

Delegation von Datenmanagement in Szenarien verteilter Verantwortlichkeiten

Ralph Stuber
OFFIS Institut für Informatik
Escherweg 2
26121 Oldenburg
+49 441 9722 141
stuber@offis.de

ABSTRACT

Um Daten aus verschiedenen Datenbeständen miteinander in Verbindung bringen und daraus zusätzliche Informationen gewinnen zu können, werden diese oftmals aus verschiedenen Quellen heraus in einem gemeinsamen Datenbestand integriert. In Fällen, in denen die Personen oder Institutionen, die die Quelldaten erzeugt haben und verantworten, nicht in den Prozess der Erstellung und Nutzung integrierter Datenbestände eingebunden werden, erhalten diese erst nach Veröffentlichung des Datenbestandes Einsicht in die Daten und können somit Änderungs- und Korrekturwünsche an den Quelldaten nicht in den vorgelagerten Integrationsprozess einbringen. So resultiert aus dem Erfordernis der Durchführung von Datenänderungen aufgrund von Aktualisierungen oder Korrekturen von Informationen der Bedarf des Delegierens von Datenpflege-Prozessen, welche es den Datenverantwortlichen ermöglicht, von ihnen verantwortete Daten im Zielsystem integrationsnachgelagert zu pflegen und ggf. zu aktualisieren.

Zur Lösung des Problems wird im Rahmen dieses Beitrages das Vorgehensmodell VD2M (Vorgehensmodell zur Delegation von Datenmanagement) vorgestellt, welches das angesprochene Delegieren von Datenmanagement unter Berücksichtigung verteilter Verantwortlichkeiten ermöglicht. Am Beispiel eines Projektes aus dem Gesundheitswesen wird dargestellt, wie das Vorgehensmodell anschließend durch den Einsatz der prototypisch implementierten Softwarewerkzeug-Sammlung WD2M (Werkzeuge zur Delegation von Datenmanagement) umgesetzt und evaluiert werden kann.

Keywords

Delegation, Datenmanagement, integrationsnachgelagerte Datenänderungen, Verantwortlichkeiten, Schreibender Zugriff auf Data Warehouses.

1. Einführung und Motivation

Die meisten Unternehmen, Institutionen, Verbände und andere Organisationen verfügen über eine Vielzahl an verschiedenen Datenbeständen. Diese wachsen zum einen stetig quantitativ bezüglich der jeweils enthaltenen Datenmengen, zum anderen jedoch meist auch qualitativ in der Komplexität und der Anzahl verschiedener, durch die Daten abgebildeter Entitäten und Beziehungen.

Da unterschiedliche Nutzerkreise oder Interessenten aufgrund stetig wachsender Datenmengen oftmals nur Zugriff auf

Teilbereiche der umfangreichen Datenbestände benötigen, gewinnt die Generierung verschiedener Sichten auf einzelne Datenbereiche zunehmend an Bedeutung [1].

Integrierte Datenbestände

Weiter existieren Datenbestände über die Grenzen von Unternehmen oder Institutionen hinaus, die sich über gemeinsame Dimensionen oder Attribute miteinander verknüpfen lassen. Um das Potential der in den verschiedenen Datenbeständen vorliegenden Informationen nutzen zu können, werden derartige Datenbestände oftmals aus verschiedenen Quellen unter Einsatz von Data Warehouse-Technologien [2] integriert und so in einem gemeinsamen Datenbestand zusammengeführt (Datenintegration). Institutionen und Unternehmen unterschiedlicher Branchen nutzen die Möglichkeit, die auf diese Weise zusammengetragenen Kennzahlen, Leistungsdaten und anderen Strukturinformationen zu verschiedenen Zwecken zu nutzen. So können integrierte Datenbestände für Analysen zum Zwecke der Unterstützung in Management-Entscheidungen genutzt werden, oder z.B. zur Präsentation von Produkten, Dienstleistungen und Unternehmensstatistiken in aufbereiteter Form beispielsweise über das World Wide Web.

Verteilte Verantwortlichkeiten

Grundsätzlich können die Rollen Datenurheber und Datenverwalter unterschieden werden. Nicht zwingend werden beide Rollen durch ein und dieselbe Person oder Institution wahrgenommen. Weisen verschiedene Datenquellen, die den integrierten Datenbeständen als Basis zugrunde liegen können, eine Urheberschaft durch verschiedene Daten verantwortliche Personen auf, so lassen sich Szenarien verteilter Verantwortlichkeiten beobachten.

