

Integration von Biobanken für Forschungsaufgaben

K.A. Kuhn¹, S.H.R. Wurst¹, D. Schmelcher¹, G. Lamla¹, F. Kohlmayer¹,
H.E. Wichmann²

¹ Lehrstuhl für medizinische Informatik
Technische Universität München
Klinikum rechts der Isar der TU München
Ismaninger Str. 22
D-81675 München
klaus.kuhn@tum.de

² Helmholtz Zentrum München – Institut für Epidemiologie
LMU – IBE Lehrstuhl für Epidemiologie
Ingolstädter Landstr. 1
D-85764 Neuherberg
wichmann@helmholtz-muenchen.de

Abstract: Zu den wichtigsten Ressourcen der translationalen Forschung zählen Biobanken, die in wohlorganisierter Weise Bioproben und zugeordnete klinische Information zur Verfügung stellen. Die Informationsintegration steht vor erheblichen Herausforderungen, da Sammlungen und Informationsquellen verteilt, fragmentiert, semantisch heterogen und unvollständig sind. Wir schlagen ein Stufenkonzept zur Integration vor, dessen Basiskomponente bereits realisiert ist und eingesetzt wird.

1 Motivation

In den Lebenswissenschaften verschwinden herkömmliche Grenzen zwischen Grundlagenforschung, klinischer Forschung und epidemiologischer Forschung. Die translationale Medizin eröffnet neue Einblicke in Krankheitsmechanismen und spielt eine immer größer werdende Rolle bei der Ermittlung persönlicher Risiken und der Festlegung personalisierter Therapien [HM07, KK08]. Zu den wichtigsten Ressourcen der translationalen Forschung zählen Biobanken, die in wohlorganisierter Weise Bioproben (Blut, Serum, Plasma, Urin, Gewebe, Zelllinien etc.) und zugeordnete klinische Information zur Verfügung stellen [YO07, WG07]. Diese Zusammenführung steht aber immer noch vor erheblichen Schwierigkeiten [HM07], u.a. weil Sammlungen und Informationsquellen verteilt, fragmentiert, semantisch heterogen und unvollständig sind [YO07]. Verschiedene Projekte haben diese Problematik erkannt und versuchen integrierte Infrastrukturen zu schaffen [YO07, HG07]. Für Deutschland hat sich der Gesundheitsforschungsrat kürzlich zum Thema Biobanken geäußert [BMBF08] und festgestellt, dass Biobanken ein unverzichtbares Instrument der krankheits- und

patientenorientierten klinischen Forschung sind. Europäische Initiativen zur Stärkung von Forschungsinfrastrukturen (European Strategy Forum on Research Infrastructures, ESFRI) tragen dieser gewachsenen Bedeutung von Biobanken Rechnung, so z.B. durch die Förderung des Projekts "Biobanking and Biomolecular Resources Research Infrastructure" [BBMRI]. Die bessere Vernetzung der deutschen Aktivitäten im Bereich "Biobanken" auf Bundes- wie auf Landesebene bildet dabei eine wesentliche Voraussetzung für die effektivere Nutzung der derzeit in Deutschland vorhandenen humanen Proben- und Datensammlungen. Im Rahmen eines Pilotprojekts für BBMRI in Deutschland werden modellhaft die Vernetzungsmöglichkeiten von Biobanken deutscher Forschungsverbände erprobt [BBMRI-DE]. Hierbei wird ein Konzept entwickelt, wie Biobanken, die im Rahmen ausgewählter Kompetenznetze und des NGFN aufgebaut wurden, vernetzt werden und somit in ihrer Nutzung eine Effektivitätssteigerung erzielen können. Dieses Konzept wird modellhaft mit einer Auswahl von geeigneten Biobanken aufgestellt und technisch umgesetzt. In diesem Artikel stellen wir einen Architekturentwurf zur Informationsintegration vor. Unter einer Portalanwendung soll ein Zugriff auf Komponentensysteme erfolgen. Die Portalanwendung befindet sich bereits im Einsatz; sie umfasst aktuell Daten über 191 europäische Biobanken [BBMRI-WP3].

