

Schlussbericht

Zuwendungsempfänger: Freie Universität Berlin - Fachbereich Mathematik und Informatik

Förderkennzeichen: 03FO2032

Vorhabenbezeichnung: „ForMaT2: Information –based Medicine“

Laufzeit des Vorhabens: 2 Jahre (+ 5-monatige Verlängerung)

Berichtszeitraum: 01.01.2010 bis 31.05.2012

Freie Universität Berlin

Fachbereich Mathematik und Informatik

BioComputing Gruppe

Prof. Dr. Christof Schütte

Arnimallee 6

14195 Berlin

Telefon: +49 (0) 30 838-75367

Telefax: +49 (0) 30 838-75412

Mail: schuette@mi.fu-berlin.de

Inhaltsverzeichnis

I.	Kurzdarstellung	1
1.	Aufgabenstellung und Voraussetzungen zur Vorhabendurchführung.....	1
2.	Planung und Ablauf des Vorhabens	3
3.	Darstellung des wissenschaftlichen und technischen Standes bei Projektbeginn	5
a)	Benutzte Konstruktionen, Verfahren und Schutzrechte	5
b)	Verwendete Literatur (Fachliteratur, Informations- und Dokumentationsdienste).....	7
4.	Zusammenarbeit mit anderen Stellen	9
II.	Eingehende Darstellung	10
1.	Zuwendungsverwendung	10
2.	Notwendigkeit und Angemessenheit der geleisteten Arbeit.....	11
3.	Nutzen und Verwertbarkeit der Ergebnisse.....	11
4.	Fortschritte Externer während der Vorhabendurchführung.....	13
5.	Erfolgte oder geplante Veröffentlichungen der Ergebnisse.....	15

I. Kurzdarstellung

1. Aufgabenstellung und Voraussetzungen zur Vorhabendurchführung

Ausgangslage: Gigantische Datenmengen werden erfasst – und bleiben ungenutzt

Fast täglich erschließt die Forschung neue Informationsquellen über den menschlichen Organismus und die in ihm ablaufenden Prozesse. Doch die explosionsartig wachsende Datenmenge stellt nicht nur eine Herausforderung für die Computerwissenschaften dar, sondern auch für die Medizin selbst: Die erhobenen Massendaten werden nur minimal genutzt; medizinisch entscheidend wichtige Zusammenhänge gehen im Datenmeer unter. Für die Medizin wird sich die Nutzung der versteckten Möglichkeiten zu einem starken Wachstumssektor entwickeln, in dem hohes wirtschaftliches Potential steckt. Viel genauere, patientenspezifische Diagnosen und Therapien werden in Zukunft möglich sein, für die auch eine hohe individuelle Zahlungsbereitschaft besteht.

Obwohl die potentiellen Möglichkeiten nicht grundsätzlich in Frage stehen, wurden auf diesem Gebiet in den vergangenen Jahren viele Erwartungen enttäuscht. So ist etwa die Euphorie um Biomarker, die auf Proteomanalysen basieren, deutlich abgeflaut, weil immer neue technische Schwierigkeiten in der Datenauswertung auftraten und sich viele als Durchbruch gefeierte Studien nicht reproduzieren ließen. Diese Probleme ergeben sich jedoch meist nicht aus einer prinzipiellen Unmöglichkeit, sondern aus Qualitätsmängeln bei der Datenerfassung und -auswertung.

Hier kann die Mathematik Entscheidendes leisten. Gemeinhin als eher praxisfern bekannt, sind ihre Ansätze besonders im Bereich der **Information-based Medicine** wegweisend. Als Teil des DFG-Research-Centers MATHEON, das „Mathematik für Schlüsseltechnologien“ entwickelt, und als Initiator des Master-Studiengangs „Bioinformatik“ an der FU Berlin steht die BioComputing Gruppe für anwendungsgetriebene, interdisziplinäre Forschung an der Schnittstelle zwischen Mathematik / Informatik und Medizin / Pharmakologie mit einem Fokus auf Modell- und Datenintegration.

Unsere Lösung: Qualitätsgesicherte Erfassung und Auswertung von Daten

Übergeordnete Zielsetzung der Prototypenentwicklung im Rahmen des ForMaT-Vorhabens war die patientenspezifische, qualitätsgesicherte Datenerfassung (Dateneingang), deren Analyse und eine domänenspezifische Reportgenerierung (Datenausgang) **für Mediziner, Pharmazeuten und in diesem Bereich tätige Unternehmen**. Der Einsatz dieser Technologien wurde auf folgenden Verwertungsfeldern entwickelt:

- Proteombasierte Fingerprints zur Krankheitsfrühdagnose
- Nicht-invasive Frühdagnose von Herz-Kreislauf-Erkrankungen
- Pharmakokinetische Modellierung in der Wirkstoffentwicklung

Alle drei Ansätze basieren auf einem gemeinsamen Fundament: Der **intelligenten Auswertung von Massendaten**. Die vorgestellten Verwertungsansätze sind die drei ersten Anwendungen der gemeinsamen Grundlage.

