

Volker Mludek  
Dr. med.

## **Modellierung der Integritätsbedingungen eines rechnerunterstützten Dokumentationssystems zur Studiendokumentation in einem Verbund multizentrischer Therapiestudien der Pädiatrischen Onkologie**

Geboren am 31.01.1970 in Kassel  
Reifeprüfung am 17.05.1989 in Idar-Oberstein  
Studiengang der Fachrichtung Medizin vom WS 1989 bis WS 1996  
Physikum am 04.09.1991 an der Universität Bochum  
Klinisches Studium in Essen und Homburg  
Praktisches Jahr in Neunkirchen/Saar und Ashton-under-Lyne (Großbritannien)  
Staatsexamen am 07.11.1996 an der Universität des Saarlandes, Homburg

Promotionsfach: Medizinische Biometrie und Informatik  
Doktorvater: Prof. Dr. rer. nat. T. Wetter

Über 90% der Patienten der Pädiatrischen Onkologie werden innerhalb von deutschlandweiten multizentrischen Therapieoptimierungsstudien unter dem Dach der Gesellschaft für Pädiatrische Onkologie und Hämatologie (GPOH) behandelt.

Es ist das erklärte Ziel der GPOH, die Datenverarbeitung und hier insbesondere die für die erhobenen Merkmale verwendete Terminologie durch Verwendung eines sog. „kontrollierten Vokabulars“ zu vereinheitlichen. Zu diesem Zweck wurde schon 1994 von der GPOH ein sog. „Basisdatensatz“ für die Pädiatrische Onkologie verabschiedet, der eine Menge von Merkmalen enthält, die in allen Studien erhoben werden sollen. Eine im Jahr 2001 aus terminologischen Gesichtspunkten überarbeitete Version des Basisdatensatzes liegt zur Zeit den Studienleitern der GPOH zur Verabschiedung vor.

Auf Grundlage des Basisdatensatzes wird seit einigen Jahren vom Institut für Medizinische Biometrie und Informatik, Abteilung Medizinische Informatik, der Universität Heidelberg in Zusammenarbeit mit der GPOH ein computerbasiertes Dokumentationssystem für die Pädiatrische Onkologie (DOSPO) entwickelt, das deutschlandweit in den behandelnden Kliniken eingesetzt werden soll. Mit DOSPO soll von der heute führenden papierbasierten Dokumentation auf computerbasierte Dokumentation und Datenübermittlung umgestellt werden.

Das Gesamtkonzept der GPOH für das Projekt DOSPO sieht vor, daß in Zukunft über DOSPO allen Studien ein kontrolliertes Vokabular zur Verfügung gestellt wird. Hierzu wird ein zentrales Merkmalsverzeichnis entwickelt, das nicht nur die Merkmale des Basisdatensatzes, sondern darüber hinaus alle weiteren Merkmale für alle Studien der GPOH enthalten soll.

Ein Entwicklungsschritt im DOSPO-Projekt ist die Einbindung von Plausibilitätsprüfungen des Basisdatensatzes durch die Definition sog. „Integritätsbedingungen“. Das Ziel der vorliegenden Arbeit war es, ein allgemeines „Modell für Integritätsbedingungen“ zu entwerfen, das die vorkommenden Integritätsbedingungen in verschiedene Klassen einteilt und den Aufbau der verschiedenen Integritätsbedingungen anhand definierter Eigenschaften beschreibt. Das Modell soll die Voraussetzungen für die Integration der Plausibilitätsprüfungen in DOSPO bzw. in das Merkmalsverzeichnis schaffen. Formale Fehler in der Studiendokumentation sollen in Zukunft schon bei der Eingabe erkannt und so der Datenverarbeitungsprozeß der Studien unterstützt werden.

Ergebnis der Arbeit ist ein umfangreiches Modell für Integritätsbedingungen, das mittels Aussagenlogik schlüssig abgeleitet und notiert wurde. Das Modell unterscheidet auf oberster Ebene die Klassen „Art der Integritätsbedingung“ und „Ordnung der Integritätsbedingung“. Die Klasse „Art der Integritätsbedingung“ teilt sich in 5 Subklassen: Die Integritätsbedingungen Datentyp unterscheiden die Merkmalsarten nach den Typen „Text“, „Zahl“, „Datum“, etc.. Die Integritätsbedingung Länge prüft die Einhaltung einer maximalen Länge der Merkmalsausprägungen. Semantische Integritätsbedingungen hingegen prüfen sowohl die Plausibilität einer Merkmalsausprägung anhand von oberen und unteren Grenzen (im Sinne von Normwerten o.ä.) als auch „semantische Zusammenhänge“ zwischen Merkmalsausprägungen verschiedener Merkmalsarten. Vollständigkeitsbedingungen kontrollieren die korrekte Anzahl an Merkmalsausprägungen zu einer Merkmalsart. Mischformen enthalten sowohl Semantische Integritätsbedingungen als auch Vollständigkeitsbedingungen. Die Klasse „Ordnung der Integritätsbedingung“ unterscheidet zwischen Integritätsbedingungen Nullter Ordnung und Integritätsbedingung Höherer Ordnung. Unterscheidungskriterium hierbei ist, ob die Definition nur auf Ebene genau einer Merkmalsart (ohne Berücksichtigung der Merkmalsausprägungen anderer Merkmalsarten) stattfindet oder ob sie eine Beziehung zwischen mehreren Merkmalsarten beschreibt. Da eine Merkmalsart (z.B. „Uhrzeit“) in mehreren unterschiedlichen Bedeutungen in einem Dokumentationssystem vorkommen kann (in der Bedeutung Uhrzeit der Messung, Uhrzeit der Medikamentenapplikation), mußte ein Unterscheidungsmöglichkeit geschaffen werden. Dazu wurde das Konstrukt „Semantischer Kontext“ eingeführt. Ein „Semantischer Kontext“ faßt mehrere Merkmalsarten zu logischen Einheiten zusammen, die einen bestimmten semantischen Zusammenhang aufweisen (z.B. „Datum, Uhrzeit, Medikamentapplikation“, „Datum, Uhrzeit, Blutdruckmessung“). Die Art und Weise, wie der Semantische Kontext in die Definition einer der Integritätsbedingung Höherer Ordnung eingebunden ist, führt zur Definition weiterer Unterklassen, die mit „Typ A“ und „Typ B“ bezeichnet werden. Weitere Eigenschaften, die eine Integritätsbedingung Höherer Ordnung näher charakterisieren, wurden definiert. Hier sind Eigenschaften wie Hauptverknüpfung (der logischen Aussage der Integritätsbedingung), Subtyp und Spannweite zu nennen, die hier nicht weiter erläutert werden sollen.