2 Anforderungen

Gefordert wurde eine Portallösung, deren Ziel es ist, Forschern eine Übersicht über vorhandene Ressourcen zu geben. Dies soll eine gezielte Kontaktaufnahme für ein Kooperationsvorhaben ermöglichen.

Bei der Identifikation von Anwendungsfällen steht die Suche nach aggregierten Daten im Vordergrund: Gesucht werden die Anzahlen der in einzelnen Biobanken verwalteten Proben oder Personen auf der Basis charakterisierender Angaben wie Krankheitscode (ICD), Geschlecht, Altersgruppe, ethnische Herkunft, Vorhandensein von Verlaufsinformation, Materialtyp, Lagerbedingungen, Größe, Gewicht. In den ersten Ausbaustufen sollen keine Identifikatoren zu Proben oder Personen an das Portal übermittelt werden. Eine Verfolgung von Personen über verschiedene Biobanken, die eine übergeordnete Identifikation erforderlich machen würde, ist derzeit nicht vorgesehen, konzeptionell aber möglich. Der vorgestellte Stufenplan ist offen für Erweiterungen. Die Übermittlung von anonymisierten oder pseudonymisierten Daten zu Proben oder Personen an das Portal bzw. an Forscher oder teilnehmende Biobanken kann ebenfalls realisiert werden.

3 Architektur

Die Portalapplikation stellt die Benutzerschnittstelle des Systems dar. Sie dient der Verwaltung von Biobank-Metadaten, deren Eingabe über Web Formulare, per File Upload oder automatisch mit Hilfe eines Dienstes erfolgen kann. Das Stufenkonzept wird in den Abbildungen 1-4 beschrieben. Abbildung 1 beschreibt eine minimale

Möglichkeit der Integration: Vordefinierte Anfragen werden manuell an die Biobanken gestellt. Die Ergebnismenge der Anfrage wird mittels eines Integrationstools aggregiert und im Portal den Forschern zugänglich gemacht. In einem weiteren Schritt kann der manuelle Vorgang automatisiert werden: In Abbildung 2 wird das Konzept auf ein föderiertes Schema erweitert. Die lokalen Biobanken stellen eine Schnittstelle zum Zugriff auf eine gemeinsame Teilmenge von Attributen zur Verfügung, welche von lokal implementierten Biobankdiensten genutzt wird. Anfragen von der Portalapplikation werden durch den Integrationsdienst an die jeweiligen lokalen Biobankdienste verteilt, die zurück gelieferten Daten werden vom Biobankdienst aggregiert und aufbereitet. Anschließend werden die lokalen Ergebnisse der einzelnen Biobanken vom Integrationsdienst zu einer globalen Ergebnismenge zusammengeführt. In Abbildung 3 werden die Daten der Biobanken nicht lokal, sondern im zentralen Integrationsdienst aggregiert. In der letzten Ausbaustufe (Abbildung 4) werden alle Attribute einer lokalen Biobank in ein generisches Schema transformiert und exportiert.

In den Stufen 2-4 wird ein Registrierungsdienst vorgeschlagen: Es handelt sich um einen Verzeichnisdienst zur Registrierung der Biobankdienste, der beispielsweise mittels UDDI oder WS-Discovery realisiert werden kann. Er beinhaltet die Abbildung der BiobankID (verwaltet durch Portalapplikation) auf die zugehörige Webservice Endpunkt - URL (URL des Biobankdienstes). Der typische Ablauf der Registrierung einer Biobank und des zugehörigen Biobankdienstes umfasst folgende Schritte: Nach Registrierung der Biobank bei der Portalapplikation wird eine eindeutige ID für die Biobank vergeben. Anschließend werden Biobank-Metadaten erfasst, die durch die Portalapplikation verwaltet werden. Auf Seiten der Biobank kann nun der Biobankdienst implementiert werden. Sobald dieser Dienst läuft, meldet er sich selbständig beim Registrierungsdienst an und wird bei der Beantwortung zukünftiger Anfragen mit berücksichtigt. Weitere zusätzliche Dienste sind denkbar, die den Integrationsdienst unterstützen können. Sie könnten z. B. für semantisches Mapping (Terminologie, Versionierung) oder für Caching, Indexing zuständig sein. Ein wesentlicher Unterschied zwischen den Stufen 1 und 2 im Vergleich zu 3 und 4 ist die Übermittlung von aggregierten Daten anstatt der Übermittlung von Probanddaten. Die Stufen 3 und 4 ermöglichen höher differenzierte Abfragen, sind aber datenschutztechnisch komplexer und machen Sicherheitsmaßnahmen (u.a. De-Identifikation, k-anonymity) erforderlich. Zudem ist eine Autorisierung für den Zugriff auf die Probanddaten notwendig, die unter Umständen spezielle Vereinbarungen oder Verträge zwischen den beteiligten Biobanken erforderlich macht.