In solchen Situationen, in denen Datenurheber und Datenverwalter verschiedene Personen oder Institutionen sind, werden die Datenurheber, die die Quelldaten verantworten, selten in den Prozess der Erstellung oben beschriebener Datenbestände durch Dritte eingebunden. Sie erhalten daher erst nach Veröffentlichung des integrierten Datenbestandes die Möglichkeit zur Einsicht, und haben somit meist keine Möglichkeit, Änderungs- und Korrekturwünsche an den Quelldaten in den vorgelagerten Integrationsprozess einzubringen. Der Bedarf an integrationsnachgelagerten Datenänderungen entsteht. Weiter ist insbesondere zu beachten, dass die Personen, die die Daten aus den Quellen der Integration verantworten, selten über detaillierte

Kenntnis der Struktur des zugrunde liegenden Datenmodells und Datenbankschemas verfügen.

Integrationsnachgelagerte Datenänderungen

Unterliegen Datenquellen stetigen Veränderungen, so erscheint die Durchführung der Datenänderungen in der Datenquelle - sofern möglich - meist sinnvoller als die Korrektur des betroffenen Datums im integrierten Datenbestand. Manipulierte Datensätze können dann unter erneuter Durchführung der Datenintegration in den integrierten Datenbestand übernommen werden. Wird die Korrektur auf dem integrierten Datenbestand ausgeführt, so würde diese bei erneuter Integration der betroffenen Daten aus der ursprünglichen Quelle verloren gehen. Um Korrekturen auf den Datenquellen durchführen zu können, ist jedoch auch Kenntnis über die Herkunft der im integrierten Datenbestand enthaltenen Informationen erforderlich.

In Fällen, in denen die Datenquellen jedoch keinen stetigen Veränderungen unterliegen, erweist sich hingegen eine einmalige Integrationsstrategie zur Erstellung eines integrierten Datenbestandes meist als sinnvoll. Die Durchführung der Integrationsprozesse ist bedingt durch die integrationsinhärenten Schritte der Beschaffung, Bereinigung, Aufbereitung und Vereinheitlichung der Daten zudem aufwändig gestaltet. Im Hinblick auf solche statischen Datenquellen, erscheint daher der herkömmliche Weg des Data Warehousing, eine Durchführung der Korrekturen in den Quelldatenbeständen und eine anschließende erneute, aufwändige Integration der nur geringfügig korrigierten Quelldaten durchzuführen, meist nicht verhältnismäßig in Bezug auf den zu investierenden Aufwand. Vielmehr ist hier die Korrektur direkt am integrierten Datenbestand naheliegend und sinnvoll.

Zudem existieren integrierte Datenbestände, die Abzüge der zugrunde liegenden Datenquellen zu bestimmten Zeitpunkten beinhalten. Sollte ein Datenfehler nun der Integration nachgelagert im integrierten Datenbestand identifiziert werden, so ist nicht sichergestellt, dass die Änderung in der Datenquelle durchführbar ist, da dort das integrierte Datum gegebenenfalls nicht mehr vorgehalten wird [3]. Eine Änderung im integrierten Datenbestand ist in solchen Fällen alternativlos.

Auch der Umstand, dass die integrierten Daten oftmals nicht mit Metainformationen zur ursprünglichen Quelle annotiert sind, erschwert eine nachträgliche Korrektur in den Quellsystemen. Ohne Wissen über die Herkunft von Daten (Data Provenance [4]), ist es nicht möglich, die Quelle zu identifizieren und eine Manipulation auf dem Quelldatenbestand auszuführen. In diesen Fällen können Korrekturen oder Manipulationen ebenfalls nur am integrierten Datenbestand vorgenommen werden.

Temporale Aspekte

Ein weiteres Problem, welches sich im Kontext der zuvor motivierten Integration verschiedenartiger Quelldatenbestände häufig beobachten lässt, besteht darin, dass die resultierenden integrierten Datenbestände oftmals als einfache relationale Schemata (sog. Snapshot-Datenbanken [5]) aufgebaut werden, die keinerlei Voraussetzungen hinsichtlich temporaler Aspekte [6,7], also der Annotation der Realdaten mit einem Transaktionszeitbezug [6], aufweisen. So ist es ohne die Existenz temporaler Annotationen nicht möglich, durchgeführte

Manipulationen am Datenbestand zu einem späteren Zeitpunkt nachvollziehen zu können, da die Daten nicht zeitbezogen getrennt abgelegt, sondern überschrieben werden, und somit die ursprüngliche Belegung nach Durchführung einer Manipulation verloren geht.

