durch Berücksichtigung der GCP-Guidelines, der Anforderungen des Methodischen Zentrums sowie der Schwachstellen des alten Datenmanagement-Konzeptes und durch Abwägen der unterschiedlichen Umsetzungsmöglichkeiten bezüglich der Arbeitsablaufgestaltung entstand schließlich als Ergebnis der Diplomarbeit ein Grobentwurf für ein neues Datenmanagement-Konzept, welches zur Zeit Schritt für Schritt verfeinert wird und teilweise bereits umgesetzt ist.

## Ablaufmodell für das neue Datenmanagement-Konzept

Als Ergebnis der oben beschriebenen Analysephase wurde folgende Vorgehensweise bei der Betreuung klinischer Studien am Methodischen Zentrum beschlossen.

Das Datenmanagement-Konzept unterscheidet zunächst grundsätzlich zwischen zwei verschiedenen Ablaufmodellen für Dokumentationsbögen und Queries, die sich gegenseitig beeinflussen.

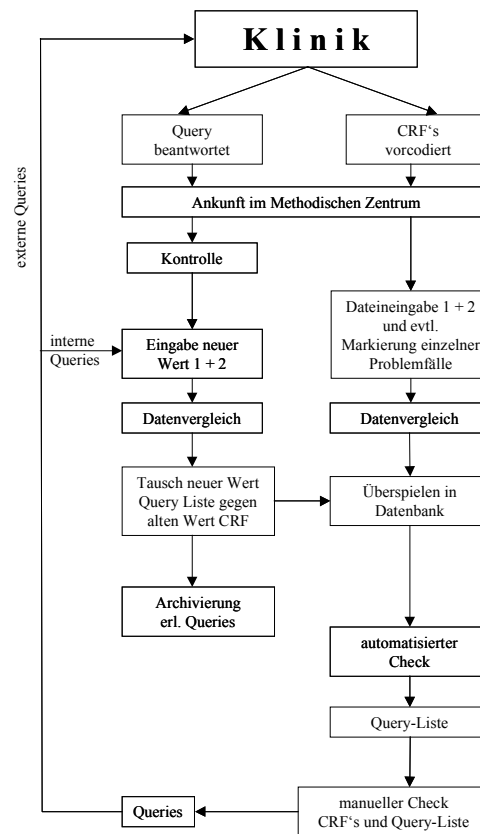


Abbildung: Ablaufplan Datenmanagement-Konzept

Die Dokumentationsbögen werden für zukünftige Studien vorcodiert gestaltet, so daß die Dateneingabe ohne den vorherigen Codiervorgang eines Dokumentars durchgeführt werden kann, was bisher viel Zeit in Anspruch genommen hat. Auch der dabei bisher durch visuelle Kontrolle durchgeführte Check der Patientendaten auf Unvollständigkeiten oder Implausibilitäten wird auf einen Zeitpunkt nach der Dateneingabe verschoben und dann weitestgehend automatisiert durchgeführt. Es findet nach wie vor eine doppelte Dateneingabe statt. Durch Datenvergleich aufgedeckte Diskrepanzen werden vom Dokumentar beseitigt.

Nach Beseitigung aller Diskrepanzen werden die Patientendaten aus den temporären Eingabedateien in die Studiendatenbank überspielt, wo sie dann mittels automatisierter Checkprogramme auf Konsistenz und Vollständigkeit überprüft werden. Patientendaten, die sich aufgrund dieser Plausibilitätschecks als implausibel herausgestellt haben, werden in einer sogenannten Query-Liste aufgeführt, die ebenfalls von diesem Checkprogramm erzeugt wird.

Jedes Query in dieser Liste wird mit einer eindeutigen identifizierenden Nummer geführt. Zu einem Query gehören weiterhin die Patienten-, die Dokumentationsbogen-, sowie die Feldidentifikation als weitere identifizierende Kennzeichen. Desweiteren werden die zu korrigierenden Daten und ein erläuternder Text aufgeführt. Jeder Dokumentationsbogen wird danach nochmals visuell gecheckt, wobei auch die Möglichkeit für den Dokumentar besteht,

die Query-Liste manuell nachzubearbeiten. Dies soll möglich sein, um eventuell Queries anzufügen, die vom automatisierten Datencheck nicht erkannt wurden, bzw. um Queries zu entfernen, die aus verschiedenen Gründen bereits vom Dokumentar geklärt werden konnten.

Letztere Queries werden in der Query-Liste als interne Queries gekennzeichnet. Basierend auf der Query-Liste werden die Queries automatisiert erzeugt, welche an die Prüfarzte gesendet werden sollen. Wenn beantwortete Queries im Methodischen Zentrum eintreffen, werden diese vom zuständigen Dokumentar überprüft und zur Dateneingabe weitergeleitet, die ebenfalls zweifach durchgeführt wird.