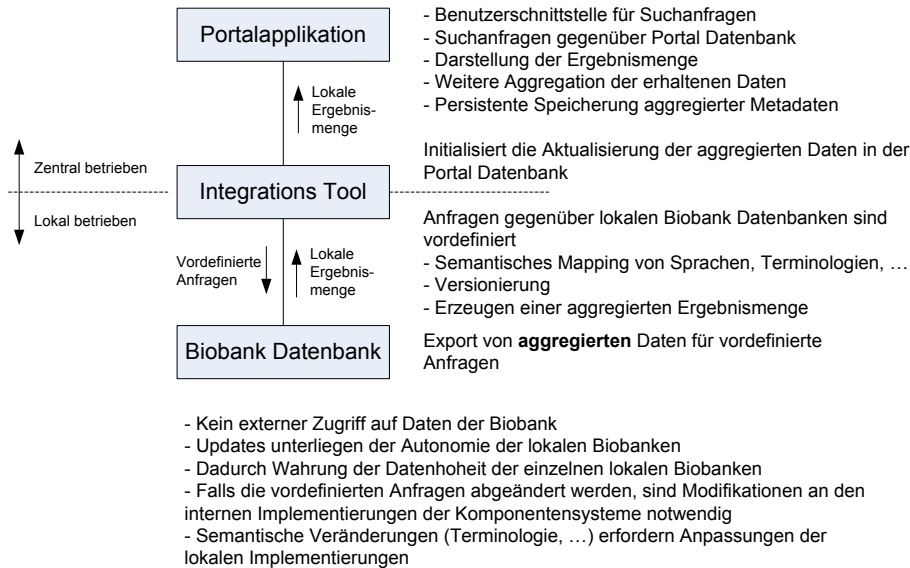


Abbildung 1: vordefinierte Anfragen, lokale Datenaggregation

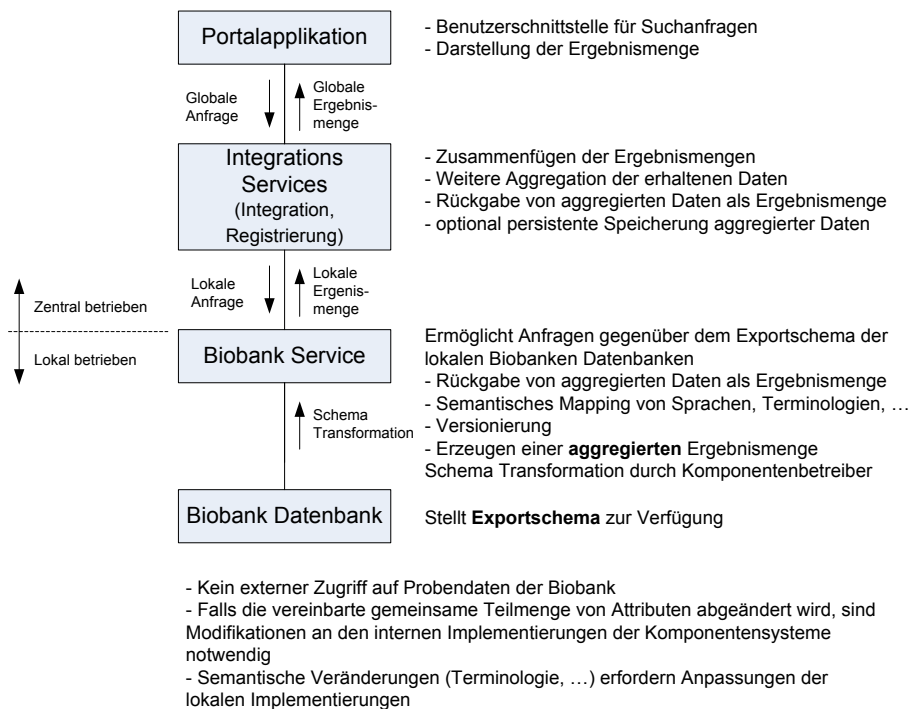


Abbildung 2: föderiertes Schema, lokale Datenaggregation

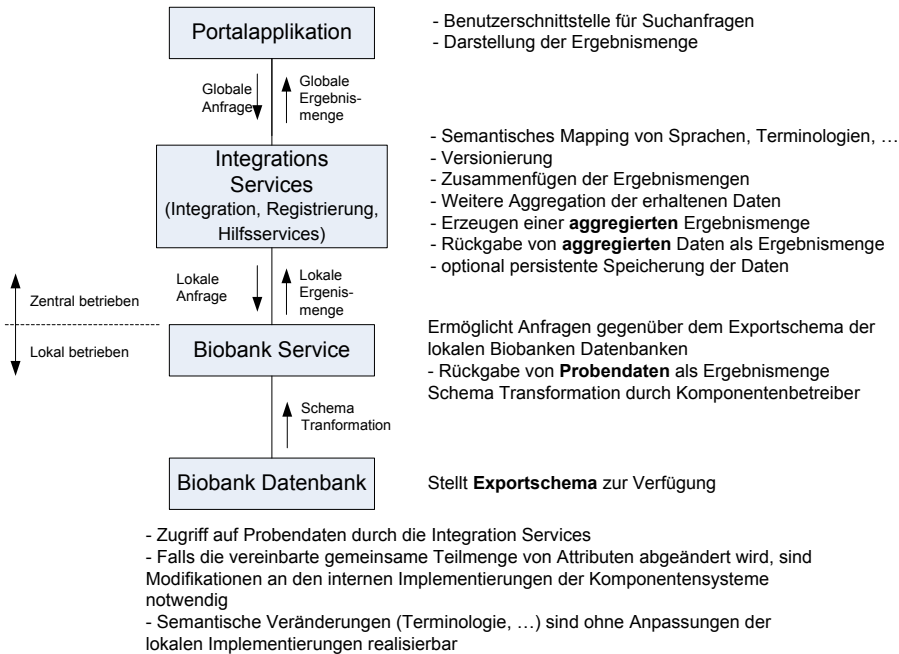


Abbildung 3: föderiertes Schema, zentrale Datenaggregation

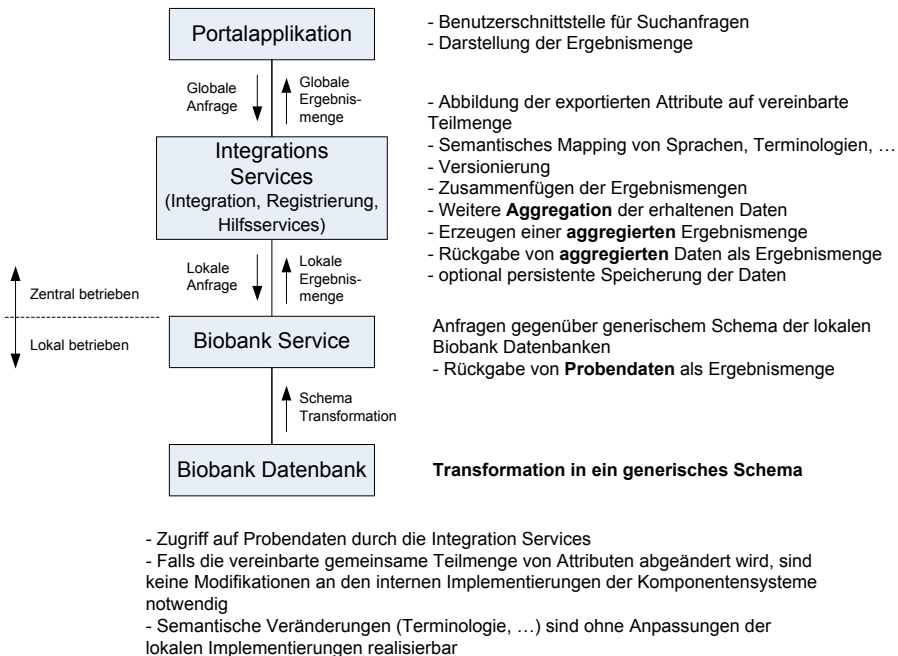


Abbildung 4: generisches Schema, zentrale Datenaggregation

4 Umsetzung

Für die Entwicklung der webbasierten Portalapplikation wird ein agiler Softwareentwicklungsprozess mit kurzen Feedbackzyklen unter enger Einbeziehung der Projektpartner gewählt, der werkzeuggestützt hohe Entwicklungsgeschwindigkeit, Anpassbarkeit und Wiederverwendbarkeit gewährleistet. Zur Verwendung kommen Open-Source-Software und Komponentenbibliotheken: Java unter Verwendung von Hibernate [Hibernate] und JSF [JSF]. Es ist kurzfristig vorgesehen, das in Abbildung 1 gezeigte Szenario in den Einsatz zu bringen. Weitere Ausbaustufen sind in Abstimmung mit dem EU-Projekt vorgesehen.