Redundante Systeme zur Wartung mehrerer Datenbestände

Darüber hinaus existieren auf Seiten eines Datenintegrators oftmals mehrere heterogene, voneinander unabhängige und inhaltlich disjunkte Datenbestände, die für verschiedene Projekte und Auftraggeber gepflegt und gewartet werden müssen. Die Entwicklung maßgeschneiderter, individueller Lösungen zur Wartung und Anpassung der verschiedenen Datenbestände birgt den Nachteil des Bedarfs an mehreren, in weiten Teilen redundanten Lösungen. Im Vergleich zum Einsatz einer generischen Lösung kann dies im Hinblick auf den zu investierenden Entwicklungsaufwand ineffizient erscheinen, sofern durch die redundante Umsetzung identischer Aufgaben vermeidbare Kosten entstehen.

Abgeleiteter Bedarf

So resultiert aus der Anforderung der Durchführung von Datenänderungen am integrierten Datenbestand in Szenarien verteilter Verantwortlichkeiten der Bedarf an einem Vorgehensmodell, welches es den Datenverantwortlichen, also den ursprünglichen Urhebern und Verantwortungsträgern der Daten aus den Quelldatenbeständen, ermöglicht, eine Manipulation oder Aktualisierung der von ihnen verantworteten Daten zu initiieren. Dabei ist insbesondere zu beachten, dass Datenverantwortliche selten über detaillierte Kenntnis der Struktur des zugrunde liegenden Datenmodells und Datenbankschemas des integrierten Datenbestandes verfügen, und zudem nur in Ausnahmefällen in dessen Prozess der Erstellung eingebunden waren.

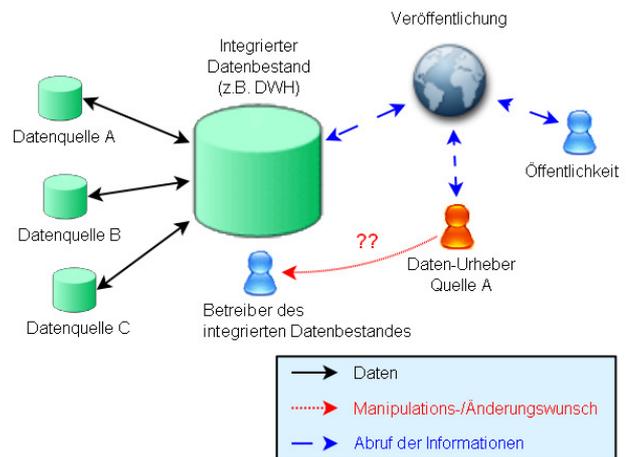


Abbildung 1: Bedarf an integrationsnachgelagertem Datenmanagement unter verteilten Verantwortlichkeiten

Abbildung 1 zeigt ein beispielhaftes Szenario, in dem ein Dienstleister die Inhalte dreier Datenquellen integriert und die Ergebnisse der Öffentlichkeit zur Verfügung stellt. Der Urheber der Datenquelle A identifiziert einen Datenmanipulationswunsch, der erst integrationsnachgelagert an den Betreiber des integrierten Datenbestandes übermittelt werden kann.

2. Problemstellung, Anforderungen und Lösungsansatz

Motiviert aus den in Abschnitt 1 aufgezeigten Umständen lässt sich folgende Problemstellung formulieren:

Ist eine Delegation von Datenmanagement in Szenarien verteilter Verantwortlichkeiten unter Berücksichtigung von Nachvollziehbarkeit der Änderungen durch Einsatz eines generischen Vorgehensmodells möglich?