Eine nicht erfüllte Integritätsbedingungen gibt entweder eine Warnung zurück, die der Benutzer ignorieren kann, oder sie meldet einen Fehler, der korrigiert werden muß.

Das komplexe „Modell für Integritätsbedingungen“ wurde über die Aussagenlogik mit der Unified Modeling Language (UML) dokumentiert, einer graphischen Sprache zur Notation von Modellen. Ergebnis dieser Arbeit sind Diagramme, die in Kombination mit einem erarbeiteten Datenbankentwurf die weiteren Entwicklungsarbeiten zur Einbindung der Integritätsbedingung unterstützen werden. Die aussagenlogischen Formeln lassen sich mit Hilfe der Diagramme sowohl für den Anwender als auch für den Entwickler leichter interpretieren, indem sie zusätzlich eine graphische Visualisierung des „Modells für Integritätsbedingungen“ zur Verfügung stellen. Damit ist - etwa zur Planung der weiteren Arbeiten - eine Einarbeitung in das komplexe aussagenlogische Modell primär nicht erforderlich.

Auf Basis des Modells wurden ca. 1250 Integritätsbedingungen für den Basisdatensatz (Version 2001) definiert. Davon waren ca. 960 Integritätsbedingungen Nullter Ordnung und ca. 290 Höherer Ordnung, also unter Beteiligung mehrerer Merkmalsarten in mehreren Semantischen Kontexten. Alle Integritätsbedingungen ließen sich auf Basis des aussagenlogischen Modells darstellen. Es zeigte sich, daß einige Klassen für die Definition nicht benötigt wurden. Diese Klassen können ggfl. zur Definition der Integritätsbedingungen für weitere Merkmalsarten der einzelnen Studien zur Verfügung stehen.

Zusammenfassend zeigt das aussagenlogische Modell eine Möglichkeit auf, Integritätsbedingungen des Basisdatensatzes einheitlich zu definieren. Diese Notation erleichtert eine Integration der Integritätsbedingungen in das DOSPO-Kernsystem und in das Merkmalsverzeichnis, da die Elemente der Aussagenlogik in allen üblichen Programmiersprachen zur Verfügung stehen. Die Einbindung der Integritätsbedingungen in das Dokumentationssystem DOSPO unterstützt die Datenverarbeitungsprozesse der Studien sowohl in den behandelnden Kliniken als auch in den jeweiligen Studienzentralen. Durch Optimierung der primären Datenqualität ist auch ein reduzierter Korrekturaufwand zu erwarten. Die Integration der Integritätsbedingungen hilft ferner, die Studiendokumentation zu vereinheitlichen, da das Dokumentationssystem den Benutzer auf eine nicht standardgemäße Dokumentation aufmerksam machen kann.

Ob mit den jetzt vorliegenden Integritätsbedingungen alle erforderlichen Bedingungen bereits definiert sind, kann zum gegenwärtigen Zeitpunkt noch nicht ermittelt werden, da die Entwicklungen im DOSPO-Projekt noch nicht weit genug fortgeschritten sind. Ggfl. erforderliche Erweiterungen oder Optimierungen der Bedingungen für die speziellen Bedürfnisse einzelner Studien der GPOH sollten anhand des Modells mit wenig Aufwand möglich sein.

Ehe die Neudefinition und Wartung der Integritätsbedingungen wie geplant in die Hände der Studienzentralen übergehen kann, wird die Entwicklung eines Regeleditors erforderlich werden, der für das komplexe aussagenlogische Modell eine leicht verständliche Benutzeroberfläche bietet und die komplexen Vorgänge bei der Erstellung der aussagenlogischen Integritätsbedingungen automatisiert.

Da die dem „Modell für Integritätsbedingungen“ zugrunde gelegten Klassen der Integritätsbedingungen nicht spezifisch für das Dokumentationssystem DOSPO oder das Merkmalsverzeichnis sind, ist zu erwarten, daß das Modell auch in anderen Dokumentationssystemen vorteilhaft eingesetzt werden kann.