Literaturverzeichnis

- [BMBF08] BMBF-GFR (Gesundheitsforschungsrat) 2008 Stärkung der Infrastruktur in der Gesundheitsforschung: Biomaterialbanken. Entschließung der 26. Sitzung am 12.12.2008
http://www.gesundheitsforschung-bmbf.de/_media/Beschluss_Biobanken_090216.doc
- [BBMRI] Biobanking and Biomolecular Resources Research Infrastructure, Website 2009; <http://www.bbmri.eu/>
- [BBMRI-DE] BBMRI-DE: Pilotstudie Biobanken deutscher Kompetenznetze. Modellhafte Erprobung der Vernetzungsmöglichkeiten von Biobanken deutscher Forschungsverbände, Gefördert vom BMBF
http://www.bbmri.de/test1/ziel_biobanken.html
- [BBMRI-WP3] Website des Katalogs für krankheitsbezogene Biobanken (WP3) in BBMRI 2009:
<http://www.bbmri.eu/index.php/catalog-of-european-biobanks>
- [HG07] Hibbert M, Gibbs P, O'Brien T, Colman P, Merriel R, Rafael N, Georgeff M. The molecular medicine informatics model (MMIM). In: Kuhn K.A. et al (eds) MEDINFO 2007: 1230-4.
- [Hibernate] Hibernate [Website] 2009; <http://www.hibernate.org/>