Daraufhin wird der korrigierte Wert gegen den ursprünglichen Wert in der Studiendatenbank ausgetauscht. Das Query wird in ein sogenanntes Query-Archiv überspielt und beschreibt durch Anfügen einiger weiterer Daten, wie beispielsweise Postausgangsdatum, Posteingangsdatum, der Identifikation des bearbeitenden Dokumentars oder dem Datum der Bearbeitung den gesamten Prozeß der Korrektur von Patientendaten als Audit-Trail.

## **Umsetzung des Datenmanagement-Konzeptes**

In diesem Abschnitt wird beschrieben, wie das eben vorgestellte Grobkonzept zur Zeit nun außerhalb der Diplomarbeit im Methodischen Zentrum weiterentwickelt und in EDV-Strukturen umgesetzt wird. Dazu hat sich ein Team aus zwei Dokumentarinnen, zwei Statistikerinnen und einer SAS-Programmiererin gebildet, die in regelmäßigen Meetings die weiteren Schritte besprechen.

### **Vorgehen**

Nach dem Entwurf des Grobkonzeptes stand zunächst die Entscheidung darüber an, für welche Studien das neue Datenmanagement-Konzept bereits eingesetzt werden kann und soll. Es wurde entschieden, daß keine der bereits laufenden Studien davon beeinflusst werden sollte. Stattdessen sollten aber alle neuen Studien direkt nach den Ideen des neuen Datenmanagement-Konzeptes gehandhabt werden.

Nach einiger Zeit wurde allerdings festgestellt, daß dieses Vorhaben mit den verfügbaren Ressourcen aus Zeitgründen nicht in die Tat umzusetzen war, da sich herausstellte, daß die Konkretisierung des Grobkonzeptes noch viele Arbeitsschritte beinhaltet. Daraufhin wurde dann beschlossen, das neue Datenmanagement-Konzept zunächst nur an einer einzigen neuen Studie im Rahmen eines Pilotprojektes umzusetzen.

Um die Entwicklung nicht unter einen zeitlichen Druck zu setzen, wird diese Studie zur Zeit parallel nach dem alten und nach dem neuen Datenmanagement-Konzept betreut. Einerseits folgt die Betreuung also den bereits bestehenden Arbeitsabläufen, wurde aber insoweit an das neue Datenmanagement-Konzept angepaßt, wie es notwendig schien, um die Studie später komplett auf das neue Konzept umstellen zu können. Parallel wird die Studie nochmals gänzlich nach den Ideen des neuen Datenmanagement-Konzeptes aufgebaut. Dies hat neben der Entwicklungsmöglichkeit ohne Zeitdruck den Vorteil, daß Ergebnisse aus Arbeitsprozessen mit dem neuen Datenmanagement-Konzept direkt verglichen werden können.

## Stand der Implementierung

Bisher wurde ausschließlich der Teil des neuen Konzeptes detaillierter ausgearbeitet und implementiert, der die Dateneingabe im weitesten Sinne betrifft.

Der Arbeitsaufwand erstreckte sich zunächst auf den Entwurf eines konsistenten Codiersystems. Die bedeutendste Änderung zum bestehenden Codiersystem bestand darin, daß bei fehlenden Patientendaten nun zwischen fehlenden Werten, die sich aus der Gestaltung der Dokumentationsbögen ergeben (logische missings), und solchen missings unterschieden wird, bei denen vom Prüfarzt bestätigt ist, daß ein erforderlicher Wert nicht bekannt ist. Letztere missings sollen in Zukunft im Unterschied zum alten Datenmanagement-Konzept für jede Art von Feld, d.h. Textfelder, numerische Felder oder Datumsfelder identisch codiert werden.

Dazu bietet SAS die Möglichkeit, die Kombination eines beliebigen Buchstabens mit einem vorausgehenden Punkt als „missing value“ zu verwenden. Für einstellige Felder besteht die Möglichkeit, die entsprechende Punkt-Buchstaben-Kombination als globales missing statement festzulegen. SAS erkennt daraufhin die alleinige Angabe des entsprechenden Buchstabens als missing.

Insbesondere für die Dateneingabe sollen komfortable Eingabemasken zur Verfügung gestellt werden, da diese wie auch die zweifache Dateneingabe dazu beitragen, Fehler bei der Dateneingabe zu vermeiden. Im Methodischen Zentrum in Freiburg werden zur Anwendungsprogrammierung in diesem Bereich die SAS-Module FSP und AF verwendet.

Innerhalb des Moduls FSP lassen sich mit der Prozedur „proc fsedit“ und der Angabe eines SAS-Datafiles sowie der Angabe eines dazugehörigen SAS-Screens Dateneingabemasken erstellen und modifizieren. Zur Verfügung gestellt werden die Dateneingabemasken im Methodischen Zentrum über das SAS-Modul AF, mit dem eine objektorientierte Programmierung möglich ist. Mit der Prozedur „proc build“ lassen sich sogenannte Frames erzeugen, auf denen verschiedene Objekte plaziert werden können.