Zur Beantwortung dieser Fragestellung besteht die Grundidee des vorgestellten Ansatzes in der Schaffung eines Vorgehensmodells, welches das Datenmanagement in Szenarien verteilter Verantwortlichkeiten unter Berücksichtigung der Historisierung der Dateninstanzen ermöglicht. Basierend auf den in Abschnitt 1 dargestellten Beobachtungen und Umständen werden verschiedene Anforderungen an das Vorgehensmodell abgeleitet, die im Folgenden aufgelistet seien:

- Das Datenmanagement muss durch Datenurheber initiiert sein, da dieser die Datenmanagement-Bedarfe identifiziert und deren Umsetzung fordert.
- Potentielle Datenmanipulationen müssen auch von „Nicht-IT-Experten“ und insbesondere ohne Detailkenntnis der Strukturen des zugrunde liegenden Datenbankschemas möglich sein, da die Urheber der Datenquelle in der Regel nicht über derartige Detailkenntnisse verfügen.
- Alle Datenmanipulationen, die durchgeführt werden, müssen umfassend dokumentiert werden und zu späteren Zeitpunkten nachvollziehbar sein, um ggf. auch zurückgenommen werden zu können (Reversibilität).
- Das Vorgehensmodell soll weitestgehend generisch auf beliebige, unterschiedliche relationale Datenbestände anwendbar sein, um individuelle Entwicklungen für verschiedene Datenbestände und damit verbundene Redundanzen in der Umsetzung der Lösungen zu vermeiden.
- Die Phasen des Vorgehensmodells sollen durch eine Sammlung von Softwarewerkzeugen unterstützt umgesetzt werden, um im Kontext eines realen Projektes im Praxisbetrieb unter realistischen Bedingungen evaluiert werden zu können.
- Aspekte struktureller Veränderungen der zugrunde liegenden Datenbankschemata sollen bei Entwurf und Umsetzung von Vorgehensmodell und unterstützender Werkzeugsammlung berücksichtigt werden; damit die jeweilige Umsetzung des Ansatzes hinsichtlich potentieller Änderungen oder Erweiterungen des Datenschemas ohne hohen Aufwand angepasst werden kann.
- Das Datenmanagement (insbesondere das Auffinden der modellierenden Struktur sowie die Spezifikation der zu manipulierenden Dateninstanz) muss auf Grundlage von natürlichsprachlichen Beschreibungen der modellierten Realwelt-Entitäten ermöglicht werden, da dem Urheber der Daten in Szenarien verteilter Verantwortlichkeiten nur diese, und nicht die Modellstruktur eines integrierten Datenbestandes bekannt ist.

3. Vorgehensmodell VD2M

Zur Lösung der in Abschnitt 2 beschriebenen Problemstellung wird das Vorgehensmodell VD2M vorgeschlagen, welches die Delegation von Datenmanagement-Anforderungen in Szenarien verteilter Verantwortlichkeiten ermöglicht, und dabei die in Abschnitt 2 formulierten Anforderungen (z.B. die Gewährleistung von Nachvollziehbarkeit und Reversibilität der Datenmanipulationen oder die Anpassbarkeit an verschiedene Datenbestände oder an strukturelle Änderungen des integrierten Datenbestandes unter Investition möglichst geringer Aufwände) erfüllt.

VD2M besteht aus zwei Gruppen von Phasen: initiale Phasen und Phasen des Regelbetriebs. Die initialen Phasen resultieren aus der Anforderung der einfachen Anpassbarkeit an verschiedene Einsatzszenarien und sind einmalig für jeden integrierten Datenbestand auszuführen, auf dem später Datenmanagement-Schritte durchgeführt werden sollen (folgend auch als „Anwendungsdatenbestand“ bezeichnet). Sie umfassen alle Aktivitäten und Aufgaben, die zur einmaligen, initialen Konfiguration der Werkzeuge anfallen. Als Ergebnis entsteht nach Ausführung der initialen Phasen ein konfiguriertes System, auf Basis dessen in den wiederkehrenden Phasen dann die Datenmanagement-Aktivitäten auf dem Anwendungsdatenbestand ausgeführt werden können.