- [HM07] Hörig H, Marincola E., Marincola F. Obstacles and Opportunities in Translational Research. *Nature Medicine*, 2007 11(7): 705-708.
- [JSF] Java Server Faces [Website] 2009; <http://java.sun.com/javaee/javaserverfaces/>
- [KK08] Kuhn KA, Knoll A, Mewes HW, Schwaiger M, Bode A, Broy M, Daniel H, Feussner H, Gradinger R, Hauner H, Höfler H, Holzmann B, Horsch A, Kemper A, Krcmar H, Kochs EF, Lange R, Leidl R, Mansmann U, Mayr EW, Meitinger T, Molls M, Navab N, Nüsslin F, Peschel C, Reiser M, Ring J, Rummeny EJ, Schlichter J, Schmid R, Wichmann HE, Ziegler S.: Informatics and medicine--from molecules to populations. *Methods Inf Med* 2008; 47(4):283-95
- [WG07] Wichmann HE, Gieger C. Biobanken. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2007 Feb;50(2):192-9.
- [YO07] Yuille M, van Ommen GJ, Brechot C, Cambon-Thomsen A, Dagher G, Landegren U, Litton JE, Pasterk M, Peltonen L, Taussig M, Wichmann HE, Zatloukal K. Biobanking for Europe. *Brief Bioinform*. 2008;9:14-24