Die Konzeption des Innovationslabors beruhte zentral auf den im Folgenden kurz beschriebenen Elementen:

Element 1: Zentrale Auswertungsplattform für alle Verwertungsansätze

Das in den letzten Jahren in unserer Forschung entwickelte Erfolgsmodell zur **Qualitätssicherung der Datenauswertung** beruht auf der **konsequenten Datenintegration**, die sich im Rahmen einer Qualitätsprüfung sowohl auf Daten aus verschiedenen Quellen (z.B. Genom und Proteom), als auch

auf die Integration von großen Patientenpopulationen bezieht. Das Kernstück des Innolabs bestand daher in der zentralen Auswertungsplattform. Sie stellte eine den drei Verwertungsgebieten gemeinsame Pipeline von Datenauswertungs-Algorithmen zur Verfügung, in der – anwendungsabhängig – geeignete Methoden für alle Verfahrensschritte von der Dateneinlese über die Auswertung bis zur Reporterstellung hintereinandergeschaltet und zu automatischen Abläufen verknüpft werden konnten. Einlese und Reporterstellung erfolgten also stark anwendungsbezogen, die Auswertungsverfahren im Inneren der Pipeline waren jedoch für alle drei Verwertungsgebiete gleich.

Element 2: Qualitätssicherung bereits in der Datenerfassung

In Anbetracht der eingangs beschriebenen Probleme wurde das Angebot des Innovationslabors nicht nur auf eine kompetente Datenauswertung beschränkt, sondern enthielt bereits die **Qualitätssicherung der Datenerfassung** von Anfang an und in jedem Schritt als gleichwertiges Element. Die gleichrangige Kombination dieser beiden Zielaspekte wurde uns nicht nur von allen befragten Experten als essentiell benannt, sondern macht auch in aktuellen Projekten der BioComputing Gruppe den Erfolg aus. Dies hat entscheidende Auswirkungen auf die Verwertungsstrategien innerhalb des Innovationslabors: Die notwendige Maßnahme zur Qualitätssicherung der Datenerfassung ist ihre **Standardisierung** (Befolgung eines Protokolls, Geräteeichung). Das Einhalten des Standards konnte nur durch eine Zertifizierung der Partner garantiert werden.

Element 3: Automatisch steigende Auswertungsqualität

Die Auswertungsqualität hängt entscheidend von unseren Verfahren zur Datenintegration ab und steigt mit der **Menge der Qualitätsdaten**. Deshalb wurde zunächst zur Kalibrierung und Validierung der Auswertungsmechanismen der notwendige Mindestumfang an Daten bereitgestellt, der für eine qualitativ hochwertige Auswertung erforderlich ist. Anschließend wurde die Datenmenge dann gemeinsam mit den Verwertungspartnern konsequent erhöht. Dieses Vorgehen bewirkte, dass sich die Qualität unserer Dienstleistung so zwangsläufig stetig verbesserte.

Element 4: Interdisziplinarität und Netzwerk

Um die vielfältigen Aufgaben zu meistern, wurde unsere mathematisch-informatische Kernkompetenz durch ein **interdisziplinäres Team** aus experimentell arbeitenden Wissenschaftlern der Biowissenschaften und Medizin und ein **starkes Kooperationsnetzwerk** aus medizinischen und pharmazeutischen Partnern ergänzt.

Vorgehensweise und Ziele: Marktorientierte Forschung in allen drei Ansätzen

Die drei genannten Verwertungsgebiete resultierten aus einer breit angelegten Analyse des Marktes im Rahmen der ersten ForMaT-Phase. Sie wurden ausgewählt, weil sie folgende Aspekte kombinierten: (a) Kombination von Datenintegration und Standardisierung der Datenerfassung, (b) Erfordernisse von interdisziplinärer Forschung zeitnah erfüllbar, (c) mittelfristige Umsetzung der wissenschaftlichen Ansätze in die wirtschaftliche Praxis möglich, (d) Konkurrenzsituation und technische/fachliche Anforderungen sprechen nicht gegen Markteintritt und (e) hohe Erfolgsaussichten aufgrund von großem Marktpotential.

Zielsetzungen der Verwertungsansätze im Rahmen der ForMaT-Förderung waren:

Im Verwertungsgebiet „**Proteom-basierte Fingerprints**“ verfügten wir bereits über sehr umfangreiche Vorarbeiten in einem oft kritisch betrachteten und recht offenen Marktsegment. Zusammen mit klinischen Partnern haben wir Standards der Blutabnahme und -bearbeitung entwickelt, die den problematischen Einfluss der Proteaseaktivität auf die Diagnose minimieren. Zusätzlich existierte bereits mit Projektbeginn eine rudimentäre Version einer Datenverarbeitungspipeline mit einem breiten Repertoire an mathematischen Methoden zur Fingerprint-Erkennung und Techniken zur multidimensionalen Massendatenverarbeitung von Tausenden von Massenspektren. Forschungsziel: Weiterentwicklung

der bisherigen Pipeline bis zum professionellen Einsatz bei der Analyse von standardisierten Blutproben in der klinischen Diagnostik; dabei vor allem verbessertes Verständnis des Einflusses von Geräteparametern, verlässliche Signifikanzanalyse von Fingerprints und Entwicklung weiterer Biomarker.