3.1 Initiale Phasen

Im Rahmen der initialen Phasen bedarf es zunächst der Erhebung einiger Metainformationen über den zu verwaltenden Anwendungsdatenbestand, um:

- eine Historisierungsinfrastruktur zur automatisierten Historisierung der zu manipulierenden Daten zum Zwecke der Schaffung von Nachvollziehbarkeit und Reversibilität sowie
- einen Suchindex zum Auffinden von Relationen und Attributen im Datenmodell anhand von natürlichsprachlicher Beschreibungen von Realwelt-Entitäten

für einen spezifischen Anwendungsdatenbestand erzeugen zu können. Folgende initiale Phasen 1-3 lassen sich identifizieren:

1. **Quellenanalyse:** Erhebung von Metainformationen aus Struktur und Inhalt des integrierten Datenbestandes zur Vorbereitung der Generierung von Suchindex und Historisierungs-Infrastruktur.
2. **Infrastruktur-Generierung:** Automatisierte Erzeugung der Infrastruktur zur Historisierung der zu manipulierenden Daten auf Basis der in der vorigen Phase generierten Informationen.
3. **Suchindex-Generierung:** Semantische Indexierung des integrierten Datenbestandes durch teilautomatisiertes Vorgehen: automatisierte Analyse der Struktur unter Extraktion der enthaltenen Informationen (Attribut- und Relationennamen, Kommentartexte, sonstige Informationen) und Erzeugung eines Suchindexes auf Basis der extrahierten Strukturinformationen, und manuelle Anreicherung des Suchindexes durch Fachexperten

um Verschlagwortungen der im integrierten Datenbestand abgebildeten Domäne, um eine Stichwortsuche unter Nutzung von beschreibenden Stichworten einer Realwelt-Entität zum Auffinden des Speicherortes im Datenmodell zu ermöglichen.

3.2 Phasen des Regelbetriebs

Neben den drei initialen Phasen, die nur einmalig zur Anpassung an einen Anwendungsdatenbestand ausgeführt werden müssen, sind wiederkehrende Phasen zu identifizieren. Sie umfassen die zum eigentlichen Datenmanagement erforderlichen Aktivitäten und Aufgaben und sollen den Prozess des Datenmanagements umsetzen und dokumentieren. Zudem sind im Falle von Änderungen am Anwendungsdatenbestand Reorganisationsschritte erforderlich, um die Anpassungen, die am Anwendungsdatenbestand durchgeführt wurden, auch zugrunde liegende System zu propagieren. Folgende wiederkehrende Phasen 4-6 können im Regelbetrieb identifiziert werden:

4. Assistierte Datenmanagement: Durchführung von assistiertem Datenmanagement mitsamt Historisierung von ursprünglichen Belegungen im Anwendungsdatenbestand.
5. Dokumentation: Dokumentation der durchgeführten Schritte zum Datenmanagement (Datenmanipulationen und Datenpflege) im Anwendungsdatenbestand.
6. Reorganisation: Erweiterung/Anpassung des bestehenden Suchindexes an strukturelle Änderungen im Anwendungsdatenbestand analog zu den initialen Phasen, jedoch unter Erhalt manuell hinzugefügter Informationen.

Abbildung 2 zeigt zusammengefasst eine grafische Darstellung der initialen und wiederkehrenden Phasen des Vorgehensmodells in UML-Notation.

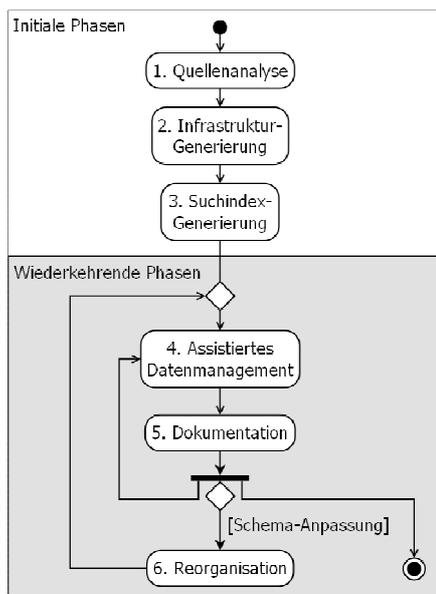


Abbildung 2: Das Vorgehensmodell VD2M

Durch das Durchlaufen der initialen Phasen erfolgt eine Anpassung des zur Umsetzung des Vorgehensmodells implementierten Systems an den integrierten Datenbestand. Die Infrastruktur zur automatisierten Historisierung sowie der Suchindex zur Abbildung der Realwelt-Entitäten auf die modellierten Strukturen des integrierten Datenbestandes werden im Rahmen der initialen Phasen auf Basis der individuellen Eigenschaften des integrierten Datenbestandes erzeugt.

