

# Einsatz von Dataspaces für die inkrementelle Informationsintegration in der Medizin

S.H.R. Wurst<sup>1</sup>, G. Lamla<sup>1</sup>, F. Prasser<sup>1,2</sup>, A. Kemper<sup>2</sup>, K.A. Kuhn<sup>1</sup>

<sup>1</sup> Lehrstuhl für medizinische Informatik

<sup>2</sup> Lehrstuhl für Datenbanksysteme

Technische Universität München

sebastian.wurst@tum.de

Die Bedeutung der Informationsintegration hat in der postgenomischen Ära massiv zugenommen, da phänotypische und genotypische Daten zusammengeführt werden können und die translationale Forschung unterstützt werden muss. [Ku08]

Für die Speicherung und Verarbeitung von strukturierten bzw. unstrukturierten Daten können unterschiedliche Paradigmen identifiziert werden. Bei einem Schema First Approach (SFA) erfolgt vorab ein Integrationsschritt für die Schemata der Komponentensysteme, was eine Erstellung komplexer Mappings und konsolidierter Schemata erfordert. Bei einer No Schema Approach (NSA) Integration werden alle Datenquellen direkt eingebunden, indem Schlüsselwortsuche und einfach strukturierte Abfragen darauf zur Verfügung gestellt werden. Eine semantische Integration der Datenquellen ist nicht notwendig, Abfragen besitzen jedoch auch keine präzise Semantik. Da das Erstellen, Abfragen und die gemeinsame Nutzung von strukturierten Daten aufwendig und teuer ist, besteht die Idee des Dataspace Ansatzes darin, die Vorzüge eines NSA mit einem SFA zu kombinieren, ohne die Vorzüge des SFA aufzugeben. Das Grundprinzip ist es, von Anfang an den Zugriff auf alle Daten zu unterstützen, dabei jedoch keine volle Kontrolle über die Daten auszuüben. Während herkömmliche Datenintegrationsansätze eine semantische Integration vor Erstellung von Services auf den Daten erfordern, stellt eine DSSP den arbeitsintensiven Aspekt der Datenintegration zurück, bis er absolut notwendig ist. Zusätzlicher Aufwand für eine engere Integration erfolgt inkrementell und bedarfsorientiert („pay-as-you-go“). [FHM05]

Zur Realisierung einer DSSP in der Medizin benötigt man ein Kernschema zur Identifikation von Patienten über Systemgrenzen hinweg und ein generisches Datenmodell, das die Integration von strukturierten, semi-strukturierten und unstrukturierten Daten ermöglicht. Data Mapping erfolgt durch Zusammenführen von Patienten aus unterschiedlichen Datenquellen. Schema Mapping erfolgt Schritt für Schritt durch Zusammenführen von Schemata bzw. Attributen durch Abbildung, Konvertierung von beispielsweise Einheit und Sprache, und durch Einbindung von Terminologien und Versionierung. Datenintegration erfolgt nach dem „pay-as-you-go“ Prinzip, aufbauend auf der Koexistenz von lose integrierten heterogenen Datenquellen und stark integrierten Kerndaten. Zum Nachweis der Machbarkeit des Ansatzes wurde eine prototypische Implementierung der beschriebenen Konzepte und Komponenten am Klinikum rechts der Isar in München durchgeführt.

[FHM05] Franklin, M. J.; Halevy, A. Y.; Maier, D.: From databases to dataspace: a new abstraction for information management. SIGMOD Record 34(4): S. 27-33 (2005).

[Ku08] Kuhn, K. A. et al.: Informatics and medicine. From molecules to populations. Methods Inf Med. 2008;47(4): S. 283-295.