

Matthias Ganzinger
Dr. sc. hum.

Datenintegration in biomedizinischen Forschungsverbänden auf Basis von serviceorientierten Architekturen

Promotionsfach: Medizinische Biometrie u. Informatik
Doktormutter: Prof. Dr. sc. hum. Petra Knaup-Gregori

In biomedizinischen Forschungsverbänden besteht der Bedarf, Forschungsdaten innerhalb des Verbundes und darüber hinaus gemeinsam zu nutzen. Hierzu wird zunächst ein Anforderungsmodell erstellt, das anschließend konsolidiert und abstrahiert wird. Daraus ergibt sich ein Referenzmodell für Anforderungen, welches anderen Forschungsverbänden als Grundlage für die beschleunigte Erstellung eines eigenen SOA-Systems dienen kann. Zum Referenzmodell wird weiterhin eine konkrete Instanz als Anforderungsmodell für den durch die Deutsche Forschungsgemeinschaft (DFG) geförderten Sonderforschungsbereich/Transregio 77 „Leberkrebs–von der molekularen Pathogenese zur zielgerichteten Therapie“ beschrieben. Aus dem Anforderungsmodell wird ein IT-Architekturmodell für den Verbund abgeleitet, welches aus Komponentenmodell, Verteilungsmodell und der Sicherheitsarchitektur besteht.

Die Architektur wird unter Verwendung des Cancer Biomedical Informatics Grid (caBIG) umgesetzt. Dabei werden die in den Projekten anfallenden Daten in Datendienste umgewandelt und so für den Zugriff in einer SOA bereitgestellt. Durch die Datendienste kann die Anforderung der Projekte, die Kontrolle über die eigenen Daten zu behalten, weitgehend erfüllt werden: Die Dienste können mit individuellen Zugriffsberechtigungen versehen und dezentral betrieben werden, bei Bedarf auch im Verantwortungsbereich der Projekte selbst. Der Zugriff auf das System erfolgt mittels eines Webbrowsers, mit dem sich die Mitarbeiter des Verbundes unter Verwendung einer individuellen Zugangskennung an einem zentralen Portal anmelden. Zum einfachen und sicheren Austausch von Dokumenten innerhalb des Verbundes wird ein Dokumentenmanagementsystem in die SOA eingebunden.

Um die Forschungsdaten aus verschiedenen Quellen auch auf semantischer Ebene integrieren zu können, werden Metadatensysteme entwickelt. Hierzu wird ein kontrolliertes Vokabular erstellt, das mit der hierfür entwickelten Methode aus den von den Projekten verwendeten Terminologien gewonnen wird. Die so gesammelten Begriffe werden mit standardisierten Vokabularen aus dem Unified Medical Language System (UMLS) abgeglichen. Hierfür wird ein Software-Werkzeug erstellt, das diesen Abgleich unterstützt. Des Weiteren hat sich im Rahmen dieser Arbeit herausgestellt, dass keine Ontologie existiert, um die in der biomedizinischen Forschung häufig verwendeten Zelllinien einschließlich ihrer Wachstumsbedingungen umfassend abzubilden. Daher wird mit der Cell Culture Ontology (CCONT) eine neue Ontologie für Zelllinien entwickelt. Dabei wird Wert darauf gelegt, bereits etablierte Ontologien dieses Bereichs soweit wie möglich zu integrieren.

Somit wird hier eine vollständige IT-Architektur auf der Basis einer SOA zum Austausch und zur Integration von Forschungsdaten innerhalb von Forschungsverbänden beschrieben. Das Referenzmodell für Anforderungen, die IT-Architektur und die Metadatenspezifikationen stehen für andere Forschungsverbände und darüber hinaus als Grundlagen für eigene Entwicklungen zur Verfügung. Gleiches gilt für die entwickelten Software-Werkzeuge zum UMLS-Abgleich von Vokabularen und zur automatisierten Modellerstellung für caBIG-Datendienste.