In der Phase des assistierten Datenmanagements erfolgt anschließend im Rahmen des Regelbetriebs eine automatisierte Historisierung von Original-Belegungen der Daten. Zudem bietet die Phase Unterstützung bei der Lokalisierung der modellierten Strukturen im Anwendungsdatenbestand unter Nutzung des für den Anwendungsdatenbestand spezifisch und individuell erzeugten Suchindexes. Unter Nutzung von Historisierung der Manipulationen und Dokumentation der durchgeführten Schritte sind Nachvollziehbarkeit und Reversibilität der durchgeführten Datenmanipulationsschritte gewährleistet.

Für den Fall einer strukturellen Veränderung des integrierten Datenbestandes bietet die Phase der Reorganisation die Möglichkeit, unter Erhalt der bestehenden Informationen aus Historisierungsinfrastruktur und Suchindex die Generierung eines an die Strukturveränderungen angepassten Suchindexes und einer angepassten Historisierungsinfrastruktur umzusetzen.

3.3 Evaluation

Um die Eignung des vorgeschlagenen Vorgehensmodells VD2M zu evaluieren, werden die Phasen des Vorgehensmodells VD2M durch eine prototypisch implementierte Werkzeugsammlung WD2M (Werkzeuge zur Delegation von Datenmanagement) umgesetzt.

Aufgaben, die von WD2M zu unterstützen sind, bestehen zunächst in der Umsetzung der initialen Phasen zur Anpassung des Systems, deren Kern in der strukturellen Analyse des zu verwaltenden integrierten Datenbestandes besteht. Die dabei akquirierten Informationen dienen zum einen dem Aufbau des Suchindexes, in dem die Strukturen des Datenmodells um semantische Realwelt-Information annotiert werden, damit der Ort der Speicherung der gesuchten Information im Datenmodell (also der modellierenden Relation sowie deren Attributen) anhand von natürlichsprachlicher Stichwortsuche identifiziert werden kann. Zum anderen dienen die Ergebnisse der strukturellen Analyse dem Aufbau einer Historisierungsinfrastruktur, die die zeitbehaftete Speicherung der zu manipulierenden Datentupel ermöglicht, um Nachvollziehbarkeit und Reversibilität der Datenänderungen zu gewährleisten.

Auch die Schritte der wiederkehrenden Phasen von VD2M sind durch WD2M zu unterstützen und umzusetzen. Die primäre Aufgabe besteht in der Umsetzung des assistierten Datenmanagements, welches mehrere Aufgaben umfasst:

- Identifikation des Ortes der Speicherung der zu manipulierenden Information unter Nutzung des Suchindexes (also das Auffinden der zu referenzierenden Relation bzw. des darin enthaltenen Attributs, welches die zu manipulierende Information modelliert).

- Spezifikation der zu manipulierenden Dateninstanz (also des zu ändernden Tupels) anhand von instanzbeschreibenden Informationen.
- Editierbare Anzeige der aktuellen Tupelbelegung für zuvor spezifizierte Relation und Dateninstanz.
- Automatisierte Historisierung der ursprünglichen Tupelbelegung und Annotation eines Zeitstempels zur Schaffung von Nachvollziehbarkeit und Reversibilität der Datenmanipulation.
- Validierung und Speicherung des manipulierten Datentupels.

Weiter muss WD2M die Dokumentation aller benannten Schritte, aller getätigten Eingaben sowie der Manipulationen des Datenbestandes gewährleisten, um den eingangs benannten Anforderungen nachkommen zu können. Schließlich muss die Funktionalität zur Reorganisation von Suchindex und Historisierungsinfrastruktur umgesetzt werden, damit auf strukturelle Änderungen des zugrunde liegenden integrierten Datenbestands reagiert werden kann, die im Laufe der Zeit vor dem Hintergrund von Schemaevolution und Schemaversionierung auftreten können.

Die Evaluation der Werkzeuge wird im Kontext eines Projektes im Umfeld des Gesundheitswesens erfolgen. Das Projekt betreut einen integrierten Datenbestand, auf dem ein Web-Portal aufbaut, welches die Inhalte den Benutzern strukturiert aufbereitet zur Einsicht zur Verfügung stellt. Die Quellen, aus denen sich der integrierte Datenbestand speist, umfassen sowohl strukturierte Informationen im XML-Format als auch Daten, die per Web-Formular direkt über das angebundene Portal erhoben werden. Im Projekt liegen dahingehend verteilte Verantwortlichkeiten vor, als dass unterschiedliche Rollen, Urheberschaften und Interessen in den Bereichen Quelleninhalte, Betrieb des integrierten Datenbestandes sowie Darstellung der Daten im Web-Portal identifiziert werden können.