Eines der wichtigsten zur Zeit verwendeten Objekte sind die sogenannten „Push-Buttons“, die beim bisherigen Stand der Umsetzung hauptsächlich zum Aufrufen eines anderen Frames verwendet werden, so daß sich das Dateneingabepersonal zur Dateneingabemaske „hinklicken“ kann. Desweiteren ist es mit der Prozedur „proc pmenu“ des SAS-BASE-Moduls möglich, Menüleisten zu erstellen, die in die Dateneingabemasken eingebunden werden können. Das Methodische Zentrum verwendet diese selbsterstellten Menüleisten, um dem Dateneingabepersonal eine komfortable Möglichkeit zur Navigation innerhalb einer Dateneingabemaske zur Verfügung zu stellen.

Die Dateneingabe wird durch sogenannte Hintergrundprogramme überprüft, die ebenfalls innerhalb von SAS-FSP hinterlegt werden können und mit einer Dateneingabemaske verknüpft sind. Zur Programmierung dieser Hintergrundprogramme wird die Screen Control Language (SCL) verwendet. Diese Programme lassen nur die Eingabe von Werten zu, die laut Codieranleitung eines entsprechenden Dokumentationsbogens zugelassen sind. Es ist auch möglich, Wertebereiche durch sogenannte „Attributes“ der FSP-Eingabemaske abzufragen, aber das Methodische Zentrum hat sich für die Lösung mit Hilfe der Hintergrundprogramme entschieden, da sich auf diesem Wege die Abfragen und Änderungen besser dokumentieren lassen.

Zur Zeit wird nun die Checkphase für die Patientendaten detaillierter ausgearbeitet. Dazu ist es zunächst notwendig, die Checks konzeptionell zu entwerfen und ihnen eine Struktur zu verleihen, anhand derer jederzeit nachvollziehbar ist, welches Query zu welchen Variablen auf einem Dokumentationsbogen aufgrund welchen Checks gestellt wurde. Im Moment wird eine Checkliste entwickelt, die alle durchzuführenden Datenchecks zu den CRFs der Pilotstudie auflistet. Aus der Checkliste gehen die betroffenen Variablen, eine eindeutige Checknummer sowie der eigentliche Datencheck hervor. Da das Checkprogramm gleichzeitig die Query-Liste generieren soll, wurde es zu diesem Zeitpunkt bereits notwendig, die gewünschte Struktur der Query-Liste festzulegen.

Letzlich wurde auch das System zum Aufbau der Datenbank sowie zum Aufbau der Codieranleitungen und Checklisten für die CRFs überarbeitet. Ursprünglich wurden die entsprechenden Daten CRF-weise vorgehalten, nun ist neuerdings die Art des CRFs für die Zuordnung zu einer Codieranleitung, einer Checkliste oder einem Speicherort für die Patientendaten ausschlaggebend. Um dies zu erreichen, werden identische oder ähnliche CRFs zu einer „Art CRF“ zusammengefaßt. Die Codieranleitung und die Checkliste liegen jeweils für eine „Art CRF“ vor und die Patientendaten werden in einer Dateneingabemaske für eine „Art CRF“ eingegeben und in einem SAS-Dataset für eine „Art CRF“ gespeichert. Die einzelnen CRFs sind über ihre Nummer identifizierbar.

## Die Zukunft

Zukünftige Projekte, die noch nicht begonnen sind, betreffen die Programmierung des Überspielprogrammes der Patientendaten aus den Eingabedateien in die Studiendatenbank, die Generierung der Queries, die dem Prüfarzt vorgelegt werden sollen, aus der Query-Liste sowie die Programmierung des Query-Archivierungsprogrammes, welches gleichzeitig den alten Wert in der Datenbank gegen den neuen korrigierten Wert austauschen soll.

Im Zeitplan zur weiteren Umsetzung des Konzeptes ist vorgesehen, daß möglichst bis Sommer diesen Jahres zumindest ein gesamter Probedurchlauf mit wenigen ausgesuchten Daten stattgefunden haben soll, d.h. erste Queries sollten in der Pilotstudie bis ins Query-Archiv gelangt sein.

## Referenzen

- [1] Schmoor et al. (1997). Arbeitsweisen des Methodischen Zentrums am Institut für Medizinische Biometrie und Medizinische Informatik der Universität Freiburg bei der biometrischen Betreuung klinischer Studien. In Informatik, Biometrie und Epidemiologie in Medizin und Biologie, 4: 253-274
- [2] Gäbler (1999). Entwurf eines Datenmanagement-Konzeptes für das Methodische Zentrum in Freiburg. Diplomarbeit, Fachhochschule Hannover
- [3] Schwarze (1995). Systementwicklung: Grundzüge der wirtschaftlichen Planung, Entwicklung und Einführung von Informationssystemen.
- [4] ICH Steering Committee (1996). ICH Harmonised Tripartite Guideline: Introduction.