Zielsetzung auf dem Verwertungsgebiet „**Nicht-invasive Frühdiagnose von Herz-Kreislaufkrankungen**“ war die Entwicklung eines medizintechnischen Geräts für Check-up-Untersuchungen, das zur Früherkennung von Strombahnhindernissen eingesetzt und direkt an medizinische Einrichtungen vertrieben werden kann. Forschungsziel: Die Verbesserung eines bereits vorhandenen Verfahrens, das die physiologischen Eigenschaften des menschlichen Herz-Kreislaufsystems beschreibt. Es sollte Rückschlüsse von den mittels Drucksensoren gemessenen Beobachtungszeitreihen auf die genaue Lage und Ausprägung von Strombahnhindernissen erlauben.

In dem Verwertungsfeld „**Pharmakokinetischen Modellierung**“ sollten unsere Forschungserfahrungen erweitert werden, um eine umfangreiche Modell- und Datenintegrationsbibliothek zu erstellen. Ziel war die Erstellung eines Softwareprodukts, das an die forschenden Pharmaunternehmen auslizenzieren und/oder für entsprechende Beratungsleistungen verwendet werden soll. Forschungsziel: Weiterentwicklung der bisherigen *in silico* Werkzeuge, um die Wirkungen von pharmazeutischen Substanzen im menschlichen Organismus abzubilden und dabei möglichst spezifisch auf die physiologischen Eigenschaften unterschiedlicher Patientengruppen eingehen zu können. Diese zu Beginn des ForMaT-Projektes definierte Zielsetzung wurde im Laufe des Projektfortschritts nachträglich auf die Anwendung der Wirkstoffentwicklung in der präklinischen und den ersten klinischen Phasen fokussiert.

Die ausgewählten Forschungsansätze sollten in der zweiten Phase von ForMaT mit Ausrichtung auf spezifische Markt- und Kundenanforderungen innerhalb des Innovationslabors (InnoLab) „Math for Diagnostics“ weiterentwickelt werden. Dies sollte durch die gemeinsame Umsetzung konkreter Verwertungskonzepte durch Forscherteams (Angiologie, Proteomik, Pharmakokinetik) und eine kaufmännische Projektunterstützung erfolgen. Erklärtes Ziel jedes der Konzepte war es, marktorientiert zu forschen und so die verfügbaren wissenschaftlichen Methoden in der Praxis einsetzbar zu machen. Als Ergebnis sollten, neben dem Ausbau des technischen Standes, am Zielmarkt orientierte Geschäftsmodelle entwickelt werden.

Basierend auf den beschriebenen Vorarbeiten und der Aufgabenstellung wurden das vorläufige Verwertungskonzept sowie eine Strategiefestlegung erstellt. Eine Budgetplanung auf Basis eingeholter Kostenvoranschläge, sowie erste Ablauf- und Meilensteinplanungen konnten erarbeitet werden.

2. Planung und Ablauf des Vorhabens

Die Planung der Arbeiten des InnoLabs wurde zunächst für jedes Verwertungsfeld separat aufgestellt, um schließlich mit einander verzahnt und bzgl. der Weiterentwicklung und Nutzung der geteilten Analyseplattform optimiert zu werden. Die Tätigkeiten der kaufmännischen Projektassistenz fanden in enger Abstimmung mit den jeweiligen Projektteams statt und waren stets an den inhaltlichen Fortschritt gekoppelt. In gruppenübergreifenden Status-Meetings wurden Fortschritte und Ergebnisse präsentiert und diskutiert, die weitere Umsetzung geplant und eventuelle Anpassungen vorgenommen.

Die Zielsetzung der Planung der Projektarbeiten des Verwertungsfeldes Proteomics bestand aus folgenden grundlegenden Modulen: (1) Indikationswahl, (2) Identifikation von proteinbasierten Diskriminierungsmerkmalen und (3) der Weiterentwicklung und der Professionalisierung der Algorithmik. Zur Bearbeitung des ersten Meilensteins wurde zunächst ein Kriterienkatalog zur Ermittlung des medizinischen Bedarfs und gleichzeitiger Erfüllung ökonomischer Gesichtspunkte entworfen. Basierend auf entsprechender Marktforschung wurde Lungenkrebs als zu untersuchende Erkrankung identifiziert. In Zusammenarbeit mit einem klinischen Partnerunternehmen und einem Labor konnten spezifizierte Proben verarbeitet und analysiert werden. Für die Validierung der Vergleichbarkeit der Ergebnisse wurden sog. Standardprotokolle in Form von Arbeits- und Verfahrensanweisungen sowohl für das