Der Einsatz von WD2M im Kontext des Projektes soll es dem Betreiber des integrierten Datenbestandes ermöglichen, integrationsnachgelagerte Datenmanagement-Anfragen auf Basis natürlichsprachlicher Beschreibung der modellierten Realwelt-Entitäten durchführen zu können. Durch die Vorgabe einer Menge exemplarischer, unterschiedlicher Manipulationsanfragen kann im Rahmen einer Studie erhoben werden, inwieweit die vorgeschlagenen Konzepte des Vorgehensmodells und deren Umsetzungen in Form der Werkzeuge WD2M zur Lösung der Problemstellung beitragen, d.h. ganz konkret, welche Quote von Anfragen unter Nutzung von VD2M/WD2M erfolgreich bearbeitet werden kann.

4. Zusammenfassung und Fazit

In diesem Beitrag wurde ein Lösungsansatz für die Frage vorgestellt, wie eine Delegation von Datenmanagement in

Szenarien verteilter Verantwortlichkeiten erfolgen kann. Zunächst wurde der Bedarf an integrationsnachgelagerten Datenmanagement-Schritten in Szenarien verteilter Verantwortlichkeiten motiviert. Anschließend erfolgte die Präzisierung der Problemstellung. Als Lösungsansatz wurde das Vorgehensmodell VD2M vorgeschlagen, dessen Phasen und Schritte vorgestellt wurden. Abschließend erfolgte die Beschreibung eines Evaluations szenarios, welches unter Einsatz der prototypisch implementierten Werkzeugsammlung WD2M quantitative und qualitative Aussagen über die Eignung des Ansatzes ermöglicht.

Durch den Einsatz des Vorgehensmodells VD2M ist die Delegation von Datenmanagement in Szenarien verteilter Verantwortlichkeiten unter Berücksichtigung der formulierten Anforderungen, dass das Datenmanagement durch den Datenurheber initiiert ist und von Nicht-IT-Experten auf Basis von natürlichsprachlicher Beschreibung von Realwelt-Entitäten durchgeführt werden kann, möglich. Die durch VD2M/WD2M durchgeführten Datenmanagement-Aktivitäten sind aufgrund der automatisierten Historisierung der Originalbelegungen und der Dokumentation des Gesamtprozesses vollständig nachvollziehbar und reversibel. Weiter sieht das Vorgehensmodell strukturelle Entwicklungen in Datenbeständen vor und kann entsprechend angepasst werden. VD2M kann durch Umsetzung der initialen Phasen weitestgehend generisch auf beliebige Datenbestände angewandt werden.

5. Literatur

- [1] Bauer, A. and Günzel, H. 2001. Data Warehouse Systeme. Architektur, Entwicklung, Anwendung. Dpunkt Verlag. ISBN 3932588762.
- [2] Inmon, William H. 1992. Building the Data Warehouse. John Wiley & Sons, Inc., New York, NY, USA. ISBN 0471569607.
- [3] Zeh, T. 2003. „Data Warehousing als Organisationskonzept des Datenmanagements: Eine kritische Betrachtung der Data-Warehouse-Definition von Inmon“, Inform., Forsch. Entwickl., Volume 18, Nummer 1, 32-38.
- [4] Buneman, P., Khanna, S. and Tan, W. C. 2000. “Data Provenance: Some Basic Issues”, FSTTCS, 87-93.
- [5] Stock, S. 2001. “Zeitbezogene Erweiterung des relationalen Datenmodells“, Wirtschaftsstudium, Vol. 30, Nummer 6, 843-852.
- [6] Snodgrass, R. T. and Ahn, I. 1985. “A Taxonomy of Time in Databases“, Proceedings of the 1985 ACM SIGMOD International Conference on Management of Data, Austin, Texas, May 28-31, 1985, 236-246. ACM Press, New York.
- [7] Harren, A. 2004. Temporale Datenintegration in Data-Warehouse-Systemen. dissertation.de - Verl. im Internet, ISBN 3-89825-868-8.